Low level feature detection in human vision

Stuart Andrew Wallis

Doctor of Philosophy

ASTON UNIVERSITY

June 2009

Aston University

Low level feature detection in human vision

Stuart Andrew Wallis

Doctor of Philosophy

June 2009

Summary

This thesis is concerned with the basic building blocks of human visual perception, which are investigated through psychophysical experiments coupled with computational modelling of the proposed underlying processes.

Influential models of edge detection have generally supposed that an edge is detected at peaks in the 1st derivative of the luminance profile, or at zero-crossings in the 2nd derivative. However, when presented with blurred triangle-wave images, observers consistently marked edges not at these locations, but at peaks in the 3rd derivative. This new phenomenon, termed 'Mach edges' persisted when a luminance ramp was added to the blurred triangle-wave. Modelling of these Mach edge detection data required the addition of a physiologically plausible filter, prior to the 3rd derivative computation. A viable alternative model was examined, and rejected on the basis of data obtained with short-duration, high spatial-frequency stimuli.

Detection and feature-marking methods were used to examine the perception of Mach bands in an image-set that spanned a range of Mach band detectabilities. A scale-space model that computed edge and bar features in parallel provided a better fit to the data than 4 competing models that combined information across scale in a different manner, or computed edge or bar features at a single scale.

The perception of luminance bars was examined in 2 experiments. Data for one image-set suggested a simple rule for perception of a small Gaussian bar on a larger inverted Gaussian bar background.

In previous research, discriminability (d') has typically been reported to be a power function of contrast, where the exponent (p) is 2 to 3. However, using bar, grating, and Gaussian edge stimuli, with several methodologies, values of p were obtained that ranged from 1 to 1.7 across 6 experiments. This novel finding was explained by appealing to low stimulus uncertainty, or a near-linear transducer.

Key words: psychophysical methods, derivatives, Mach edges, Mach bands, non-linear transducer

Acknowledgements

I would like to thank Mark Georgeson for all aspects of his supervision. His approachability, infectious enthusiasm, clarity of communication, patience and all-round helpfulness have contributed greatly to my immensely enjoyable PhD experience.

I would also like to thank the other members of our lab for all of their help and support, and (for most of them) their many hours of patiently sitting as a subject - Tim Meese, Rob Summers, Tim Yates, Daniel Baker and Kirsten Challenor.

Finally, I would like to thank my wife, Sue, for her support during my PhD, and for reminding me that there is a world outside Vision Science.

Contents

List of Figures

List of Tables

Chapter 1

Introduction to the thesis and to feature detection

1.1 Thesis overview

This thesis addresses the issue of how edges and bars are processed, early in the human visual pathway. Psychophysical methods will be applied with images that vary along one dimension, whose luminance profiles are carefully-designed, in an attempt to identify the cues that observers use. Computational modelling will be used in an attempt to infer underlying visual processes. An outline of this thesis follows.

Chapter 1 introduces the reader to some of the previous literature, with a particular focus on edge detection. It also introduces signal detection theory, which is applied in several results sections.

Chapter 2 provides the methodological details that apply to most of the experiments. A psychometric fitting program is also described here, as it is used with most of the data in chapter 5 onwards.

Chapter 3 starts the experimental work, by first examining whether observers can perceive edges where there are no peaks in the 1st derivative, nor zero-crossings in the 2nd derivative. Observers *do* perceive these 'Mach edges', and their locations are nicely predicted by a non-linear 3rd derivative model. Next a yes-no method is used to strenuously test this 3rd derivative approach with an image set that contains a classic Mach ramp stimulus. The model predicts one edge for some stimuli where observers perceive two. So the model is refined by the addition of a pre-filter whose properties are based on retinal ganglion cells, and this revised version fits the data well. A feature-marking method is then applied with the same image set. The resulting data support the revised model, but also a competing model. However, the results of a final experiment allow rejection of the competing model, leaving a 3rd derivative approach as the most likely account of Mach edges.

Chapter 4 addresses the perception of Mach bands. Data are collected using detection and feature-marking tasks with a set of edges that vary along one dimension, some of which produce Mach bands. Five different models are fitted to the data. Previous research suggests a primary role for the 2nd derivative of the luminance profile, so all 5 models employ 2nd derivative filters, or approximations thereof. The best-fitting model is a scale-space edge-bar model.

Chapter 5 covers the perception of luminance bars in 2 sets of images. The first is a compound (f + 3f) grating, which is a sinusoidal grating added to its 3rd harmonic. The second image set consists of a narrow Gaussian bar on an inverted wide Gaussian bar background. Observers indicate the presence or absence of a thin central bar at various contrast levels. For the f + 3f stimuli, the data from one observer are dissimilar to the data from the other observer. However, for the Gaussian stimuli, the data from both observers suggest a simple rule for detection of the thin bar. Novel near-linear performance is observed at low contrasts.

Chapter 6 is the longest and final experimental chapter. It examines the degree of linearity of low contrast detection performance in a series of 6 experiments. In previous research, discriminability (d') is typically reported to be a power function of contrast, where the exponent (p) is 2 to 3. However, using bar, grating, and Gaussian edge stimuli, and a number of different methodologies, values of p are obtained that range from 1 to 1.7, where a value of 1 indicates totally linear performance. Such a low value of p is a very novel finding. It is explained by appealing to low stimulus uncertainty, and this account is supported by modelling (although a transducer with exponent p is also a viable account). Support is also presented for independent channels for dark and light features, and for some predictions of signal detection theory that have received little previous attention in visual psychophysics.

Chapter 7 summarises the thesis and suggests the direction of future research.

1.2 Existing publications

Some of the material in this thesis has appeared in previous publications. A large part of chapter 3 has been published as a journal article (Wallis & Georgeson, 2009), and smaller sections of chapter 3 have been presented in poster form at 3 conferences; two at AVA (Wallis & Georgeson, 2007a,c), and one at ECVP (Wallis & Georgeson, 2007c). From chapter 6, experiment 10 has been presented at 2 conferences; as a poster at AVA in the UK (Wallis & Georgeson, 2007b), and as a talk at VSS in the USA (Georgeson & Wallis, 2008). Some of the figures from the journal article and the posters have been reproduced in this thesis. However, all of the material presented in this thesis is my own work, and is presented in this form for the first time.

1.3 The building-blocks of vision

Human vision is a surprisingly complicated process. Even though it seems utterly effortless, repeatable and accurate, there is no current model that captures it all. Instead, advances have been made by breaking it down into smaller, more manageable parts. This reductionist approach, which has been so useful in many areas of science, leads us to the question 'what are the fundamental building-blocks of vision?' It might be thought that the obvious answer is light, photons or luminance. But the visual system is more sensitive to relative than absolute luminance, for two reasons: 1) a sophisticated system of gain control adapts the visual system to the local luminance level (Saito & Fukada, 1986; Shapley, 1986), and 2) most retinal ganglion cells have a receptive field that is excited by light in one region, yet also inhibited by light in an other region (De Monasterio et al., 1975; Kuffler, 1953). Although these regions are often not weighted equally, it is still the case that these retinal ganglion cells are more sensitive to luminance changes than absolute luminance *per se*. Thus a luminance border (an edge)

is a likely candidate for a building-block of vision. Models of edge detection will be reviewed below, and some of them will be shown to include the analysis of bars as well as edges. But the fundamental units of vision may not be best conceptualised as edges or bars in the spatial domain. An alternative approach considers the fundamental units of vision to be the image's Fourier components, implying that vision is best analysed in the spatial frequency domain. This approach has been extremely influential, so will be discussed first.

The spatial-frequency approach was popularised by the suggestion of Campbell & Robson (1968) that the visual signal is filtered by a set of bandpass spatial frequency filters, and that simple Fourier theory applies. They measured the contrast needed for detection of a sinusoidal grating at a function of spatial frequency and obtained an inverted U-shaped contrast sensitivity function (CSF). When the sinusoid was replaced with a same-contrast square-wave grating, sensitivity increased by a factor of about 1.3 across all spatial frequencies above 1 cycle per degree (cpd). This amount of increase was consistent with the proportionally higher amplitude of the square-wave's fundamental. Below 1cpd, as spatial frequency was reduced, sensitivity was expected to be progressively higher for the square-wave grating, because sensitivity to its 3rd harmonic was now much greater than to its fundamental (i.e. the 3rd harmonic was at the peak of the CSF and the fundamental was on its descending left limb). Square-wave sensitivity did progressively increase, but at a slower rate than expected from the known CSF, which suggested that performance could not be based on the properties of a single broad-band filter. In a second experiment, square-wave and saw-tooth gratings could not be distinguished from sinusoidal gratings until their contrast has been raised to a level at which the higher harmonic components reached their independent threshold. Taken together, these results suggested the existence of a number of separate spatial frequency channels, each tuned to a relatively narrow range of frequencies. Thus the CSF could be their envelope.

A particularly convincing demonstration of the existence of spatial frequency channels was provided by Blakemore & Campbell (1969) who showed that adaptation of a sinusoidal grating caused a localised dip in the CSF, centred on the adapting spatial frequency. The spatial frequency bandwidth (half width at half-height) of the dip was just over one octave. The adaptation effect was able to transfer across eyes (adapt one eye, test the other). This suggested a cortical locus, because the visual cortex is the first site where cells combine information from both eyes (Hubel & Wiesel, 1962, 1968). Other researchers have also studied adaptation and obtained results that support the idea of spatial frequency tuned channels (Blakemore & Nachmias, 1971; Georgeson & Harris, 1984).

Sub-threshold summation studies have also provided supporting evidence for spatial frequency channels (Graham & Nachmias, 1971; Kulikowski et al., 1973; Sachs et al., 1971). The method used here was to examine how the sensitivity to a target grating of one spatial frequency was affected by a second grating of a different spatial frequency or orientation that was of sub-threshold contrast. Thus, the sub-threshold grating was used as a probe to delimit the orientation and spatial frequency bandwidths of the filter that detected the target grating. The sub-threshold grating reduced target thresholds when its orientation was within about 3° of orientation, but not outside this range. A sub-threshold grating whose frequency was 3 times that of the target grating did not affect the latter's detection threshold. Subsequent work has suggested a wider bandwidth of $15 - 30°$ for orientation (Phillips & Wilson, 1984). This appears to contradict the narrow bandwidths obtained in the early 1970s, but the apparent discrepancy was resolved by incorporating probability summation across space into the modelling (Bergen et al., 1979). The stimuli used in the early 1970s typically contained several cycles

of grating, and so would be expected to stimulate a number of mechanisms. A group of neighbouring mechanisms that are aligned along a grating contour would all contribute to detection. Thus as a grating's orientation was changed, the part of a contour that is remote from the axis of rotation would quickly be shifted away from a mechanism's receptive field. So when the pooled output is considered, the whole group of mechanisms makes the observer very sensitive to changes in orientation. This accounts for the narrow orientation bandwidths observed in the early 1970s, even though the individual mechanisms had much broader bandwidth. By consideration of a number of mechanisms orthogonal to the grating's contours a similar argument would apply to spatial frequency bandwidths - probability summation across a group of broad bandwidth mechanisms would produce apparent narrowband behaviour.

A complementary approach to sub-threshold summation is that of contrast masking. Here, the non-target grating is high contrast, well-above threshold, and its effect is to reduce the observer's sensitivity to the target grating. Such masking studies have also supported the multiple channel hypothesis because the mask grating affected the target grating only when its spatial frequency and orientation were similar to the mask's (Campbell & Kulikowski, 1966; Legge & Foley, 1980; Wilson et al., 1983).

Some physiological support for spatial frequency channels was provided by Enroth-Cugell & Robson (1966), who found that cat retinal ganglion cells possessed spatial frequency band-pass properties. Furthermore, the receptive field of their central region varied in size between 0.5° and 8° Wiesel (1960), which is consistent with a preference for high and low spatial frequencies respectively. However, more convincing support has been provided by examination of cortical cell receptive fields. Hubel & Wiesel (1959, 1962) reported the existence of 2 types of cells: 'simple' cells and 'complex' cells. Simple cells had receptive fields that were divided into separate, side-by-side excitatory and inhibitory ('on' and 'off') areas, and the responses of these cells were orientation specific. They also appeared to perform linear summation across these areas, and their responses could be predicted from a knowledge of their receptive field. The responses of complex cells, however, could not be predicted from their receptive field. They responded to a suitably oriented stimulus that was flashed anywhere in their receptive field.

If the visual system performed a type of Fourier analysis, then the ability to distinguish between slightly different spatial frequencies would be optimised by using mechanisms that have a very narrow spatial frequency bandwidth. A narrow bandwidth was shown to be consistent with a large number of alternating positive and negative lobes in the cat and monkey simple cell receptive field by De Valois et al. (1985). They measured each cell's spatial frequency bandwidth and the number of cycles in a sinusoidal grating that produced its maximal response. A high negative correlation was observed between spatial-frequency bandwidth of simple cells and optimum number of cycles. Narrowly-tuned simple cells have been found in V1 of cat (Robson et al., 1988) and in V1 and V2 of monkey (von der Heydt et al., 1992). Mullikin et al. (1984) used light bars to probe cat simple cell receptive fields and found as many as 6 adjacent excitatory and inhibitory regions. These cells would be expected to have a bandwidth (half-width at half height) of 0.4 octaves.

This psychophysical and neurophysiological work has formed the basis of a multiple channel model of human vision, whereby the retinal image is filtered at each spatial location by filters that are tuned to a narrow band of spatial frequencies and orientations (Campbell & Robson, 1968; Graham & Nachmias, 1971; Robson, 1975; Watson & Robson, 1981; Wilson & Gelb, 1984). The model proposed by Wilson & Gelb (1984) was inspired from the observation in previous research that spatial-frequency discrimination was not a smooth function of frequency but rather exhibited shallow, alternating peaks

and troughs. This observation not only supported the existence of spatial frequency channels, but also provided a means of estimating their number from the number of peaks. On this basis, Wilson & Gelb employed 6 channels in their model, whose peak spatial frequencies ranges from 0.8 to 16cpd. Based on previous research, as spatial frequency increased (on a log axis) the channels' bandwidth slightly decreased. Each channel's sensitivity was based on the mean CSF of 3 observers. The filter outputs were then passed through a contrast nonlinearity that was accelerating at low contrast and compressive at high contrast and whose parameters differed for each channel. The final stage of the model performed Minkowski summation across each channel and neighbouring spatial locations ($f_i(x)$), which can be expressed as

$$S = \left(\sum_i \sum_x f_i(x)^Q \right)^{\frac{1}{Q}} \tag{1.1}$$

where Q was a free parameter that defined the type of summation. If $Q = 1$, summation was linear, if $Q = 2$, summation was quadratic, and if Q was very large, then the summation was dominated by the largest value of x, so that the net effect was to take the maximum. Wilson & Gelb (1984) showed that this model provided a reasonable fit to the discrimination data from 3 previous studies, with $Q = 2$.

However, evidence for the incompleteness of the channel approach has been provided from studies of plaids. A plaid is a sinusoidal grating superimposed on a second sinusoidal grating of different orientation. When two 1cpd components were at $\pm 45°$ from vertical, the plaid appeared to contain blurred horizontal and vertical edges (Georgeson, 1992, 1994, 1996; Meese & Freeman, 1995; Meese & Georgeson, 1991). Yet the most active channels would be at ± 45, and not horizontal and vertical, so a simple channel account should presumably predict only oblique edges.

Meese (1995) used vertical sinusoidal gratings that consisted of 2 different frequencies (f and nf) in a 2AFC phase discrimination task where f + nf was in one interval and f - nf was in the other interval. The range of n was 1/3 to 10. The contrast of f was fixed at 16% and the contrast of nf was varied by staircase to obtain the threshold for 75% correct performance. Meese also obtained the threshold for harmonic detection (f + nf *vs* f) and plotted the discrimination threshold relative to detection threshold as a function of n. The resulting function was U-shaped, with a plateau at about n = 1 to 5. The plateau was at about -6dB, which indicated that discrimination thresholds were about half of detection thresholds. This was consistent with the fact that the change in luminance was twice as large for phase discrimination as it was for compound detection.

In a second condition, the f grating was replaced by a plaid that consisted of 2 gratings at $\pm 45°$ from vertical. This plaid's components were of the same spatial frequencies as the grating it replaced. But its horizontal spatial repetition (SR) was now $f/\sqrt{2}$. Yet there was no energy at this frequency in the plaid's Fourier spectrum. When these data were plotted as a function of SF ratio (nf/f) the curve describing the plaid condition appeared to be a leftward shift of that describing the grating condition. However, when the results were plotted as a function of SR ratio (nf/SR), the functions superimposed. The conclusion reached was that spatial, not Fourier, attributes of the stimulus appeared to govern visual performance.

Akutsu & Legge (1995) performed an experiment that was specifically designed to test the spatial-frequency channel and local feature accounts of suprathreshold pattern discrimination. The stimulus was a f + 3f or f + 5f compound grating and the task was discrimination of a yoked contrast change (c_f, c_{3f}) *vs* $(c_f + \Delta c, c_{3f} - \Delta c)$. The compound gratings were of peaks-add or peaks-subtract phase. The resulting data were compared the predictions of the Wilson and Gelb channel model and three

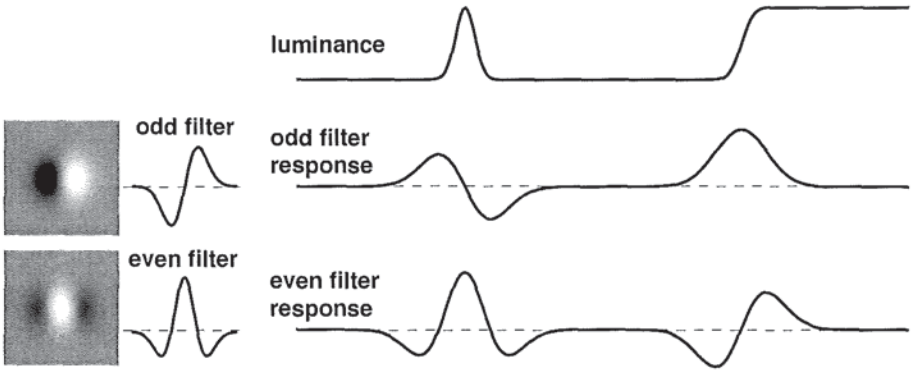

Figure 1.1: Receptive fields of odd-symmetric and even-symmetric filters (left, in plan view and cross-section), and their responses (right) to the blurred bar and edge shown in the top row. Note that each filter responds to both features.

local-feature models. The Wilson and Gelb model accounted for the peaks-add data, but failed to predict the peaks-subtract data. None of the local-feature models explained the results by themselves, but most of the data fell within an envelope composed of the lowest thresholds of the three feature-detector models. The conclusion reached was that suprathreshold pattern discrimination appeared to be mediated by mechanisms responsive to spatially localized features, and that more than one type of feature was used.

To summarise the above research, while the evidence is overwhelming for the existence of channels tuned to a narrow range of spatial frequencies and orientations, these channels may not govern visual perception directly. Rather than performing an analysis in the Fourier domain, the visual system may derive its information about spatial structure more from the features that are apparent in the spatial domain.

It is widely accepted that the retinal image is filtered by even- and odd-symmetric spatial filters at various scales, early in the visual pathway (Burr et al., 1989; Field & Nachmias, 1984; Hubel & Wiesel, 1968; Pollen & Ronner, 1981; Ringach, 2002) but exactly how the filter outputs are used to derive features is still an open question. Early psychophysical work proposed an edge-detector role for odd-symmetric filters and a bar-detector role for even-symmetric filters (Kulikowski & King-Smith, 1973; Shapley & Tolhurst, 1973). While these roles may be partially correct, they cannot define filter function completely, because the 'edge-detectors' also respond to bars, and the 'bar-detectors' respond to edges, as shown in figure 1.1. An alternative view is to consider these early mechanisms as filters that perform a computation of the spatial derivatives of the luminance profile. The next section describes how this process may be achieved.

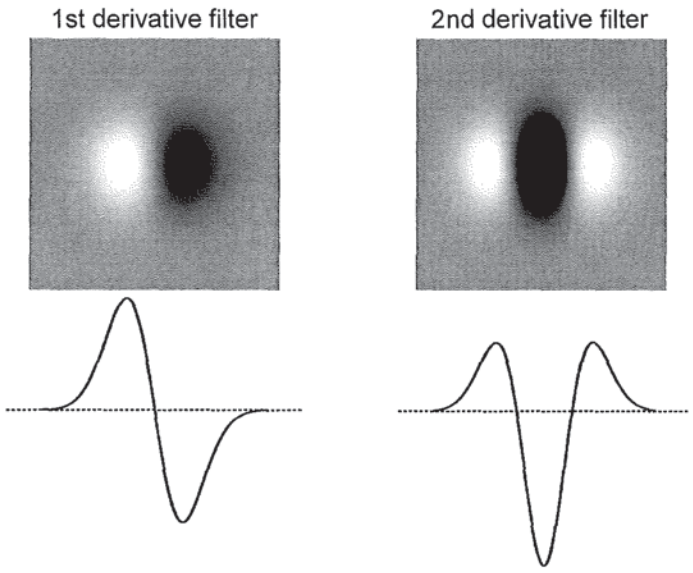

Figure 1.2: 1st and 2nd derivative filters and their horizontal profile.

1.4 Derivative filters and convolution

One general view of receptive fields in early vision is that they act as spatial derivative operators (ter Haar Romeny, 2003; Lindeberg, 1994; Marr & Hildreth, 1980; Watt & Morgan, 1985; Young & Lesperance, 2001; Young, 1985). Figure 1.2 (left) shows a 1st derivative filter with a horizontal section through its profile. It is the 1st derivative of a Gaussian. Suppose that the output of this filter is the simple sum of the responses of its 2 regions, each of which has equal weight. Since the 2 regions are of opposite sign, the filter output will be the difference in luminance between the 2 regions. If this filter is 'swept along' (convolved with) a 1D image, so that it makes this comparison between every point and its neighbour, then the output at every point is proportional to the rate of change of luminance, i.e. its 1st derivative, after a degree of smoothing that is determined by the scale (size) of the receptive field. A similar argument applies to obtain filters that compute the 2nd, 3rd or any higher derivative, but with different weights applied to different regions. Thus the 2nd derivative filter shown in 1.2 (right) has a central dark region that has twice the weight of each flanking light region.

1.5 The importance of edges

Edges are critical features in the visual scene. They signify not only object boundaries, but also other information such as surface markings, surface texture, reflectance changes or shadows. Their usefulness is demonstrated by the ease with which we can recognise items in a simple line drawing that represents edge locations, but where many other visual cues have been removed, such as colour, shading and depth. The importance of edges was recognised by probably the most influential vision scientist of the last century, David Marr, who proposed that the human visual system processes the signal from the eyes in a hierarchical sequence of stages (Marr, 1976; Marr & Hildreth, 1980; Marr, 1982). In this 'bottom-up' approach to visual processing, such low-level features are crucial, as they are the

foundations upon which the rest of vision is built. For example, the perceived lightness of a surface is affected by 'filling-in' from its bordering edges (Hung et al., 2007; Mumford et al., 1987). Edges are crucial for the segmentation of regions of different texture (Nothdurft, 1985) and for the perception of motion transparency (Zanker & Meso, 2008). They are considered essential in the perception of natural images (Geisler et al., 2001). Furthermore, they are required for emmetropisation during post-natal ocular development (Hess et al., 2006).

But how are edges extracted from the eye's neural output? This review will examine 5 approaches to this question. After a brief definition, each one will be described in two parts: the machine vision research that forms its basis, followed by psychophysical, computational and neurophysiological work in biological vision.

1.6 Difference of boxes

This approach compares the luminance at each point or region in the image with its immediate neighbour. A certain intensity difference indicates that an edge lies between the two points or regions. This method can also be conceptualised as convolution of the image with a Difference of Boxes (DOB) filter that consists of a rectangular box abutting an identical box of opposite sign.

1.6.1 Machine vision

An algorithm developed by Roberts (1965) performed this luminance comparison between adjacent pairs of pixels in the image, and applied a low threshold to deal with image noise. This stage delivered a set of candidate edge points. Next, for each point, four correlations were performed to find the direction of the short line (if any) that spanned the point. Four lines, each 5 pixels in length, and of slopes 0, 1, ∞ and -1 were correlated with the data around the point. The ratio of the best to the worst fit was taken and if above threshold, an edge was output, with the orientation of the best fit. These short edges were then combined into longer contours by looking for adjacent edges that were within $\pm 23°$ of each other, and further smoothing steps were applied.

Although restricted to simple block-like images with straight contours, this algorithm was successful enough to be used in later stages of processing to determine 3-D structure. It demonstrated the usefulness of applying a threshold, and obtaining differences in intensity between adjacent pixels. However, Rosenfeld (1970) noted that this method failed to locate a substantially noisy edge that was still visible to the human observer. So he proposed that the luminance comparison be made, not between individual pixels, but between the averages of two short, horizontal strips of pixels. This filtering was applied with strips of various lengths (scales). Small-scale filters localised edge position well, but produced many false-positives from the images noise. Large-scale filters localised edge position less well, but were more noise-resistant. So the product of the filter outputs at each point was used to obtain discrete locations for edges of any scale. This operation was also carried out vertically, and the maximum of the horizontal and vertical outputs was used to assign an edge of the appropriate orientation at each point. While this algorithm works for sharp edges embedded in noise, it does not specify the location of a large-scale (blurred) edge very well, as its output at a point near the centre of the edge is almost as high as the output at the exact mid-point of the edge. So subsequent work added a compensatory rule; the output for each point was compared to all other outputs within half

of the box function's width (Rosenfeld, 1970). Only the maximum output survived.

1.6.2 Biological vision

Recordings from the visual cortex of cat (Hubel & Wiesel, 1959, 1962) and monkey (Hubel & Wiesel, 1968) established the existence of elongated cells that have an excitatory area adjacent to an inhibitory area. This approximates the DOB profile, and therefore suggested that these cells could be edge detectors. The existence of separate excitatory and inhibitory areas had previously been observed in cat retinal ganglion cells by Kuffler (1953). But the areas here were arranged in a concentric manner. Kuffler observed 2 versions of this arrangement. Some cells were excited by a light spot in their central region, and inhibited by a spot in a surrounding annular region. Kuffler termed these 'on-centre' cells. Conversely, 'off-centre' cells were also observed, which were inhibitory in their central region, and excitatory in the surround. But since these receptive fields were isotropic, they appeared to be unlikely candidates for edge detectors.

1.7 1st derivative

The first derivative of the luminance profile forms peaks (and troughs) where the luminance gradient is most positive (or negative). For a symmetrical blurred edge, this peak will be at its mid-point. So, peaks in the first derivative locate dark-to-light edges and troughs locate light-to-dark edges.

1.7.1 Machine vision

Rosenfeld's box function, described above, produces a maximum output for a step edge that fits its shape precisely. However, the point was made by Macleod (1972) that many edges are not step edges, but have a luminance profile that is smooth (i.e. blurred edges). A filter that is discontinuous, such as the DOB, would provide poor discrimination of low-resolution edges from background noise.

Macleod suggested that edge processing required processing at both fine and coarse scales. Therefore, better filter shape would be one that is able to exclude fine scale information when a large scale filter is required. He suggested a Difference of offset Gaussian (DOOG), illustrated in figure 1.3, left (Macleod, 1970). This is an upright Gaussian added to an offset inverted Gaussian (each shown by dotted lines in figure 1.3, left). Although it was not described as a first derivative filter, its shape is clearly similar when the offset between the 2 Gaussians is small (figure 1.3, right). In fact, at a particular offset, the DOOG filter is mathematically the discrete form of the continuous Gaussian derivative filter (Young et al., 2001).

However, the notion that a smooth filter is optimal was disputed by Argyle (1972, 1971). He argued that a central discontinuity was essential when the image contained step edges, because the optimal filter would be one that matched the profile of the edge. Nevertheless, he maintained that if an edge was blurred, a smooth filter would be optimal. Thus it seems that an ideal edge detection mechanism would employ a number of filters that span the full range of edge blurs. 'Edge blur' is equivalently described as 'edge scale', because a blurred edge becomes visibly sharper when it is made smaller (e.g. by increasing the viewing distance). The use of filters of many scales is a core feature of the scale-space approach, which is discussed below (section 1.10).

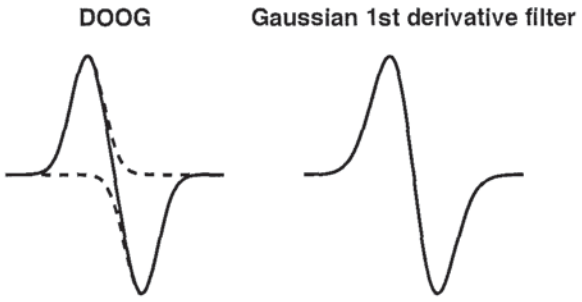

Figure 1.3: A Difference of offset Gaussians filter (left) and a Gaussian 1st derivative filter (right).

Fram & Deutsch (1975) used noisy edge images to perform a quantitative evaluation of the DOB and DOOG filters, in order to determine their performance limitations. Performance was measured with 2 criteria. One was an estimate of the signal-to-noise ratio in the filter's output, which was equally good for both DOB and DOOG filters. The other was a measure of the distribution of the filter's output edge signal over the length of the input edge. Here, the DOOG performed better. Fram & Deutsch also presented images that were reconstructed from the filter outputs, where the input images were a satellite photograph that contained blurred edges with various amounts of added 2D noise. The original images were more faithfully reproduced by the output of the DOOG filter than the DOB filter.

An influential peak finding model that employed 1st derivative filters was proposed by Canny (1986). He used numerical optimisation to obtain the optimal filter profile for a noisy step edge stimulus that balanced good localisation with good signal-to-noise ratio. This filter was similar to a 1st derivative of a Gaussian or a DOB and was shown to be optimal at any scale. The image was convolved with filters of various scales and the outputs were combined by 'feature synthesis'. This process began by locating edges using the smallest scale filter that had an acceptable signal-to-noise ratio. Then these edges were used to predict the output of the larger scale filters. The predictions were compared with the actual outputs of the larger filters. Then additional broad-scale edges were added at locations where the actual output was 'significantly' (not defined) larger than the prediction.

Canny noted that a single threshold applied to a low-gradient contour could cause streaking (blank portions caused by its fall below threshold, due to noise). But a threshold at some level was required to remove noise. So the method of 'hysteresis thresholding' was developed. It dictated that if an edge was above a high threshold it was output as an edge, along with all connected edges that lay above a low threshold. This reduced streaking, which would now occur only if the contour fluctuated above the high threshold and below the low threshold (which was 1/2 to 1/3 of the high threshold).

Canny also determined that combining outputs along the length of a contour improved detection, localisation and signal-to-noise ratio. This was accomplished by first locating an edge. Then the outputs of a small number of filters that line up with it were summed, thus forming an elongated filter. The variance of the small filters' response was high if the contour was curved or not aligned with the first edge, but low if a contour ran along the elongated filter. If the variance was high, the summed output was suppressed and the non-elongated filter output was used to identify a short edge.

Many machine vision algorithms require that outputs of filters of various scales are combined. This

is computationally relatively expensive. In an effort to reduce the computational burden, Bergholm (1987) suggested that a multiple-scale analysis is needed only at those regions of the image that are close to features, rather than across the whole image. His algorithm located edges by using first derivative of Gaussian filters, initially at the broadest scale. Edges that survived blurring at a large scale were assumed to relate to significant features. The filter scale was then progressively reduced, while the features were tracked, in order to accurately determine their spatial location.

However, Elder & Zucker (1998) pointed out two main problems with this 'edge focusing' method. Firstly, they take issue with the assumption that all significant edges would survive to high scale (when heavily blurred). Edges that are high contrast, isolated and in focus would generally survive, but many significant edges may occur at low contrast, close to neighbouring edges, or be out of focus. Secondly, they questioned the assumption that the smallest scale filter necessarily leads to the best localisation of an edge. If edges are widely spaced, they argue, localisation accuracy in fact increases with scale, because determination of edge location is less affected by noise. But this conclusion presupposes that no other method is used to deal with the noise, such as the application of a threshold. Furthermore, edges are often not widely spaced, such as those that make up texture elements.

Inspired by Bergholm's demonstration that it is not necessary to filter the whole image at all scales, Morrone et al. (1995) presented an algorithm that was also designed to reduce the computational burden. It used a filter that changed in size dynamically in order to keep signal-to-noise ratio constant across space. The filter size was small where the output was high (optimising resolution), and large where the output was low (optimising sensitivity). This automatic scale selection was implemented at each pixel by using the amplitude of its output at the previous pixel. At the first pixel, the scale was set to maximal.

This algorithm was applicable to any feature detection model that employed filters of more than one scale. Morrone *et al* applied it to a bank of Gaussian 1st derivative filters, to emulate a Canny detector. They presented modelling, using a noisy step edge, that showed that their algorithm performed better than filters of fixed scale, where the performance measure was the ratio of the model's output at the edge to its summed output across all marked edges. However, they do not make a direct comparison with a full Canny detector that employed feature synthesis, so it is not clear whether automatic scale selection would really perform better.

1.7.2 Biological Vision

The results of Hubel & Wiesel (1959, 1962, 1968), described in section 1.6.2, also apply here, as the different areas of the simple cell receptive field were categorised as 'on' or 'off', without intermediate levels of response. At this low level of resolution the DOB and DOOG are equivalent. Jacobson et al. (1993) examined macaque simple cells and found that some had a biphasic spatial profile, which were subsequently shown to provide a good approximation to 1st derivative Gaussian filters by Young & Lesperance (2001).

Furthermore, Heggelund et al. (1984) described cat simple cell receptive fields as formed from excitatory and inhibitory regions that were partially overlapping. The receptive fields of alert macaque V1 have been well-fitted by a model that employed a DOOG (Livingstone, 1998).

Psychophysical evidence has also suggested a DOOG filter shape. Subthreshold summation was used

by Kulikowski & King-Smith (1973) to probe the receptive field of edge and bar detectors. In the case of an edge, the threshold contrast for detection of a sharp edge and flanking light and dark regions was determined in the presence of a pair of flanking subthreshold thin bars at various separations. The bar was light on the light side of the edge and dark on the edge's dark side.

When the bars coincided with the edge, their luminance cancelled, so they had no effect on the edge detection threshold. As the bar separation increased, the edge's detection threshold was progressively reduced. This facilitation was maximal at about 3mins from centre and gradually fell to zero at about 10mins on each side of the edge. Thus the sensitivity profile as a function of distance had a shape that was similar to a DOOG (figure 1.3), where its zero-crossing indicates the edge's centre. The same profile was also generated from a second method, described as follows.

The subthreshold bars were replaced by a subthreshold sinusoidal grating and again the detection threshold contrast for the edge was obtained. This procedure was repeated for many different grating contrasts (all $\leq 0.5\%$). The threshold edge contrast was a linear function of grating contrast, and this function was then extrapolated to zero edge contrast in order to obtain the sensitivity of the putative edge detector to a grating of this spatial frequency. This whole procedure was then repeated for many other grating spatial frequencies, and the edge detector's CSF was plotted. This was then Fourier transformed to obtain the filter's shape, which was almost identical to that obtained by the previous method. The conclusion reached was that edges are detected by a mechanism of this shape. However, these results were subsequently explained without recourse to such filters; probability summation between narrow-band spatial frequency channels provided an adequate account (Graham, 1980, 1977).

1.8 2nd derivative

The location of a vertical blurred edge that is formed by a dark region on its left and a light region on its right can be determined by finding where the gradient (1st derivative) of its luminance profile forms a peak. But if the dark and light regions are reversed, then the edge forms a 1st derivative trough, not a peak. So if a single approach is required for edges of both polarities, it becomes necessary to locate 1st derivative extrema. These extrema points can be found by determining where the 2nd derivative is zero. This is the basis of the zero crossing edge detection method.

However, a limitation of this method is that it can report spurious edges that may not visible to humans, because not all zero-crossings are caused by the process described above. For a dark-to-light edge, the 1st derivative forms a peak that has a positive value, because the luminance profile is ascending (figure 1.4, left column). However, consider a luminance profile who's central portion descends slightly, abutted on each side by portions that descend rapidly (figure 1.4, 3rd column). The 1st derivative of this profile will also form a central peak, of negative value, that still provides a 2nd derivative zero-crossing. Equivalently, a luminance profile that ascends in a similar manner will form a positive-value trough in its 1st derivative and a zero-crossing in its 2nd derivative (figure 1.4, right column). Clark (1989) referred to these spurious edges as 'phantom' and the visible edges as 'authentic' and then addressed the issue of how to distinguish between them, using the 3rd derivative.

A positive 3rd derivative indicates that the 2nd derivative is ascending as it crosses zero and a negative 3rd derivative a decending zero-crossing. So an authentic edge has a 3rd derivative extremum whose sign is opposite to that of the 1st derivative extremum, and a phantom edge has a 3rd derivative

extremum whose sign is the same as that of the 1st derivative extremum, as shown in figure 1.4. Thus if the signs of the 1st and 3rd derivatives are multiplied together, then a positive product indicates a phantom edge and a negative product indicates a authentic edge. This method - based on standard calculus - was used by Clark to weed out the phantom edges that were generated by an isotropic Laplacian filter. This improved the resulting edge-maps of some periodic and natural images.

However, the notion that humans do not see phantom edges may be incorrect. Richter & Ullman (1986) demonstrated the existence of cells in the visual cortex of both cats and monkeys that respond to the phantom edge found in the center of a double step. A psychophysical study by Georgeson & Freeman (1997) showed that a phantom edge was perceived by 3 out of 4 observers, where the luminance profile was a plateau flanked by gradients of the same sign.

1.8.1 Machine Vision

A self-correcting algorithm that employed 2nd derivative zero-crossings was developed by Chen & Medioni (1989). They recognised that edges often occur close to one another, thus causing filter outputs to produce zero-crossings at distorted positions. They still applied 2nd derivative filters at the first stage of their model, but then corrected the edge positions by an iterative process that estimated the edge amplitudes and then combined them with the estimated edge positions to synthesize the signal. This modified signal was then filtered a second time close to the first estimate of edge positions, to produce zero-crossings at slightly amended positions. This whole process was repeated several times

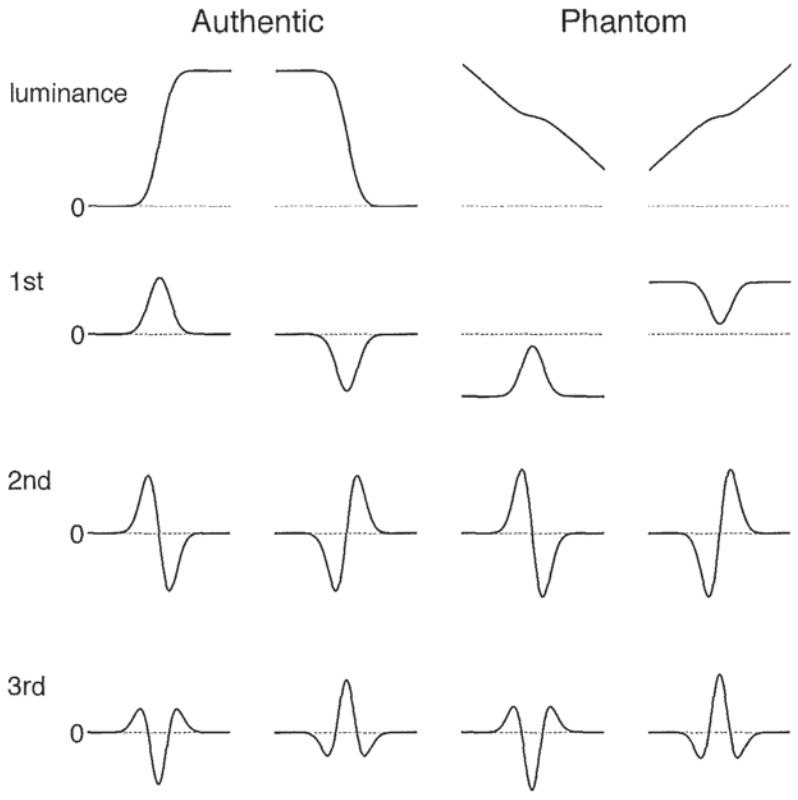

Figure 1.4: Authentic and phantom edges and their first three derivatives

until the edge positions stabilized. A low threshold was applied to remove the false edges caused by points of inflection in the filter output, which were due to low contrast edges. Chen & Medioni showed that this algorithm analysed and reconstructed a 2D image that looked similar to the original. The algorithm was applied vertically and horizontally and the results passed through an OR gate for each position.

Elder & Zucker (1998) suggested an edge detection algorithm that used the 1st and 2nd derivative in series. The algorithm first computed gradient estimates at octave intervals of scale, at each point, starting with the smallest scale at which the gradient estimate exceeded a 'critical value'. This was defined as the minimum response that could be considered to be statistically reliable, for a given level of sensor noise and filter scale. It was inversely proportional to the square of the filter scale for these 1st derivative filters. This stage of the model returned the 1st derivative of the luminance profile, so it would have been possible to then locate its peaks. But Elder & Zucker argued that although the use of the critical value meant that the 1st derivative would be reliably non-zero, the 1st derivative peaks may nevertheless be unreliable, due to noise. But a peak in the 1st derivative corresponds to a zero-crossing in the 2nd derivative. Thus, an edge could be located where the 2nd derivative reliably changes sign. So Elder & Zucker added a 2nd stage to their model, whereby the smallest reliable 2nd derivative filter was applied along the gradient, whose direction had been identified by the 1st derivative filter. Edges were located at 2nd derivative zero-crossings. The model also returned the edge blur, which was derived from the distance between the 2nd derivative extrema that flanked the zero-crossing, after correcting for the blur that had been caused by the filter itself.

Elder & Zucker showed that their algorithm correctly returned the amount of blur of an edge whose blur increased gradually along its length. It also coped very well with natural images that contained a wide range of edge blurs. However, May (2003) pointed out that one possible drawback is that it could only return a single scale of edge at each location. Thus if a shadow (blurred edge) fell across a region of fine detail (sharp edge), one edge would be suppressed. This limitation is absent in the N3+ scale-space model discussed below in section 1.10.2.

1.8.2 Biological vision

Kuffler's work on cat retinal ganglion cells was extended by Enroth-Cugell & Robson (1966, 1984), who examined the response of these cells to stationary, counterphase flickering or drifting sinusoidal gratings that were of various contrasts, spatial frequencies and phases (relative to the cells' receptive field (RF) centre). The responses of the cells fell into 2 general categories, and thus identified 2 cell types, termed X and Y. For flickering gratings, an on-centre X cell gave a maximal sustained response to a grating whose light bar fell across the cell's central excitatory region and whose flanking dark bars fell across the inhibitory surround (phase 0°). When the grating's phase was reversed (to 180°), the cell firing rate decreased a large amount while the grating was present, and returned to its resting level soon after grating offset.

However, when the grating was odd-symmetric about the RF centre (phase 90° or 270°), the X cells did not respond at all to the grating onset or offset. Thus the increase in luminance in one half of the receptive field had nullified the effect of an equivalent decrease in luminance in the other half of the receptive field. The conclusion reached was that X cells linearly summate across their RF area. Y cells, in contrast, responded to all phases of grating, with a brief increase and decrease of firing rate

at grating onset, offset or both. Y cells were clearly non-linear with respect to area summation.

Enroth-Cugell & Robson modelled X cell receptive fields as a difference of Gaussians (DOG), which is a positive Gaussian added to a wider negative Gaussian centred on the same point (figure 1.5, right). This filter shape is strikingly similar to a 2nd derivative filter (figure 1.5, left) and thus suggests a possible locus for 2nd derivative computation. A DOG model for cat X cells was also proposed by Rodieck & Stone (1965).

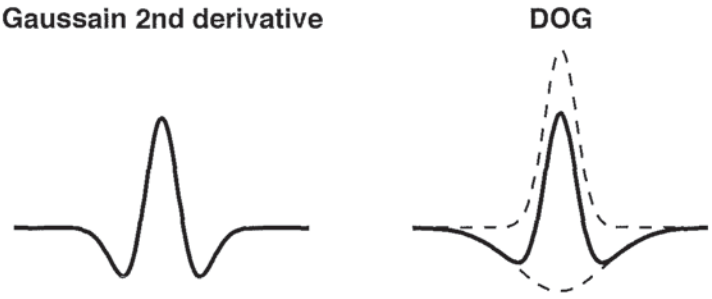

Figure 1.5: A Gaussian 2nd derivative and a DOG filter. The dashed lines in the right panel show the DOG's component Gaussians.

The DOG model has not only been applied to the responses of cat retinal ganglion cells. It has also been applied to primate retinal ganglion cells, although for primate, the division into different cell types has not directly followed the X and Y categorisation. De Monasterio & Gouras (1975) identified 3 types of rhesus monkey retinal ganglion cells: colour-opponent, broad-band, and non-concentric. The colour-opponent cells had the same antagonistic receptive field shape as the cat X cells, but the excitatory and inhibitory regions had peak sensitivities to different wavelengths of light. The broad-band cells were similar to the colour-opponent cells except that the regions' peak sensitivities were to the same wavelengths of light. The non-concentric cells did not have concentrically organized receptive fields. Some of these responded only briefly to stimulus onset, offset or both onset and offset, and had no spontaneous activity. But others had regular spontaneous activity and were responsive only to a moving stimulus. The colour-opponent and broad-band cells are now known as P and M cells respectively, because they project to the parvocellular and magnocellular layers of the LGN.

P cells are like X cells in that they have a sustained response to stimuli, and M cells, like Y cells, have a transient response. But the direct comparison between cat and primate breaks down when the linearity of area summation is considered. Virtually all P and most M cells show the linear area summation of X cells, and only a few M cells are non-linear like Y cells (Shapley & Perry, 1986). Nevertheless, the receptive fields of both M and P retinal ganglion cells have been fitted with a DOG model by Croner & Kaplan (1995). The DOGs were described by 4 parameters: the standard deviation and sensitivity of each of the 2 Gaussians. For foveal P cells, the relative weight of the inverted Gaussian was about 0.6 of the central upright Gaussian. This is less than the value of 1 that characterises a Gaussian derivative filter. Moreover, the standard deviation of the DOG's inverted Gaussian was larger than expected from a Gaussian derivative filter. Thus the DOG model for P-cells only partially resembled a 2nd derivative filter.

Moving from retina to cortex, the response profiles of simple cells have been mapped at relatively high resolution and found to approximate a Gabor function in cat (Jones & Palmer, 1987) and monkey (Ringach, 2002). A Gabor function is a sinusoidal 'carrier' grating enveloped by a Gaussian. When a Gabor contains about 1.5 cycles of its sinusoidal component, in cosine phase relative to its Gaussian envelope, its 1D profile closely approximates a DOG, as shown in figure 1.6. Cells of this profile were reported in both of the above papers.

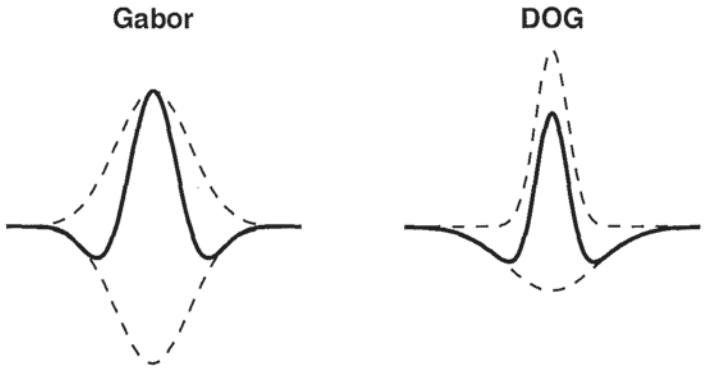

Figure 1.6: A 1D profile of a Gabor and a DOG filter. The dashed lines on the Gabor show its Gaussian envelope, and the dashed lines on the DOG show its component Gaussians.

1.8.3 Marr & Hildreth's Laplacian of Gaussian

A very influential model of feature detection was developed by Marr & Hildreth (1980). It was expressed within a broader theory of vision by Marr (1982), whereby the problem of interpreting the neural signal from the eye was considered to be a complex information-processing problem. This problem required an understanding at 3 different levels: computational, algorithmic, and implementational. The computational level specified the precise goal of the computation and the logic of the strategy by which it could be carried out. The algorithmic level specified how the computational problem could be solved, including the representations for the input and output. The implementational level specified how the computation could be achieved physically.

In the context of edge detection, the computational level required the detection of luminance changes in the image. The algorithmic level specified that these could be located from the zero-crossings of a Laplacian of Gaussian (LOG) filter, which is the 2nd derivative of a Gaussian, rotated about its peak so that it is 2 dimensional. LOG filters have the advantage of being circular rather than directional, thus eliminating the need to perform separate operations at each orientation. Furthermore they seemed biologically plausible because cat LGN (Dreher & Sanderson, 1973) and retinal ganglion cells (Rodieck & Stone, 1965) had been found that had a DOG profile, which is similar to a LOG (see figure 1.5). Threshold detection studies also pointed to DOG profiles (Wilson & Giese, 1977; Wilson & Bergen, 1979).

At the implementational level, Marr & Hildreth (1980) suggested that the LOG filter was implemented by cells that were driven by on-centre and off-centre retinal ganglion and LGN cells. The on-centre and off-centre cells would carry the signals that are transmitted to the positive and negative regions of the

LOG respectively. If an on-centre and an adjacent off-centre LOG were both firing, then a zero-crossing was assumed to be located between them. Moreover, if the outputs of two parallel rows of on-centre and off-centre LOGs were combined by a logical AND operation, then the resulting output would be positive when a luminance edge (of the correct orientation and polarity) was present. Marr & Hildreth proposed that simple cells in V1 were likely to perform this AND operation. Interestingly, Marr & Hildreth defined the amplitude of the edge as the slope of the 2nd derivative at the zero-crossing, which is a 3rd derivative operation. The role of the 3rd derivative will be discussed further below (section 1.10).

The LOG filters produced several zero-crossing maps, one for each scale of filter. A 'spatial coincidence assumption' was applied, which specified that if zero-crossings across two or more adjacent scales were coincident, then a candidate edge was assumed to be present. The final output of the algorithm was a 'raw primal sketch', which formed the sole input to the next stage of visual processing. This sketch consisted of early representations of edges (and bars and blobs) that were contained as 'primitives' (or building blocks). The sketch provided a richer description than a simple line drawing, however, as its primitives had the attributes of contrast and blur in addition to position, orientation and width. More importantly, it was a symbolic representation, which had transformed the retinal image into a form amenable to subsequent processing.

Although Marr & Hildreth's model was consistent with existing physiology and psychophysics, it has since been criticised on a number of counts. Robson (1983) argued that a LOG filter did not form a good approximation to the receptive field profile of retinal ganglion cells in cat, because their surround was too broad. However, Marr & Hildreth proposed that the LOG operation was probably performed in LGN, rather than retinal ganglion cells. If a filter computes the 2nd derivative, then its spatial frequency tuning curve will have a slope of 2 at low spatial frequencies, when plotted on logarithmic axes, because the response amplitude is proportional to the square of the input frequency (May, 2003). May pointed out that the slopes derived from retinal ganglion cells are about 1 or less (Enroth-Cugell & Robson, 1966), which appears to rule out the retina as as a locus for LOG filters. Furthermore, he added that slopes of about 2 have been reported for LGN X cells (Maffei & Fiorentini, 1973; Troy, 1983), which is consistent with Marr & Hildreth's model.

Marr & Hildreth's model assumed that zero-crossings were encoded at a stage that was not orientation-specific (in LGN), and that orientation specificity came later (in V1). However, psychophysical work by Georgeson (1992) suggested that zero-crossings were computed at a stage that was orientation-specific. First, he established that the perceived edges in a plaid composed of two or three sinusoidal gratings, corresponded to zero-crossings in its isotropic (non-oriented) 2nd derivative. Then he found that the perceived edge locations were distorted by prior exposure to a masking or adapting grating. Yet no distortion would be expected from isotropic filters, because they are not orientation-specific. Masking or adaptation would be expected merely to reduce an isotropic filter's overall response. Georgeson also found that the distortion was nulled by increasing the contrast of the masked/adapted component. He proposed a model that consisted of orientation-selective 2nd derivative filters that were linearly summed to form a circular filter, and then the zero crossings in the combined output were obtained. To conclude, although Georgeson's results were evidence against the order of stages in Marr & Hildreth's model, his model still incorporated isotropic filters as an essential component of the edge localisation process. These filters were 2nd derivative operators.

Marr & Hildreth suggested that V1 cells derive their orientation specificity from parallel rows of on-

and off-centre LGN afferents, and that a logical AND operation is carried out. Whilst the inputs from LGN do appear to be organised in this way (Ferster & Miller, 2000), evidence for the AND operation has been lacking. Schumer & Movshon (1984) interpreted the AND operation to mean that a simple cell would not fire unless a critical number of its afferents were firing, and that this predicted a discontinuous edge length threshold. However, when the edge length was gradually increased, the simple cell's response also increased in a smooth manner. Moreover, an increase in edge length could be compensated by a decrease in contrast, which was consistent with linear summation rather than a logical AND. Tolhurst & Dean (1987) also found little evidence of AND gates in cat simple cells.

Further problems with Marr & Hildreth's model have arisen as a result of its method of combining across scales. Firstly, it relies upon a spatial coincidence assumption, which states that if a zero-crossing exists across a range of scales at the same location and orientation, then a real edge can be assumed. Conversely, if the zero-crossing locations differ in location or orientation, then an edge is not assumed. While this strategy works fine for an isolated edge, it may not be so effective for a more complex image, because its zero-crossings often shift position as scale increases (Witkin, 1983).

The second undesirable consequence of Marr & Hildreth's method of combing across scale is partly due to the use of isotropic LOGs rather than orientation-specific 2nd derivative filters. Zero-crossings from a LOG are known to 'swing wide' of luminance corners (Torre & Poggio, 1986) by a distance that increases with increasing LOG scale. This causes problems for the spatial coincidence assumption, because the zero-crossings of different scale will never coincide. However, even if this problem is ignored, it might be thought that the property of 'swinging wide' would lead to the percept of a corner that is not in line with the rest of its component edges. Yet sharp-edged corners appear perfectly regular (figure 1.7, top left). However, this regular appearance may occur because the distortion caused by the smallest scale filter is imperceptibly small. But if the square is blurred, then this distortion may become measurable. A future experiment could measure the perceived distortion in a set of images that are like the ones shown in figure 1.7.

Although a LOG filter of any scale will encode the location of an isolated edge correctly, if the edge is replaced by a narrow bar, then 2 edges will be detected that are displaced outwards from their veridical location by a distance that increases with increasing LOG scale (figure 1.8). This is not just an example of Helmholtz's irradiation effect (Helmholtz, 1911/2000, p172-204; Mather & Morgan 1986), because the same distortion applies to dark bars. Marr & Hildreth suggested that in this circumstance, the larger scale zero-crossings be ignored, and the edge location be derived just from the smaller scale zero-crossings that do superimpose. However, May (2003) asserted that this strategy completely undermined Marr & Hildreth's multi-scale approach because it would make all but the smallest scale redundant.

Perhaps the most fundamental limitation of Marr & Hildreth's algorithm arises from the different number of zero-crossings at each scale. When presented with an image of a sculpture, large scale LOGs produced relatively few zero-crossings, and small scale LOGs produced many more zero-crossings (Marr & Hildreth, 1980). This difference in the number of zero-crossings means that the spatial coincidence assumption must inevitably fail to capture many small scale edges. Therefore its application loses information in the image.

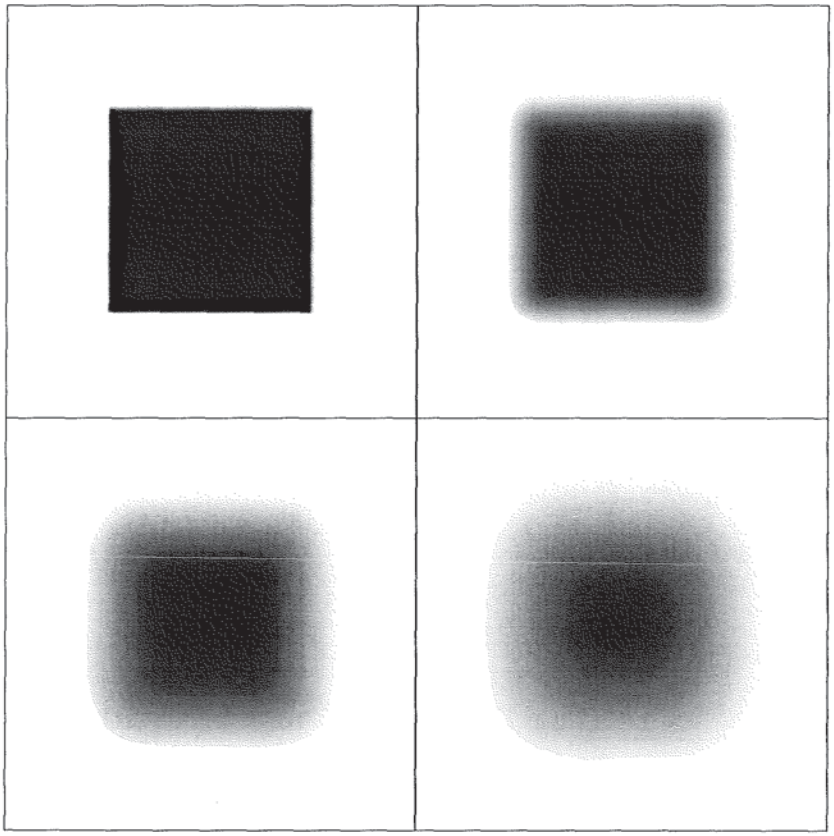

Figure 1.7: A sharp-edged square that is blurred with a Gaussian isotropic filter whose scale increases in successive panels from top left to bottom right.

1.8.4 MIRAGE

Watt & Morgan (1985) proposed a feature detection model, named MIRAGE, that employed 2nd derivative filters. It was different from the above models in that it did not assume that 2nd derivative zero-crossings identified potential feature locations directly. Rather, it used these zero-crossings to delimit regions of activity, which were then subject to further processing before feature locations were obtained. This model encompassed both edge and bar detection, and its stages are described as follows.

First, the image was filtered with a bank of 4 normalised Gaussian 2nd derivative filters defined as

$$f(x) = \left(\frac{x^2}{\sigma^2} - 1\right) exp\left(\frac{-x^2}{2\sigma^2}\right) \tag{1.2}$$

where σ is the standard deviation of the Gaussian. The expression in the left pair of brackets normalised the filters to a common height. The values of σ ranged from 0.35 to 2.8mins, which produced filters that were maximally sensitive to sinusoids of 27.3 to 3.4cpd. This range of spatial frequencies is much narrower that a typical CSF, but the predictions of the full model provided a good match to the whole CSF (Watt & Morgan, 1985), due to the later non-linear combination of filter outputs. Each filter's output was then separated into its positive and negative constituents, by half-wave rectification. Next, the constituents of each sign were separately summed across scale. This stage yielded 2 outputs; one positive (T+) and the other negative (T-). Each output consisted of one or more peaks (or troughs) that were separated by plateaux at zero response level.

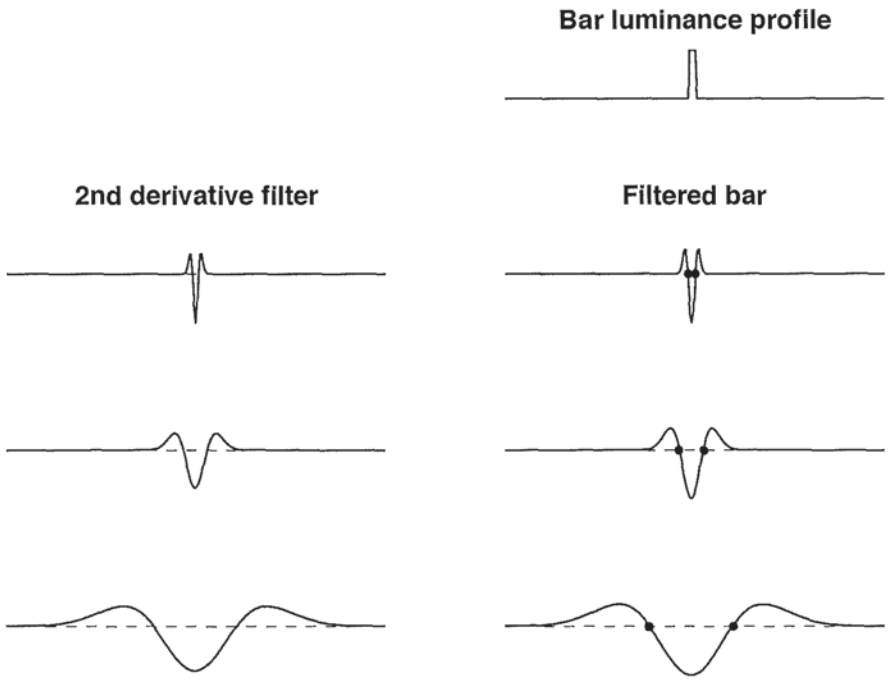

Figure 1.8: 2nd derivative filters of 3 different scales and their responses to a narrow bar. The black dots show zero-crossings.

If the original image was noisy, then the half-wave rectification stage resulted in outputs that had non-zero means. Then when these signals were summed across scale, the resulting signals (T+ and T-) had a d.c. shift, which would lead to very few regions of zero activity. This was undesirable, because such regions were crucial in the next stage of the model. To rectify this problem, the model suppressed the noise by subtracting the mean noise level from T+ and T-.

The T+ and T- signals were each separated into sequences of 2 types of primitive. A zero-bounded peak or trough was defined as a 'RESP primitive' (which indicated a response), and a plateau was a 'ZVSP primitive' (which indicated a region at zero value). Each RESP had the properties of mass (integral under the peak), centroid (point of zero 1st order moment) and standard deviation. In order to remove the RESPs due to noise that survived the above suppression, RESPs whose mass fell below a low threshold were vetoed.

Next, a set of parsing rules was applied as follows.

- A 'null' corresponded to a ZVSP, which indicated a luminance plateau.
- the boundary of an edge corresponded to a RESP with a ZVSP on one side only. Thus an edge between 2 luminance plateaux would correspond to the primitive sequence of ZVSP, RESP+, RESP-, ZVSP.
- A bar corresponded to a RESP with a ZVSP on both sides, or a RESP on both sides.

The polarity of each edge and bar response depended on the RESP's sign. The location, contrast and

width of each bar were determined by the RESP's centroid, mass, and standard deviation respectively. An edge was assumed to be located at the mid-point between its 2 RESPs' centroids, and edge blur was the distance between these 2 centroids.

MIRAGE has received substantial experimental support from its authors (Watt & Morgan, 1983, 1984, 1985; Watt, 1988). In one experiment, they measured blur discrimination thresholds, using a reference edge whose blur and edge profile was varied (Watt & Morgan, 1983). As blur increased from zero, the blur discrimination threshold first decreased and then increased in a power law manner, with exponent 1.5. This pattern was similar to a typical 'dipper' function seen in contrast discrimination experiments (Legge & Foley, 1980), although the exponent in those experiments is typically in the range 0.5 to 1. Watt & Morgan's dipper functions for the 3 different edge profiles did not superimpose when plotted as blur threshold *vs* reference edge blur, but did approximately superimpose when replotted as blur threshold *vs* separation between peak and trough in the 2nd derivative. This supported a 2nd derivative metric, and MIRAGE's computation of edge blur from the distance between RESP+ and RESP-.

In a 2nd experiment, Watt & Morgan (1983) generated a set of edges that ranged between a Mach ramp (a luminance ramp flanked by two plateaux) and a smooth Gaussian edge. Thus the stimuli were Mach ramps with 'corners' that had been smoothed to various degrees. The threshold amount of smoothing was obtained that allowed discrimination from a pure Gaussian edge, for a given edge width. Inspection of the luminance profiles of the threshold edges for 4 edge widths and 2 observers did not reveal the cue that allowed discrimination from a pure Gaussian edge. But the 2nd derivatives of these profiles all had a central narrow region of zero activity, that would not be present in the 2nd derivative of a pure Gaussian edge. This supported MIRAGE's use of zero-valued stationary points (ZVSP) in the 2nd derivative of the luminance profile.

MIRAGE's measure of contrast was obtained from the mass (integral under the peak). Thus contrast discrimination amounted to a comparison of 2 masses. The noise suppression stage in the model reduced the mass by a small amount, which would have greater impact on small RESPs than large RESPs. Therefore, predicted contrast discrimination performance would be selectively impaired at low contrasts only. So as contrast increased from zero, a plot of threshold *vs* contrast would have an initial negative gradient that became a plateau at moderate and high contrasts. When this was combined with Weber's law performance (threshold directly proportional to contrast), the resulting pattern of thresholds was a typical dipper shape. Watt & Morgan (1985) showed that the predicted thresholds provided a reasonably good fit to the contrast discrimination data of Legge & Kersten (1983).

A distinctive feature of MIRAGE is that it combines information across scale before primitives are obtained. This suggests that there may be no access to the information from the individual scales; only their summed output may be available. Watt & Morgan (1984) tested this suggestion psychopysically by studying the effect of a high frequency grating mask on vernier acuity performance for a pair of Gaussian blurred edges, positioned one above the other. The mask was applied to one edge only and was a high frequency grating enveloped in one of two ways. Either it was enveloped by a Gaussian, so that the resulting mask was superimposed on the edge, or it was enveloped by its complement, so that only the flanking regions of the edge were masked. Watt & Morgan described the response of three MIRAGE filters (σ =0.5, 1 or 2mins) to the 2 types of masked edges. The 2 smaller scale filters responded only to the high frequency mask, and therefore would be uninformative in the vernier task. However, the σ =2mins filter responded only to the edge, and its output was identical for the 2 types of mask, except for a small phase shift. Therefore, if the observer had access to the individual

scales, performance should be similar for both types of mask. But if the observer had access only to the combined signal, like the MIRAGE output, then performance was expected to be different for each mask type. The results were different for each mask type, which supported an inflexible filter combination process.

The evidence for MIRAGE has not all been supportive. Hesse & Georgeson (2005) reported that MIRAGE provided a relatively poor fit to the perceived location of edges and bars in their 1D images. They used a feature-marking method (Georgeson & Freeman, 1997) with a set of images whose amplitude spectrum was that of a partially blurred square-wave, and whose Fourier components all had a common phase. In different conditions, the phase was shifted in 45° steps. When the phase was zero, the image looked similar to a blurred triangle-wave, and when the phase was 90°, the image looked like a blurred edge dividing a light and a dark region. Observers marked a central edge only when the phase was 90°. At all other phases tested, observers marked a central bar flanked by a pair of opposite-polarity edges. The MIRAGE predictions matched the number, type and polarity of features seen by observers. However, the predicted edge locations were too central; they always lay within about ±1.7mins, but the marked edges extended to over ±4mins. The bar locations were generally correctly predicted. The success with bars and the failure with edges may be a property of MIRAGE's use of only even-symmetric filters. Watt (1988) stated that MIRAGE correctly predicted the presence of Mach bands on a Mach ramp (a luminance ramp flanked by 2 plateaux), and the appearance of the Chevreul illusion (figure 1.9). However, its behaviour with these 2 illusions has been challenged (Fiorentini et al., 1990; Kingdom & Moulden, 1992; Pessoa, 1996; Ross et al., 1989) and this will be discussed further in chapter 4.

1.9 Local energy model

All the above approaches rely on derivatives of the luminance profile, and they each use filters of only one shape (albeit of various scales). In contrast to this, the local energy model does not use derivatives. It uses two types of filter: even-symmetric and odd-symmetric. A feature is located at the peak of the energy profile (E), which is defined by

$$E = \sqrt{e^2 + o^2} \tag{1.3}$$

where e is the output of an even-symmetric filter, and o is the output of an odd-symmetric filter. The relative magnitudes of e and o dictate whether the feature seen is an edge, a bar, or some combination of both. A high value of E is caused by a high degree of phase congruency of the Fourier components of the image. For example, all of a bar's Fourier components coincide in cosine phase at its midpoint, but only a few components coincide at other locations.

1.9.1 Machine Vision

Morrone & Owens (1987) noted that all of the Fourier components of an image are in sine phase at an edge, and in cosine phase at a bar or line. So the location of a feature could be obtained by using these points of 'phase congruence'. One method of obtaining an output that is maximal at these points is to add the square of the image Fourier components to the square of their Hilbert transform (which shifts phase by 90° but leaves the amplitude spectrum unaltered). But since calculation of the Hilbert transform required a great deal of computation, Morrone & Owens suggested a more efficient

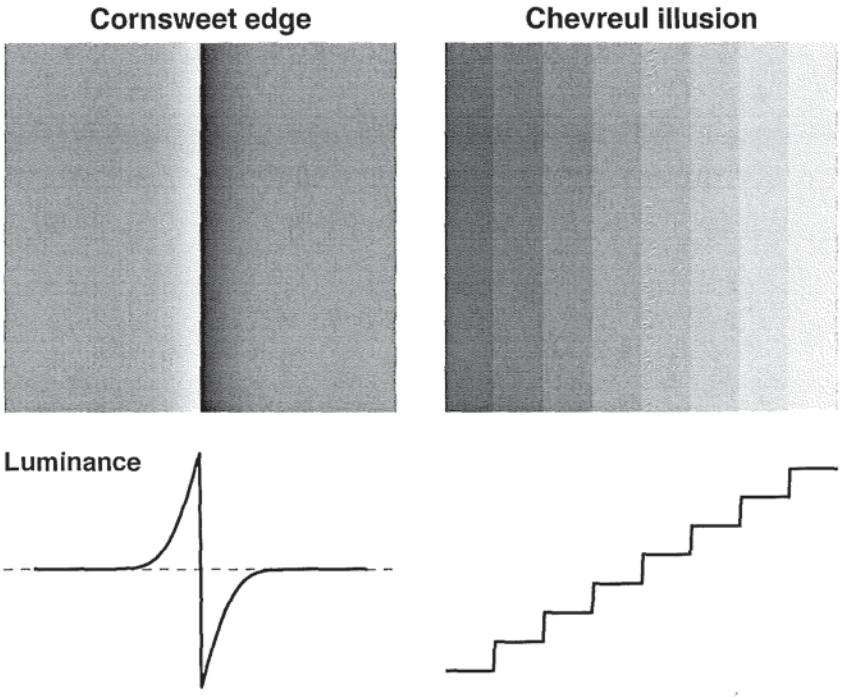

Figure 1.9: Image and luminance profile of a Cornsweet edge and the Chevreul illusion.

algorithm. The image was filtered simultaneously with two filter types. One was odd-symmetric, produced the same output as the Hilbert transform, and was maximally responsive to edges. Its profile was similar to that of a Cornsweet edge (Cornsweet, 1970), which is illustrated in figure 1.9 (left). The other was even-symmetric, and was maximally responsive to bars. Its profile was similar to a Gaussian 2nd derivative filter, but with very pointed peaks and troughs. The shapes of the filters were constrained by the requirements that their amplitude spectra should be equal for all frequencies, and that the phase spectra should remain at 0° for one filter and 90° for the other. Furthermore, to prevent aliasing, the spectra should fall smoothly to zero just before the nyquist limit. Several sets of filters met these constraints, and all produced clear energy peaks at edges, bars and the locations of Mach bands in a Mach ramp waveform; all with no false positives. They correctly identified each step on a staircase-shaped waveform, but did not produce any other energy peaks that might account for the Chevreul illusion's appearance.

The Chevreul illusion is the effect of an illusory sloped appearance of the steps of a luminance staircase, whose luminance profile consists of a series of steps that are each completely flat. It is illustrated in figure 1.9 (right). The reader may observe that each edge appears to be abutted by a faint light bar on its right side and a faint dark bar on its left side. These bars may cause each step to appear curved or of negative gradient. The bars are entirely illusory, as indicated by the flat luminance profile of each step, which is illustrated at the bottom of figure 1.9 (right).

Morrone & Owens were specifically trying to emulate human vision, so it is surprising that they did not attach any weight to a failing of this algorithm; it predicted no features at all for a sinusoidal grating, because energy (E) is constant across space. Perhaps they regarded a grating as a special case that was too different from most natural images to warrant further investigation. They extended their

algorithm to 2D by filtering horizontally and vertically, and applied a low energy threshold to deal with noisy images. This version provided a realistic 'line drawing' of edge locations of a coffee mug in diffuse lighting. This was achieved by using only 2 orientations and 1 scale.

Morrone & Owens' algorithm was criticised by Kovesi (2000) on the basis that its measure of phase congruency was a quantity that depended on image contrast. He proposed a more sensitive phase congruency model that was contrast independent. The derivation of congruency was explained with reference to the complex plane of the Fourier transform. Here, each axis denotes the response magnitude of each type of filter. So when the filters' responses are plotted as a vector, its length denotes the energy (E) and its direction denotes its phase. This is illustrated as vector **a** in figure 1.10. This vector represents the vector sum of all of its Fourier components. Thus if the components were of several different phases, then their individual vectors traversed a long distance when summed (vector **b**). But if the components were of very similar phases, then their individual vectors traversed a distance that was little more than the direct route from the origin (vector **c**). Koveski defined phase congruency (PC) as E divided by the sum of the lengths of the Fourier components. When all components were exactly in phase $PC = 1$, and PC tended to zero as phase became less congruent.

Koveski's algorithm was very sensitive to noise. A step edge of unit height with added noise of amplitude close to 0.1 resulted in noise PC peaks that were of equal magnitude as the edge PC peak. As Koveski remarked 'this is the price one pays for using a normalised measure such as phase congruency'. But he then proposed a correction whereby the level of E that arose from the noise was subtracted from the original value of E, which effectively imposed a low energy threshold. When the algorithm was applied to a number of natural images, its output was a realistic line drawing of the original image. A Canny detector performed less tidily on the same images; it produced twin responses on features that were not edges, where Koveski's algorithm produced just one response. Although Koveski presented

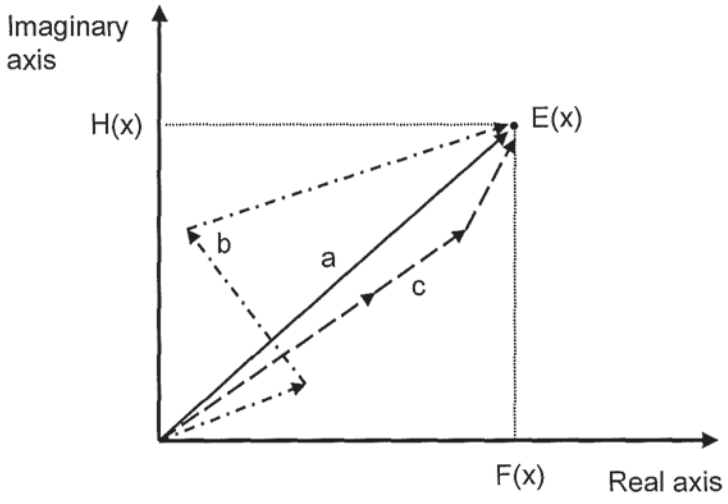

Figure 1.10: Complex plane representation of a signal's Fourier components. Vector **a** shows the signal's energy ($E(x)$). Vectors **b** and **c** show signals that have low and high phase congruency respectively. After Kovesi (2000).

this comparison in support of his algorithm's superiority over Canny's, it more properly reflects the fact that the Canny detector was designed purely to detect edges, so would predict 2 edges for each bar.

1.9.2 Biological Vision

The local energy model employs quadrature pairs of filters of even- and odd-symmetric filters. Early neurophysiological work in cat (Hubel & Wiesel, 1962) and monkey (Hubel & Wiesel, 1977) established that V1 simple cells had either even- or odd-symmetry, and this finding has been replicated (Maffei et al., 1979). The quadrature pair arrangement was supported by Pollen & Ronner (1981), who recorded the responses of adjacent pairs of simple cells to drifting sinusoidal gratings. Both cells of a pair were tuned to the same orientation and spatial frequency but their responses differed in phase by approximately 90°.

The local energy model requires that the visual system is sensitive to phase. Evidence for this was produced by Maffei & Fiorentini (1973). A cat striate cortex simple cell was exposed to a contrast-reversing sinusoidal grating. A shift in grating phase caused a corresponding phase shift in the cell's response.

A sinusoidal grating has constant energy. This can be understood mathematically, by the fact that substitution of $e = cos$ and $o = sin$ into the equation 1.3 results in

$$E = \sqrt{cos^2 + sin^2} = 1 \tag{1.4}$$

Therefore a cell that provides a constant output to a sinusoidal grating (above its spontaneous discharge rate) may be computing energy. This unmodulated, non-linear response to a grating was observed in complex cells of cat and monkey (Movshon et al., 1978; Skottun et al., 1991), and has been generally accepted as an effective means of allocating cells into the 'simple' and 'complex' categories, because the response of a simple cell follows the luminance variation within the grating in a linear manner. The unmodulated response of complex cells inspired Adelson & Bergen (1985) to propose a model of motion perception that filtered the image with quadrature pairs of filters, computed energy, and then employed an opponent stage. This model was subsequently extended, by the inclusion of mutual inhibition between cells (Heeger, 1992), which made the responses contrast-invariant.

The feature detection local energy model relies upon the existence of mechanisms that are sensitive to the phase of the Fourier components of an image. Psychophysical support for phase sensitivity was provided by Burr (1980), who reported that observers were able to discriminate the phase relationship between the two sinusoidal components of a compound grating, where the components had spatial frequencies f and 3f. The spatial frequency of f was selected from 1 to 30cpd. Although discrimination thresholds were high (a 20° phase shift was undetectable), the key finding was that the thresholds were approximately constant when expressed as a function of the phase of the 3f component, across all spatial frequencies tested. But when they were expressed as a function of retinal displacement, the thresholds varied by a factor of 16. The conclusion reached was that discrimination was mediated by mechanisms sensitive to phase, rather than to the retinal position of local features. However, an alternative account was that observers were simply using the variation in contrast between two selected points in the image (Badcock, 1984a).

Morrone & Burr (1988) also provided psychophysical evidence in support of the local energy model. They used 1D images that were constructed from the amplitude spectrum of a blurred square-wave, with all Fourier components at a common phase at the image centre. The phase at image centre was then varied in different conditions. When the phase was zero, the central feature looked like an dark-to-light edge, and when the phase was 90°, the feature looked like a light bar. When the phase was 45°, the feature looked like a combination of edge and bar. Morrone & Burr instructed observers to indicate the position and identity (bar or edge) of all features. They applied a version of the local energy model that employed several scales of filters; a feature was predicted at an energy peak at any scale[1].

The local energy model correctly predicted the position, type and polarity of the marked features. When the phase was 45°, observers reported that the central feature appeared to flip between edge and bar in a manner that had been previously described as 'monocular rivalry' (Atkinson & Campbell, 1974). For this image, the response of local energy model's even- and odd-symmetric channels was equal, so the feature was not identified as a pure edge or bar, but rather as a composite of both. Thus the local energy model was consistent with the observed perceptual instability. More importantly, it predicted that the feature was always located at the image centre, whereas models based on luminance or 1st derivative peaks predicted slightly displaced locations that did not accord with the observers' responses.

For this image set, the local energy model was shown to defeat 2 competing models. When Marr & Hildreth's LOG filter was applied to the images, the lack of coincidence of zero-crossings meant that no features (edges) were marked in the phase 90° and 45° images, even though observers reported a bar and edge-bar composite respectively. The effect of filtering with Watt & Morgan's MIRAGE was similar to full-wave rectification, which meant that it predicted a series of wide bars that did not match the observers' reports. Morrone & Burr concluded that derivative-based models were defeated by the local energy model.

However, Georgeson & Freeman (1997) argued that this conclusion was premature. They proposed that Morrone & Burr's use of unlimited duration stimulus presentation meant that eye movements would be likely to cause negative after-images that interacted with their image and altered its appearance. They modelled the effect of a 4mins lateral eye movement on Morrone & Burr's phase 45° waveform, and demonstrated that the addition of the pre-movement after-image to the post-movement stimulus would result in a combined waveform that had an edge-like or bar-like profile at its centre, depending on the direction of the eye-movement. This meant that a derivative-based approach could now provide a viable alternative to the local energy model in the interpretation of Morrone & Burr's results.

Georgeson & Freeman used the same feature-marking method as Morrone & Burr, with compound (f + 3f) gratings of moderate contrast and spatial frequency that had several phase relations between f and 3f. Stimuli were flashed repeatedly (duration 216ms, inter-stimulus interval 1067ms) to prevent the after-image confound described above. Observers marked 6 edges and 6 bars in most conditions, at locations that were predicted by peaks in the 1st derivative (for edges) or 2nd derivative (for bars). In stark contrast, the local energy model predicted only 2 features, not 12. The conclusion reached was that human vision does not use energy peaks to select feature locations.

However, the f + 3f compound grating used by Georgeson & Freeman may have been too spectrally

[1]This rule for combination across scale has been replaced by probability summation in a later version of the local energy model (Ross et al., 1989).

sparse to have been a fair test of the local energy model. Burr & Morrone (1994) stated that the local energy model was particularly successful with broadband images, because they generate strong energy peaks. But the the f + 3f gratings contained only 2 spatial frequencies, and were thus not broadband at all. Therefore, Hesse & Georgeson (2005) performed a feature-marking experiment with spectrally broadband stimuli that were the same as those used previously to illustrate the local energy model by its authors (Morrone & Burr, 1988). The aim was to provide a more thorough test of the derivative and local energy models of feature detection.

Hesse & Georgeson presented Morrone & Burr's stimuli (described above) with a flashing presentation to prevent the after-image confound (also described above). Each stimulus was selected from a range of scales, contrasts and phase shifts of all Fourier components. Observers marked the polarity and position of all perceived edges and bars. As stated in section 1.8.4, observers marked a central edge only when the phase was 90°, and at all other phases, observers marked a central bar flanked by a pair of opposite-polarity edges. When the position of each feature was plotted as a function of phase, each feature had a clear near-linear trajectory; its location was progressively shifted in tandem with phase shifts, by up to ±20arcmins. These shifts, and the polarity and type of both edges and bars, were accurately predicted by a simple model that used extrema in a slightly smoothed version of the luminance profile to derive bar locations and extrema of its 1st derivative to derive edge locations. The local energy model incorrectly predicted a single feature that was always at image centre, irrespective of stimulus phase.

In a 2nd experiment, Hesse & Georgeson examined a more global aspect of the perception of these stimuli: their apparent position. Observers were presented with 3 vertically-arranged elements, each drawn from the same set of profiles used previously. The outer pair were both of phase 0° or 180°, and the central element was phase shifted from the outer pair by ±45°. The task was to indicate whether a salient feature roughly in the middle was to the right or left of the outer pair. The horizontal position of the central element's profile was randomised between trials. When the energy peaks of the 3 elements were in alignment, the apparent position of the central element was shifted laterally by an amount that was consistent with the use of 1st derivative peaks as a cue. However, the local energy model predicted no shift at all. Taken together with the results of the 1st experiment, this appears to be damning evidence against the local energy model.

However, Hesse & Georgeson's model, as they admit, would not predict Mach bands. Yet Ross et al. (1989) showed that the local energy model correctly predicted the threshold contrast required for seeing Mach bands on 1D, trapeziodal-wave images that had been low-pass filtered to various degrees. The local energy model also predicted the observed slight threshold change when the lower Fourier harmonics were removed from the trapezoidal-wave. It also correctly predicted the threshold ramp width required for Mach band perception. In contrast, a derivative-based model (MIRAGE) incorrectly predicted a threshold ramp width that was 6 times wider.

The local energy model was supported by Hess & Holliday (1992) and McGraw et al. (1999), who examined the perceived location and motion of Gabors. Observers were able to accurately align 2 Gabors, even when their carrier differed in spatial frequency or orientation. This implied that observers did not employ cues derived from the carrier's local luminance or its derivatives; the envelope itself appeared to be the salient cue. The location of the envelope's peak would be returned accurately by the local energy model, whereas derivative models would predict locations based only on the carrier. While the local energy model performs well with the perception of contrast (Hess & Holliday, 1992;

McGraw et al., 1999), motion (Adelson & Bergen, 1985), texture segmentation (Bergen & Adelson, 1988; Bergen & Landy, 1991; Bovik et al., 1990; Malik & Perona, 1990), and binocular disparity (Fleet et al., 1996; Hibbard, 2008; Ohzawa et al., 1990), it performs relatively poorly with 1st order feature detection.

1.10 Scale-space

This approach provides an elegant solution to the problem of how to combine image representations of different scales. The solution is to construct a scale-space representation, which is a domain that is constructed (for a 1D image) as follows. The image is filtered with a derivative filters at a range of scales. Next, the filtered luminance profile from the largest filter is plotted in the z direction (out of the page) as a function of spatial position. Then, the luminance profile from every other filter is plotted above that, in the y direction, in order of increasing scale. The resulting 3D plot has the (x,y,z) axes of spatial position, scale and response amplitude. In order to display the response amplitude, scale-space maps sometimes show contours of equal response value, like height in a conventional map. Subject to further constraints, a peak in this map may denote a feature.

1.10.1 Machine Vision

Linear scale-space was first proposed by Taizo Iijima in 1959 (Weickert et al., 1999). But as he published his work in Japanese, it did not receive wide recognition. The first Western proponent of scale-space filtering was probably Witkin (1983). He described the construction of scale-space (as outlined above) and proposed that a Gaussian derivative should be the filter of choice, because it is 'well-behaved'. This term means that as scale increases, no new features can appear; in other words, the act of filtering does not generate spurious features. Witkin stated that the Gaussian derivative was the *only* filter that fulfilled this criterion, but he only cited an unpublished paper in support. The proofs for 1D and 2D images have since been published (Babaud et al., 1986; Koenderink, 1984; Yuille & Poggio, 1986). Witkin examined the behaviour of zero-crossings in the scale-space map, rather than the peaks described above. But since this approach applies to Gaussian derivative of any order, a zero-crossing in the nth derivative corresponds exactly to a peak in the derivative of order $n-1$.

Korn (1988) presented an algorithm that tracked edges through scale-space. He used Gaussian 1st derivative filters of various scales, that were normalised by multiplying their response by $\sigma\sqrt{2\pi}$, where σ is the scale of the Gaussian. This normalisation meant that the filters' responses to a step edge would have a common amplitude. First, the algorithm obtained peaks in the 1st derivative with filters of only the smallest scale. Then these edges were tracked through scale-space, and the peak of activity at each scale was used to form a function that mapped filter scale onto filter response. The first peak in this final function, moving from small to large scales, denoted the scale of the edge. Korn showed that the scale returned by the algorithm did correspond well to the scale of the input edge. However, May (2003) remarked that there is a flaw in this method. When the input is an isolated Gaussian edge the final function would not form a peak, but rather would increase monotonically with scale (Zhang & Bergholm, 1997). A peak would only occur when the edge was close to its neighbour, and the scale at which it occurred would be affected by the distance between edges. Korn had applied his algorithm only to complex scenes, so may not have spotted this flaw. May also pointed out that since Korn's

algorithm tracks from small to large scales, it would be very sensitive to noise.

Zhang & Bergholm (1997) suggested an algorithm that was very similar to that of Korn (1988). But they tracked edges through scale-space from large to small scales, rather than small to large, in order to reduce susceptibility to noise. They categorised edges as a 'step edge' or a 'double edge'. A step edge was a monotonic increase in luminance, and therefore encompassed any degree of blur. A double edge was defined as 2 neighbouring luminance changes. The algorithm assumed that a discrete peak in scale space denoted a double edge, and monotonically increasing function denoted a step edge. The double edge category was further subdivided into 'bar edge' and 'line edge'. A bar edge was flanked by regions of approximately equal luminance, and the regions differed in luminance for the line edge. The algorithm distinguished between these 2 edge types on the basis that a bar edge formed 2 peaks in scale-space that moved apart symmetrically as scale increased, whereas for a line edge, one peak did not survive to such large scales as the other.

Once the edge types had been determined, the algorithm then estimated their attributes. A step edge and line edge had the attributes of amplitude and blur. A bar edge had blur and width attributes, which were used to determine its amplitude. The attributes were obtained by matching the scale-space trajectory with various templates to determine the best match. Each template was labelled with the appropriate attribute values. Zang & Bergholm showed that when their algorithm was applied to a number of complex images, it produced line drawings that were less noisy that those produced by some previous algorithms (e.g. Mallat & Zhong, 1992).

A substantial contribution to the scale-space approach has been made by Lindeberg (1990, 1993, 1994, 1996, 1998). He demonstrated that the appropriate well-behaved filter for a continuous image was a Gaussian (or its derivative), and for a discrete image was a modified Bessel function of integer order. The Bessel function would be appropriate when the image is pixelated, which is always the case for digital images that are presented on monitors. However, the retinal image is subject to optical (Campbell & Gubisch, 1966; Williams et al., 1994) and neural (Levi & Klein, 1990a,b) blur, which effectively transforms the discrete image onto a continuous form, providing the viewing distance is sufficiently large, and the image features are sufficiently blurred. Therefore only the Gaussian filters are considered further.

Lindeberg (1998) used peaks in the 1st derivative scale-space to locate edges. The amplitude of these derivative filters was normalised by multiplication with σ^{γ}, where σ defined the filter's scale and γ was a parameter that affected the placement of peaks in scale-space along the scale dimension. When $\gamma = 0$, peaks always occurred at the smallest scale, and when $\gamma = 1$, the peaks were always at the largest scale. Lindeberg presented a proof that when $\gamma = 0.5$, the scale-space 1st derivative peaks occurred at the scale that matched the blur of the edge. He also examined the effect of replacing 1st derivative filters with 3rd derivative filters. For those filters, $\gamma = 1.5$ provided similar scale-dependent performance. Lindeberg presented the model's line drawings of complex images, for both 1st derivative and 3rd derivative versions, and noted that the 3rd derivative version returned some edges that the 1st derivative version had missed. This was due to the 3rd derivative's narrower response properties to edges; in other words, it had finer resolving power. Therefore his algorithm employed 1st derivative filters to locate edges, followed by a stage where 3rd derivative filtering returned their amplitude and blur. The 3rd derivative was not used alone because it would return spurious flanking edges, as described in the next section. The final stage of the algorithm provided a measure of the saliency of each edge by integrating the edge strength along the contour. By selecting a restricted number (such

as 100) of the most salient curves, the algorithm produced an effective line drawing of a variety of test images.

1.10.2 Biological Vision

A scale-space model that detected edges using only the 3rd derivative was presented by Georgeson et al. (2007). It might seem that the use of the 3rd derivative would be problematic, because the 3rd derivative of a simple blurred edge has 3 extrema, not 1 (figure 1.11a), and so would predict 2 flanking edges of opposite sign to the central 'real' edge. However, Georgeson *et al* overcame this

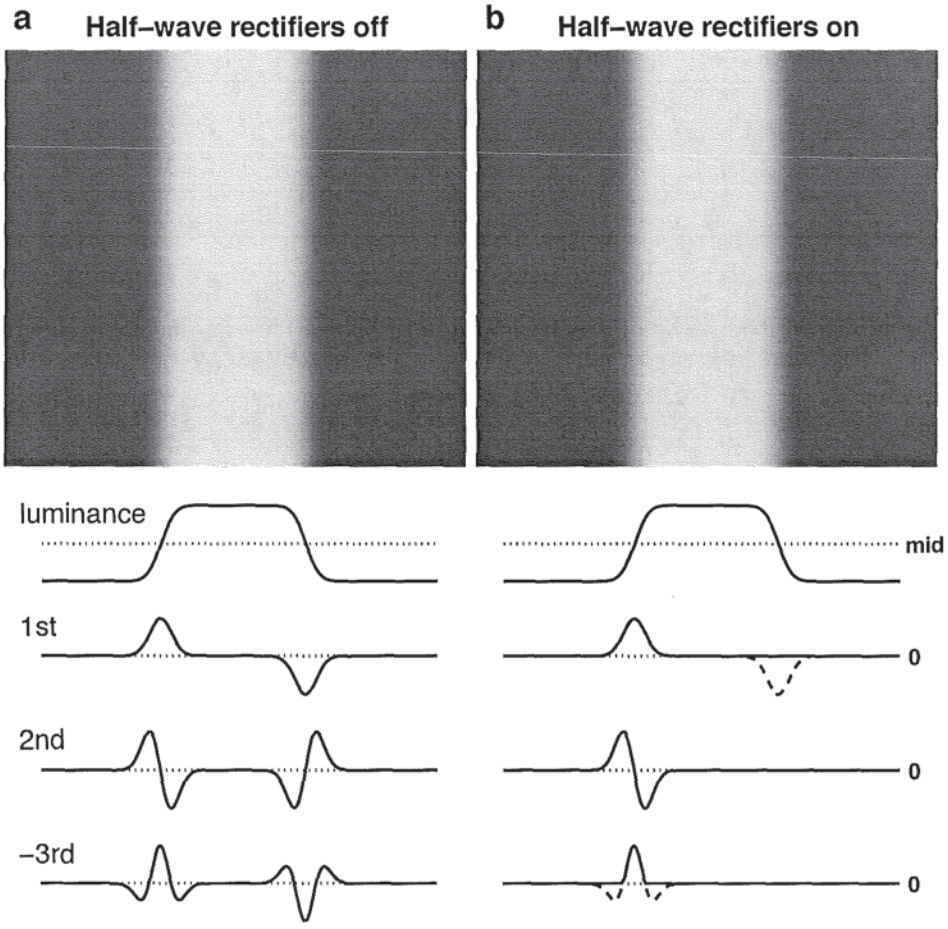

Figure 1.11: Two blurred edges and their first 3 derivatives. The 3rd derivative is shown inverted. a) Sequence of derivatives computed without half-wave rectification. b) Sequence of derivatives computed with half-wave rectification after the 1st and -3rd derivatives were obtained. Responses suppressed by rectification are shown by dashed lines.

problem by splitting the computation of the 3rd derivative into 2 stages: a 1st derivative followed by a -2nd derivative. A half-wave rectifier followed each stage. The first half-wave rectifier ensured that the mechanism was sensitive to edges of only 1 polarity. Therefore a 2nd complementary channel was needed for opposite polarity edges. The 2nd half-wave rectifier vetoed the spurious flanking extrema that are characteristic of the 3rd derivative of a blurred edge. Figure 1.11b shows the effect of this 2-stage sequence of rectification, and figure 1.12 shows the model's structure.

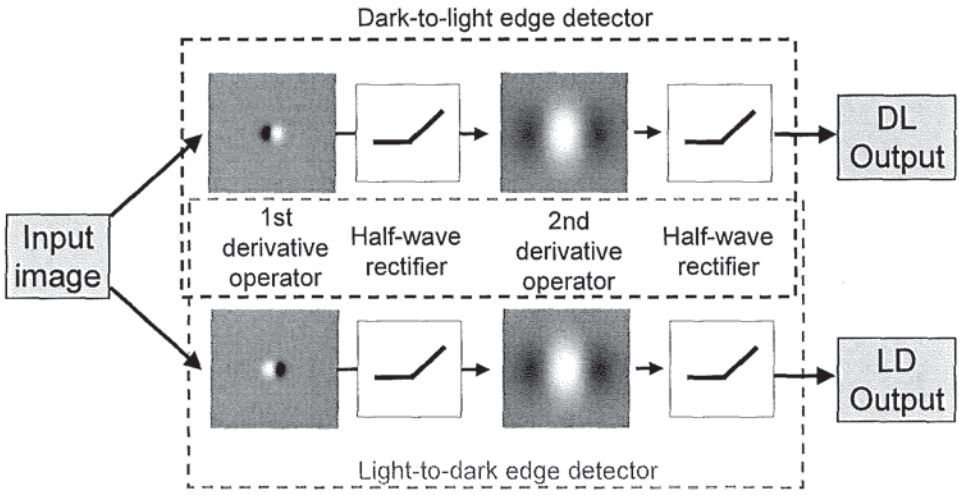

Figure 1.12: Structure of the N3+ model. Each operator shown represents a whole set of operators at a range of scales.

Following Lindeberg (1998), Georgeson *et al* applied this filtering process at a wide range of scales, and then normalised the filter responses by $\sigma^{1.5}$. Peaks in scale-space indicated the location, and scale of each edge. Edge polarity was indicated by the channel's identity. The complete model was termed 'N3+', because it used Normalised 3rd derivative filters + rectification. The model was tested against 6 other derivative-based models in a series of blur-matching experiments, as follows.

All experiments used a 2AFC procedure, where one interval contained a fixed test edge, and the other interval contained a comparison Gaussian edge whose blur was adjusted according to the observers response. The aim was to find the comparison blur that matched the test blur. Contrast levels were fixed at 30% or 32%. In experiment 1, the test edge was a composite of 2 Gaussian edges of the same polarity but different blurs. The test edge was selected from a range of blurs and contrast ratios. Experiment 2 used the same procedure as experiment 1 except that the test was a single Gaussian edge that was 'sharpened' to different degrees. Experiment 3 used a test edge that was a half period of a sinusoid flanked by luminance plateaux. In different conditions it was replaced by a single cycle of, or a full sinusoidal grating and was of various spatial frequencies. In experiment 4, the test was the central edge of a Gaussian derivative profile of odd order (-1, 1, 3, 5) at one of 3 different scales. Experiment 5 examined the effect of changing the length of the test edges used in experiment 4.

When the observers' matched blur was plotted as a function of contrast ratio of the stimulus components (experiment 1) or degree of edge sharpening (experiment 2), the functions were well-fitted by the N3+ model. The model had no free parameters. The predictions of other models fitted less well. Rejected models included a Gaussian edge template match to the test edge, a 1st derivative scale-space model, and 3 single-scale derivative-based models, where the derivatives were obtained at the finest scale possible. The first of these 3 assumed that edge blur was derived from the widths of the 1st derivative peaks. The second model derived blur from the separation between the adjacent peak and trough in the 2nd derivative (Watt & Morgan, 1983). The third of these 3 models took the ratio of 1st to 3rd

derivative as the measure of blur (Georgeson, 1994). Finally, for experiment 1 only, a rejected model was the contrast-weighted average of the 2 component Gaussian edge blurs.

In experiment 3, the matched blur of every sine-edge configuration and spatial frequency was also well-fitted by the N3+ model. Here, there were only 2 competing models considered; a scale-space 1st derivative model, and a version of the N3+ model that had no rectifiers (termed N3). Both models fitted less well than N3+, and this was also the case for the data of experiment 4, which had used the Gaussian derivative profiles of odd order. N3 and N3+ fitted the data of experiment 5 equally well, but the 1st derivative scale-space model provided a poor fit. To summarise all results, only the N3+ model fitted all of the data from a comprehensive set of edge types. The advantage of N3+ over other models was clearest for periodic images that contained adjacent edges of opposite polarity. The conclusion reached was that the N3+ model correctly captured some of the human visual processing of blur.

The N3+ model returns not only the amount of blur of each edge, as tested above, but also edge locations. In order to more thoroughly test the model, it was applied to the waveforms used in the previous feature-marking experiment of Hesse & Georgeson (2005), described above. The stimuli were 1D images derived from a blurred square-wave and manipulations of it, where the phase of its Fourier components were all shifted in 45° steps. Georgeson et al. (2007) demonstrated that the N3+ model predictions provided a good fit to the perceived edge locations across a range of phase shifts and edge blurs. In contrast, the local energy model predicted the absence of all these edges.

May & McIlhagga (2009) provided evidence that supported the use of 3rd derivative filters in the perception of blurred edges. They applied a methodology that is sometimes known as 'reverse correlation' (Ahumada & Lovell, 1971; Eckstein & Ahumada, 2002). This method employs test stimuli embedded in a field of background noise, and then correlates the observer's responses with the exact pattern of noise over a large number of trials. By averaging the noise pattern over a given subset of trials (e.g. where the response was that the target was present) a 'classification image' is built up, which identifies the template (or filter) that the observer used to identify the test stimulus. May & McIllhagga used a test stimulus that was a Gaussian blurred edge in noise and a comparison stimulus that was a step edge in noise. They obtained the mean difference between the test and comparison noise profiles for the correct trials and separately for the incorrect trials. The difference between the two defined the classification image, whose shape approximated a 3rd derivative filter. May & McIllhagga then used the N3+ model to derive a simulated prediction of the trial-by-trial performance and the resulting classification image. These predictions fitted remarkably well to the psychophysical data, with only a modest increase in the (now fitted) scale normalisation factor (previously 1.5).

The scale-space approach is consistent with known physiology. The receptive fields of cortical cells occur at a wide range of sizes, and they are bandpass tuned in spatial frequency (De Valois et al., 1982; Movshon et al., 1978), which means that they are scale-selective. The first stage of the N3+ model consists of linear 1st derivative filtering followed by half-wave rectification. Some supportive evidence for a 1st derivative filtering role for cortical cells has been described above (section 1.7.2), so will not be repeated here. Maffei (1985) described cortical cells in cat that responded in proportion to the luminance of light (or dark) bars of a drifting or flickering grating and therefore described these cells as linear with half-wave rectification. The recordings from complex cells showed evidence of full-wave rectification when presented with a flickering grating, but their response was unmodulatory to a drifting grating. Therefore these cells were described as highly non-linear. Movshon et al. (1978) also

described cat simple cell responses as linear filters followed by a half-wave rectifier.

Georgeson et al. (2007) presented a receptive field analysis of the output of the N3+ model's 2nd stage of filtering and half-wave rectification (before construction of the scale-space map). The resulting receptive field had an excitatory region and an inhibitory region that overlapped considerably. This is reasonably consistent with the observed absence of separate (or equivalently, the complete overlap of) ON and OFF regions in the cat complex cell (Hubel & Wiesel, 1962). Martinez et al. (2005) recorded from cat V1 cells that had an ON region adjacent to an OFF region and applied an overlap index (Schiller et al., 1976) that had a value of 1 when the regions overlapped entirely, zero when the regions were abutting, and was negative when the regions were separate. A subset of cells whose index was 0.32 to 0.84 was defined as complex. While Georgeson *et al* did not calculate an overlap index for the N3+ receptive field, inspection of its profile (their figure 9) suggests that it appears to fall within the range 0.32 to 0.84. Mata & Ringach (2005) examined V1 cells in the anaesthetised monkey and also calculated the overlap index. The pattern of index values formed a continuous and bimodal distribution from -1 to 1, which clearly encompasses Georgeson *et al*'s N3+ receptive field. A similar finding was reported by Kagan et al. (2002) who used alert monkeys.

The 2-stage nature of the N3+ model also has physiological support. Hubel & Wiesel (1962) posited that simple and complex cells represented distinct stages in a feed-forward visual process. It has been found that some cat V1 layer 4 cells project to layer 2+3 (Hirsch et al., 1998; Martinez & Alonso, 2001; Martinez et al., 2005), where the majority of cells are complex. Furthermore, complex cells have receptive fields that are consistent with a number of simple-cell components (Anzai et al., 1999; Movshon et al., 1978; Szulborski & Palmer, 1990).

To return to psychophysical work, the N3+ model was developed further by Georgeson (2006) by the addition of a bar scale-space map that was derived from 2nd derivative Gaussian filters. Edge and bar scale-space maps were produced in parallel and then combined to produce a composite map. A candidate feature was identified by a peak in the individual edge or bar map. But it was deleted in the composite map if its peak was occluded by a peak derived from the other feature map. This behaviour accounted for mutual inhibition between edges and bars (Tolhurst, 1972b), but only where the edge and bar were sufficiently similar in scale and spatial location.

This edge-bar model correctly predicted the presence of Mach bands on a Mach ramp and a sine edge, and their absence on a Gaussian edge. The model also correctly predicted the observed suppression or enhancement of Mach bands caused by edges or bars that were brought close to the Mach ramp (Ratliff et al., 1983; Ratliff, 1984). Moreover, when the model was applied to the waveforms used in the feature-marking experiment of Hesse & Georgeson (2005), it correctly predicted the location and polarity of the bars in all conditions where they were marked. It also correctly predicted no bars when the waveform's phase was 90°. In this condition, a large peak in the edge map occluded 2 smaller neighbouring peaks in the bar map, so the bar responses were suppressed. The model also predicted a new phenomenon: 'Mach Edges', which are illusory edges that are seen when the gradient of the luminance profile (rather than the luminance profile itself) is a ramp bounded by two plateaux. One main aim of this thesis is to explore these ideas and findings further.

Before launching into the experimental work, which commences in the next chapter, a brief tutorial on signal detection theory will be presented. This is included here, because signal detection theory is applied extensively in the analysis of the experimental results in chapters 5 and 6.

1.11 Signal detection theory

Signal detection theory is a framework that has been extensively used in psychophysics (Green & Swets, 1966; Macmillan & Creelman, 2005). One of its chief benefits is that it can separate out the effect of criterion shifts from experimental data, so that a true measure can be obtained of the discriminability of different stimuli. It is based on a number of assumptions. Firstly, the experimental design should include events where a stimulus is either present or absent, and in either case, the observer should be able to respond that the stimulus is present or absent. The observer is assumed to collect all of the information from a stimulus into (in the 1-dimensional case) a single value on a single axis in 'decision space'. But on subsequent trials that present the same stimulus, the response value varies somewhat, due to noise, which is assumed to be Gaussian. Over several trials, a given stimulus thus forms a Gaussian probability distribution in decision space, whose mean represents the most likely response value.

To take the example of a yes-no task, on each trial the observer is assumed to respond 'yes' if the response value in decision space falls above a fixed criterion level, and 'no' if it falls below the criterion. When a Gaussian distribution of response values is considered, then the probability of a 'yes' response is the area of the distribution that falls above (to the right of) the criterion. This is illustrated by the shaded region in the centre panel of figure 1.13. The decision axis units are the standard deviation of

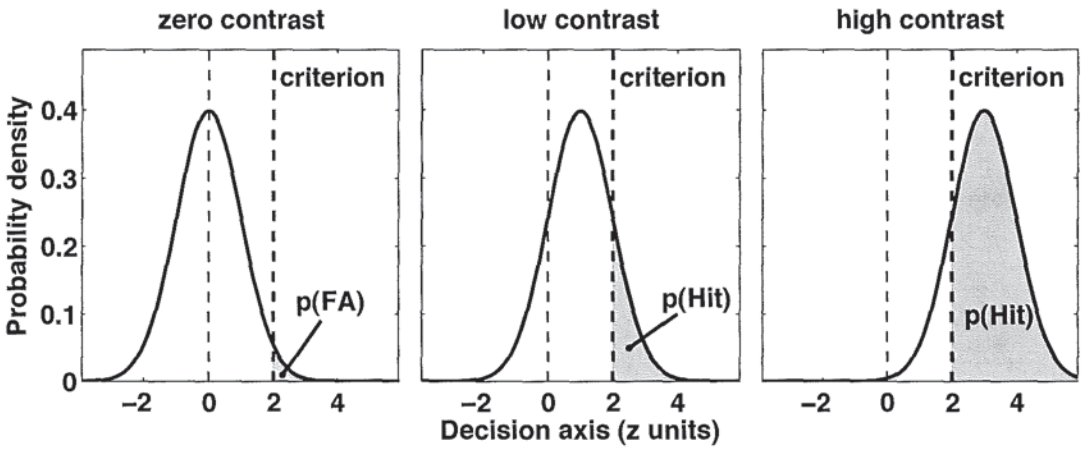

Figure 1.13: Probability density functions of internal responses in decision space that arise from repeated presentations of a single stimulus that is zero contrast (left), low contrast (centre) or high contrast (right).

the distribution. This means that they are z-score units. When the stimulus contrast is above zero, the proportion of 'yes' responses is known as the 'Hit rate', because the observer has made correct 'yes' responses. When the stimulus contrast is zero, the proportion of 'yes' responses is known as the 'False Alarm rate' (FA), because the observer has made incorrect 'yes' responses (figure 1.13, left panel).

The discriminability (d') between the zero contrast ('non-signal') and above-zero contrast ('signal') stimuli is represented by the distance between the peaks of those distributions, shown in figure 1.14. It can be calculated from

$$d' = z(Hit) - z(FA)$$

because $z(Hit)$ is the distance between the criterion and the peak of the signal distribution, and

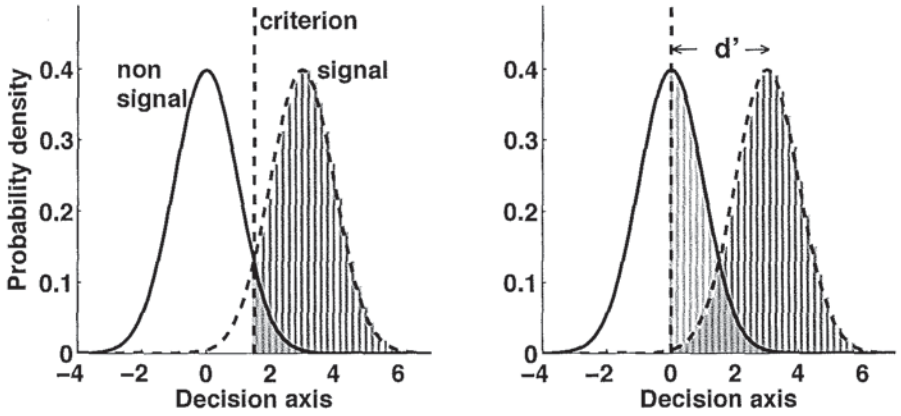

Figure 1.14: Non-signal and signal probability density functions with a criterion position that is optimal (left) or biased (right).

$-z(FA)$ is the distance between the criterion and the peak of the non-signal distribution. Thus, for a fixed signal and non-signal, the measure of d' is independent of the criterion position.

The right panel of figure 1.14 shows the same distributions as the left panel; only the criterion's position has changed. This change would increase the Hit rate. Therefore, if performance was measured simply from the proportion of correct responses to the stimulus (Hits), then an improvement would be apparent, which is simply an artifact of the criterion shift. It is the removal of this artifact that is the huge benefit gained by using d'.

This account can be extended to provide an explanation of how a sigmoidal psychometric curve can be predicted from stimulus values. The right panel of figure 1.13 shows the internal response distribution of a high contrast stimulus. The Hit rate is higher (shaded region is larger) than in the low contrast stimulus case. As stimulus contrast varies, the criterion is assumed to remain fixed. Therefore the observer's Hit rate increases smoothly with contrast, and by an amount that is defined as

$$Hit = 1 - \int_{-\infty}^{c} G(x - x_0)dx \tag{1.5}$$

where c is the criterion level and $G(x)$ is the unit area Gaussian, whose peak is at x_0.

When the observer is unbiased, the psychometric function of hit rate *vs* contrast level is centred on zero, as shown in figure 1.15 (dark solid line, top left panel). But if the observer is biased, then the criterion is shifted along the decision axis, and this shifts the psychometric function horizontally along the stimulus axis, as shown by the other lines in the top left panel of figure 1.15.

The above explanation assumes that the mapping from stimulus value (contrast) onto the observer's decision axis is linear; in other words, that the observer's transducer (the function that maps stimulus value onto mean response) is linear. However, if the transducer is non-linear (eg a square-law, $r = sign(c).c^2$), then the resulting psychometric function contains a central kink, as shown in top right panel of figure 1.15. The nature of this kink is clearer when the z-scores on the psychometric function for Hits are considered. The middle row of this figure shows the z-scores of the curves in the top row. The left panel z-scores are linear, as expected from a linear transducer, and the right panel z-scores accelerate as the square of the absolute stimulus value. The lower row of figure 1.15 shows

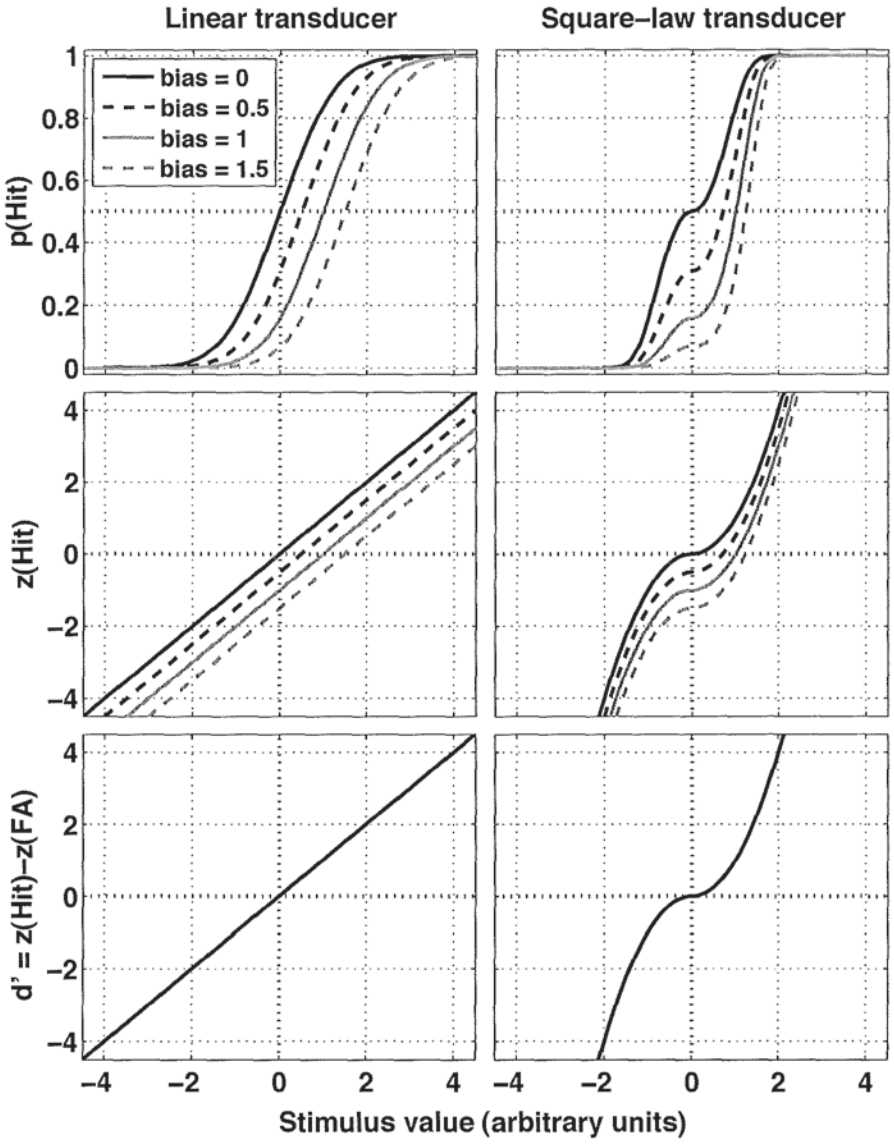

Figure 1.15: Effect of bias on ideal observer responses that are based on a linear transducer (left) or a square-law transducer (right).

discriminability (d'). The subtraction of $z(FA)$ has the effect of vertically shifting each of the curves in the middle row of figure 1.15 by an amount that causes them to superimpose. The fact that the shift is vertical rather than horizontal is unexpected, as a criterion shift appears to cause a horizontal shift in the linear transducer case. However the middle left panel of 1.15 shows that the shift in z-scores caused by a criterion shift can be described equally well as a vertical or horizontal shift when a linear transducer applies.

To conclude, when stimulus values range from negative through zero to positive, a non-linear transducer is expected to cause a psychometric function for Hits that contains a central kink. Observer bias is expected to cause a vertical shift of the whole z-score function and a vertical shift of the central kink of the p(Hit) psychometric function.

Chapter 2

General methods

2.1 Psychophysical methods

All of the experiments in this thesis employed psychophysical methods. The term 'psychophysics' was coined by Fechner (1860), and relates to the scientific study of the relation between physical stimuli and how they are perceived. Since its inception, the study of vision has expanded to include methods that are no longer defined as psychophysics, such as neurophysiolgy and imaging techniques. However, despite many undoubted successes, these other methods have certain disadvantages. Neurophysiology is an invasive technique of recording or stimulating neurons that are, in the present context, involved in visual perception. As such, it is predominantly performed on animals, and can supply information regarding only a small number of neurons at one time. Psychophysics is non-invasive, is performed predominantly on humans, and encompasses a large part of the visual system. Although imaging techniques, such as MEG and fMRI, are non-invasive, they only provide information regarding the site of brain activity (and to a limited resolution). They do not tell us what kind of processing is being performed. However, psychophysics, in combination with a suitable model, *does* provide information about how the visual signal is processed in the human brain.

Early ('classical') psychophysical methods include the method of constant stimuli (MCS) which uses a pre-determined regime of stimulus presentation that is not affected by the observer's responses (for a review of psychophysical methods see Gescheider, 1997). The data it produces are typically used to produce a 'psychometric function', which is a plot of the observer's response as a function of stimulus magnitude. To take the example of a simple detection experiment, the psychometric function would show the proportion of 'I see it' responses *vs* stimulus contrast. At low contrasts, the observer makes few (if any) 'I see it' responses, and at high contrasts he makes many such responses. As contrast progressively increases from zero, the transition from few to many responses is gradual. This results in a S-shaped psychometric function, which can be summarised by its horizontal position and slope. The horizontal position is defined by the ('threshold') contrast level that produces a certain response proportion (e.g. 0.84).

However, MCS can be argued to be inefficient because every point on the psychometric function is derived from the same number of trials, even though the resulting threshold and slope are influenced much more by the data close to threshold than data remote from it. A more efficient method is to devote

a large number of trials near threshold and relatively few at the tails of the psychometric function. An adaptive method can be implemented by using a 'staircase' procedure, whereby the observer's response determines whether the contrast level is shifted up or down a level on the next trial.

Although staircases are widely-used in visual psychophysics, they have not been employed in any experiment in this thesis, because their focus on near-threshold contrast levels means that we can be less certain, compared to MCS, of the psychometric function's shape at low contrast levels. This region of the function is of critical importance when considering the linearity or nonlinearity of performance, which is the theme of chapter 6.

2.1.1 Feature marking

In this thesis, the feature marking method is employed in experiments 1.1 to 1.4, 3 and 5.2. It is a 'subjective' method (subject to criterion effects) whereby the observer indicates the attributes of particular features in the image, such as the position and polarity of all perceived edges. When feature position is recorded, an on-screen marker is typically employed, which is moved across the image by the observer until it coincides with a perceived feature (Hesse & Georgeson, 2005; Morrone & Burr, 1988; Wallis & Georgeson, 2009), and its position is then recorded. Alternatively, the marker is fixed (e.g. at screen centre) and the image is moved relative to it (Georgeson & Freeman, 1997; Meese, 1993; Tyler & Gorea, 1986).

2.2 Apparatus

All of the experiments in this thesis were run on an Apple MAC computer using Matlab software (version 5.2.1.1421) within a classic OS9 environment. Stimuli were passed through a Cambridge Research Systems Bits ++ box, running in Mono++ mode, which made 16382 grey-levels available (14-bit resolution). Such a large number of grey-levels ensured that the low-contrast images always appeared smooth and free of artifactual edges. All stimuli were grey and were displayed on an Eizo Flexscan 6600-M monitor, which had been previously calibrated. Since there is no look-up table (LUT) with Mono++ mode, the digital stimuli were gamma corrected before being passed through the Bits++ box.

2.2.1 Monitor calibration

Cathode ray tubes are controlled by a digital-to-analogue converter (DAC) that linearly converts the digital input to a voltage that drives the beam intensity. However, the luminance of the displayed image is not linearly related to the values in the digital input. Typically, luminance increases in proportion to the digital input raised to a power (gamma) of between 2 and 3. In order to compensate for this non-linear behaviour, the monitor was calibrated and the digital input images gamma-corrected. Calibration refers to the process of adjusting a device to a desired configuration. As part of that process, the display equipment needs to be characterized, which refers to the process of modeling the equipment and measuring its properties to allow accurate rendering.

The relation between the digital input values and the observed luminance was modelled by:

$$\frac{L-k}{L_{max}-k} = \left(\frac{j-j_0}{j_{max}-j_0}\right)^{\gamma}, \text{for} j \geq j_0, \text{ else } 0 \tag{2.1}$$

where L is the luminance, j is the DAC input value, and L_{max}, k, j_0 and γ are fitted parameters that are estimated from the calibration, and j_{max} is the highest permitted value (Brainard et al., 2002). With 14-bit resolution, the theoretical value for j_{max} is $(2^{14}-1)$[1]. However, this value (16383) is odd, which means that that there is no single value that will provide a luminance that is half of maximum. All of the experiments in this thesis use a background luminance that is exactly half of the maximum. So to ensure that a single DAC value existed for that luminance, j_{max} was set to $(2^{14}-2)$.

Characterization was implemented with a custom Matlab program supplied by Mark Georgeson and a Minolta LS110 digital photometer. The displayed luminance was measured for 64 different grey levels, spanning the full range. The luminance was then plotted as a function of DAC value, which was an accelerating function, as expected. A downhill Simplex algorithm was used to minimise the squared error between this empirical curve and the fitted curve predicted by equation 2.1. The best fitting values of L_{max}, k, j_0 and γ were stored, and recovered each time a stimulus was generated, and used to generate a LUT that linearised the relation between luminance and the requested grey levels. Thus the stimuli were gamma-corrected.

2.3 Experimental set-up

All experiments were performed in a windowless room, where the only source of light was from the CRT. The stimuli were viewed binocularly and with natural pupils. A constant viewing distance was maintained by using a chin-and-forehead rest, except for experiment 4, where the relatively long viewing distance meant that slight movements would have a minimal effect on image size. All stimuli were viewed directly, except for experiment 4, where a plane mirror was used.

2.4 Observers

All of the experiments in this thesis used the author (SAW) as the principal observer. He had a standard eye examination shortly before and during the course of the PhD. In both instances this revealed a slight astigmatism in one eye ($\leq$ 0.5DC) that did not require correction. His vision was otherwise normal. All other observers had normal or corrected-to-normal vision. The strongest correction (approx. -9.00DS) was worn by observer PRM, who was used in experiment 10 only. She wore contact lenses for all data collection, in order to prevent the considerable image minification that would have occurred with spectacles.

[1] We subtract 1 here because the lowest value of j is zero.

2.5 Data analysis

Various methods of data analysis were applied to the data in this thesis, depending on the experimental design. However, the experiments in chapter 5 and part of chapter 6 all employed a common program to fit a curve to the resulting psychometric functions and obtain confidence intervals. That program will be described next.

2.5.1 Psychometric fit

It is desirable to fit a curve to empirical psychometric functions, because data invariably contain some noise that is assumed to represent random experimental fluctuations from an underlying smooth function that relates stimulus magnitude to response performance. Furthermore, the parameters of the fitted curve can be a useful summary of the data set. The fitting program described here was written in Matlab and supplied by Mark Georgeson. It used a downhill Simplex algorithm to maximise the likelihood of the data set given its fitted curve (Wichmann & Hill, 2001a,b). The curve was generated in a 2-stage process. First, the proportions of 'stimulus seen' responses were generated from the stimulus contrast levels by using the formula

$$p_1 = \Phi\left(\text{sign}(c).k|c|^m - \text{sign}(c_{50}).k|c_{50}|^m\right) \tag{2.2}$$

where c is the experimental contrast level, c_{50} is the contrast level that provides a 50% responses rate, k is a fitted 'gain' parameter, and Φ is the standard normal integral that transforms a z-score into a proportion. The terms to the right of the minus sign represent the influence of observer response bias. Next, the lapse error rate (*lapse*) was incorporated into the resulting probabilities by using the formula

$$p = p_1(1 - lapse) + \frac{lapse}{2} \tag{2.3}$$

This has the effect of vertically shifting all the data points towards the 0.5 response point, by an amount that is proportional to their distance from it. When the lapse error rate is 0.1, the resulting probabilities cannot extend beyond 0.05 and 0.95.

The combination of equations 2.2 and 2.3 has 4 free parameters (m, c_{50}, k and *lapse*). When m is fixed at 1, the resulting fit is a Gaussian. The contrast level that provides a 50% responses rate (c_{50}) represents the 'null point' or category boundary between positive and negative responses for the feature or task in question. The threshold (at $d' = 1$) was defined as the stimulus level 1 z-unit above c_{50} (giving 84% positive responses), where c_{50} was the natural choice. However, the pattern of data sometimes suggested a natural null point that was at 0% contrast, rather than c_{50}. On those occasions, the threshold was defined as 1 z-unit above the response to the 0% contrast stimulus.

The best-fitting values of the free parameters were obtained by using maximum likelihood estimation. This technique maximises the log likelihood (L) of obtaining a particular set of data, assuming that it was derived from the fitted curve. It was implemented by using a downhill Simplex algorithm to maximise the value of L returned by

$$L = \sum_{c=1}^{c=n} (y_c.log(p_c) + n_c.log(1 - p_c)) \tag{2.4}$$

where n is the number of contrast levels, y_c is the number of positive responses at a given contrast level indexed by c, n_c is the number of negative responses, and p_c is the proportion of positive responses predicted by the fitted curve. The computation within the brackets was performed for every stimulus contrast level, and the resulting n values of L were summed, to arrive at the final log likelihood.

Once the best-fitting parameter values had been obtained, a parametric bootstrapping procedure using a large number of runs was employed to obtain the 95% confidence limits.

2.5.2 Bootstrap

Bootstrapping is a method of estimating the properties of a data set (such as its variance) by drawing many samples from an approximating distribution, and measuring those properties in the whole set of samples (Efron, 1981). It can be conceptualised as a practice of using the existing set of data to simulate many re-runs of the experiment in order to measure the variability in the simulated data, and so determine the position of the original experimental data within that range.

In the nonparametric bootstrap, the approximating distribution is the experimental data itself. So any outliers that exist in the data will tend to influence each drawn sample. This problem can be avoided by using a smoothed version of the data as the approximating distribution, such as that generated by fitting a Gaussian or power function. This is the basis of the parametric bootstrap (Efron, 1979), which was implemented here as follows.

One set of synthetic data was generated from binomial sampling of the fitted curve, and then these new data were fitted and the values of the fitted parameters were recorded. This process was repeated for many runs, thus generating many values for each parameter. For each parameter, the values were placed in rank order and the values at the 2.5% and 97.5% points indicated the 95% confidence limits. This method, however, relies on the assumption that the fitted curve is an accurate indication of the underlying function that relates stimulus to response. In order to provide an estimate of the goodness of fit of the fitted curve to the experimental data, the deviance of each set of simulated data from the fitted curve was obtained and plotted as a distribution. This was examined to determine whether the deviance of the data from the fitted curve fell within the 95% limits of the distribution.

Chapter 3

The 3rd derivative model is tested and refined

3.1 Experiment 1.1 - Mach edges

3.1.1 Introduction

We saw in chapter 1 that the derivative approach to edge detection has predominantly focused on peaks in the 1st derivative or zero-crossings in the 2nd derivative, and that both methods have their limitations. A recent development was described, that employs the 3rd derivative of the luminance profile for edge localisation (Georgeson et al., 2007). These researchers re-examined previous feature-marking data (Hesse & Georgeson, 2005), where edge positions marked by observers had been well predicted by the peaks in the 1st derivative, and found that a scale-space non-linear 3rd derivative model (the N3+ model) provided similar predictions.

While the blur predictions of the N3+ model have been found to fit experimental data better than competing models (Georgeson et al., 2007) the position predictions have not been shown to out-perform a model based on 1st derivative peaks. The stimuli used by Hesse & Georgeson had peaks in the 3rd derivatives at similar locations as the 1st derivative peaks, so the edge position data could not easily distinguish between the two models. The present experiment aims to provide this distinction by using stimuli that do separate out the peak positions of these two derivatives in a single waveform.

An example of this waveform is shown in figure 3.1, with its first three derivatives and an image that has been generated from it. It is a single cycle of a triangle-wave that has been blurred. The blurring has transformed its apex into a smooth curve of constant convex curvature. The width of this blurred apex is determined by the amount of applied blur, and will therefore be termed 'blurwidth'. Beyond the apex, the luminance falls away linearly until it meets the concave curvature caused by the blurred trough at each side of this periodic waveform. The linearity of the decrease in luminance is more clearly seen as a plateau in the 1st derivative (which is the gradient of the luminance profile). Note that the 1st derivative has no peaks; only a succession of plateaux, interleaved with linear ramps. This property is important, as it means that any edge-detection model that relies on 1st derivative peaks will predict

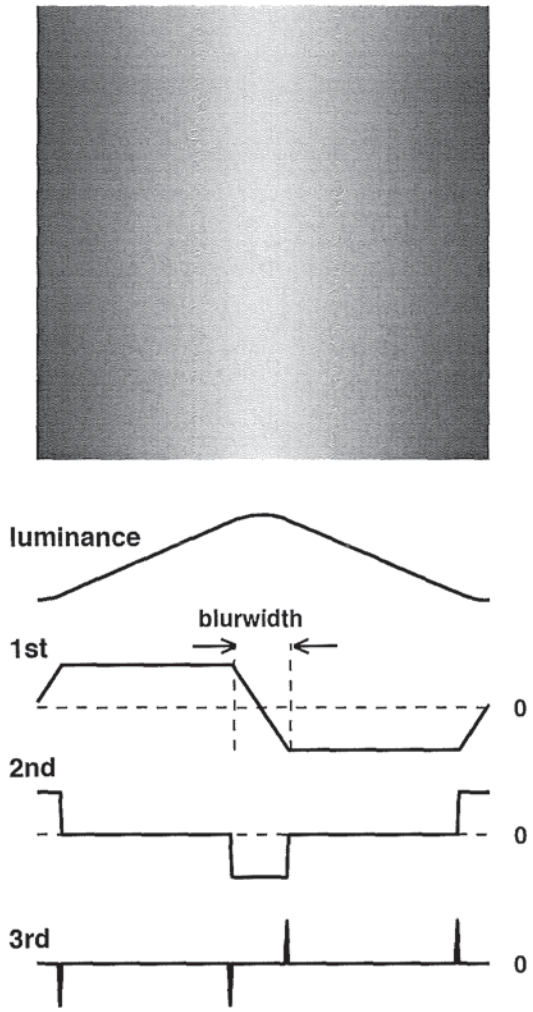

Figure 3.1: High contrast example of an experiment 1.1 image (blurwidth 32 pixels, peak central) and its first three derivatives.

no edges in this image[1].

The gradient of the 1st derivative is the 2nd derivative, which represents the curvature of the luminance profile. Figure 3.1 shows that the 2nd derivative remains at zero where the luminance profile is linear, and abruptly changes to a plateau where the luminance profile begins to curve. Crucially, it has no zero-crossings, only step-changes to or from zero. Therefore, any edge-detection model that relies on 2nd derivative zero-crossings will predict no edges in this image.

The 3rd derivative of the luminance profile is shown at the bottom of figure 3.1. This does contain clear peaks and troughs that are located where the luminance curvature changes. So any edge-detection model that relies on 3rd derivative peaks will predict edges at these specific locations. The first experiment's aim is to test this prediction. Given the absence of 1st derivative peaks and 2nd derivative zero-crossings, it is not clear whether observers will perceive any edges at all.

[1]Strictly speaking, at a very large scale there is a peak on the left side of the 1st derivative profile and a trough on the right side. So a 1st derivative model may predict a large scale edge centred on each plateaux, but it would be of low amplitude.

3.1.2 Method

Stimuli were generated by construction of an initial waveform that formed the 1st derivative of the luminance profile. This was integrated using the Matlab function 'cumtrapz' to obtain the luminance profile. The initial waveform was generated from a single period of a square wave that was convolved with a box function whose width was chosen from 1, 2, 4, 8, 16, 32 or 64 pixels. This convolution produces a trapeziodal-wave whose ramps have a horizontal extent that is equal to the width of the box function. This initial waveform was integrated to form a blurred triangle-wave. An equivalent stimulus description is of a triangle-wave blurred by a box function, whose width is from 1 to 64 pixels. Example waveforms and their first three derivatives are shown in figure 3.2. There are no 1st derivative peak and no 2nd derivative zero-crossing, so no edges are predicted by any model based on these derivatives. However, the 3rd derivative predicts two edges of opposite polarity near the centre of each image.

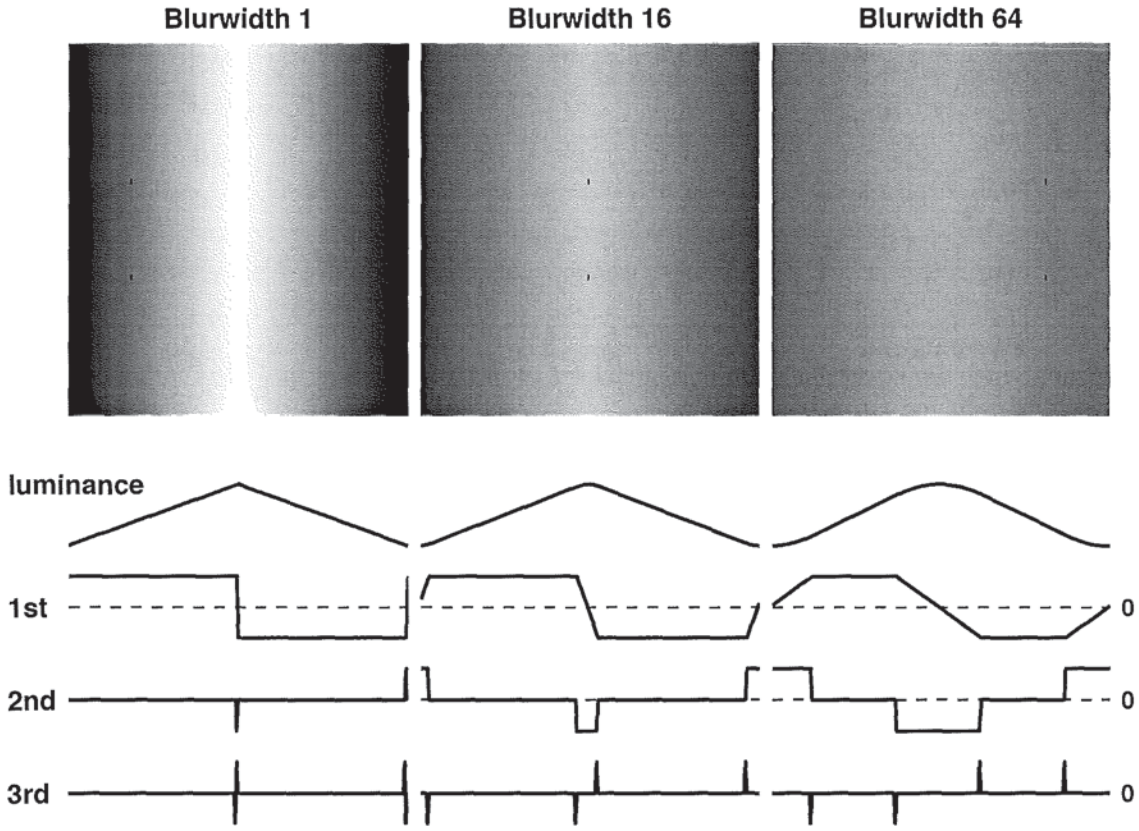

Figure 3.2: High contrast examples of three experiment 1.1 images (peak central) with marker, and their first three derivatives. The left and right panels show the minimum and maximum blurwidths (respectively) used in this experiment, and the left-most and right-most (respectively) permitted marker positions.

Waveforms had one of two phases: 'peak central' (see figure 3.1 or 3.2), or an inversion of this named 'trough central'. The waveforms were used to form vertically-uniform 1D images that were 256 by 256 pixels. They subtended 4 degrees at the viewing distance of 131.6cm, thus remaining within the

observers' fovea. The images were surrounded by a full-screen mid grey of luminance $40.7cd/m^2$, and were viewed at a contrast of 0.2 or 0.4. The images were displayed flashing in order to reduce adaptation (presentation 300ms, isi 600ms). The presentation time was short in an attempt to reduced the fluctuations of perceived contrast caused by the combination of negative afterimages and eye movements (Georgeson, 1984). The inter-stimulus display was of a full screen of mid-grey.

The task was to indicate the position and polarity of all edges seen in each image. The position was identified by moving a marker across the image and pressing a Keyspan remote button when the marker was over an edge. A second button-press indicated the polarity of the edge as light-to-dark (LD) or dark-to-light (DL). Once all the perceived edges had been marked, the observer initiated the next trial.

The marker consisted of two black dots, each of 1 x 3 pixels. One was centred 32 pixels ($\frac{1}{2}$ degree) above, and the other 32 pixels below, the horizontal mid-line of the image. The observer was instructed to fixate midway between the two dots, in order to reduce any effect that the marker itself may have upon the location of the perceived edges. The marker position was 32 pixels to the left of the image centre at the start of the first trial. Its starting position alternated between left and right on each trial. Its movement was constrained to the central 2.5 degrees of the image, in order to prevent responses to the peripheral troughs in the waveform (which were shifted to the image centre in 'trough central' conditions). The left and right panels of figure 3.2 shows the peripheral limits of marker position.

The 28 conditions (7 blurwidths, 2 phases, 2 contrast levels) were presented in randomised order, blocked by contrast. This procedure was repeated a further two times in one experimental session, which took about 30 minutes to complete. The three observers (SAW, DHB and TAY) were all psychophysically experienced and had normal uncorrected vision. They viewed the stimuli binocularly, with natural pupils, in a darkened room, with their head supported by a chin and head rest. They each completed three sessions.

3.1.3 Results

Each observer marked a pair of opposite polarity edges near the image centre, in every condition. Since there were 9 trials for each condition, an average position of each edge was calculated from 9 values. These average positions are shown in figure 3.3 plotted against blurwidth, separately for the three observers and two image phases. The left column is from peak-central waveforms and the right column is from trough-central waveforms. Each row shows data from one observer. Black symbols denote DL edges, and grey symbols denote LD edges. Triangular symbols denote contrast 0.2 images and circular symbols denote contrast 0.4. The plots show little or no effect of contrast, as the two symbol shapes generally overlap for each blurwidth and edge polarity.

The plots also show that perceived edge position remains approximately constant for blurwidths 1 to 4, at about ± 4 pixels. As blurwidth increases above 4, the marked edges spread smoothly and symmetrically away from the image centre to a maximum of approximately ± 20 pixels (for SAW) or ± 30 pixels (for DHB and TAY). Error-bars (± 1 standard error) are plotted behind symbols, and are smaller than symbol size, except for some of the blurwidth 64 data.

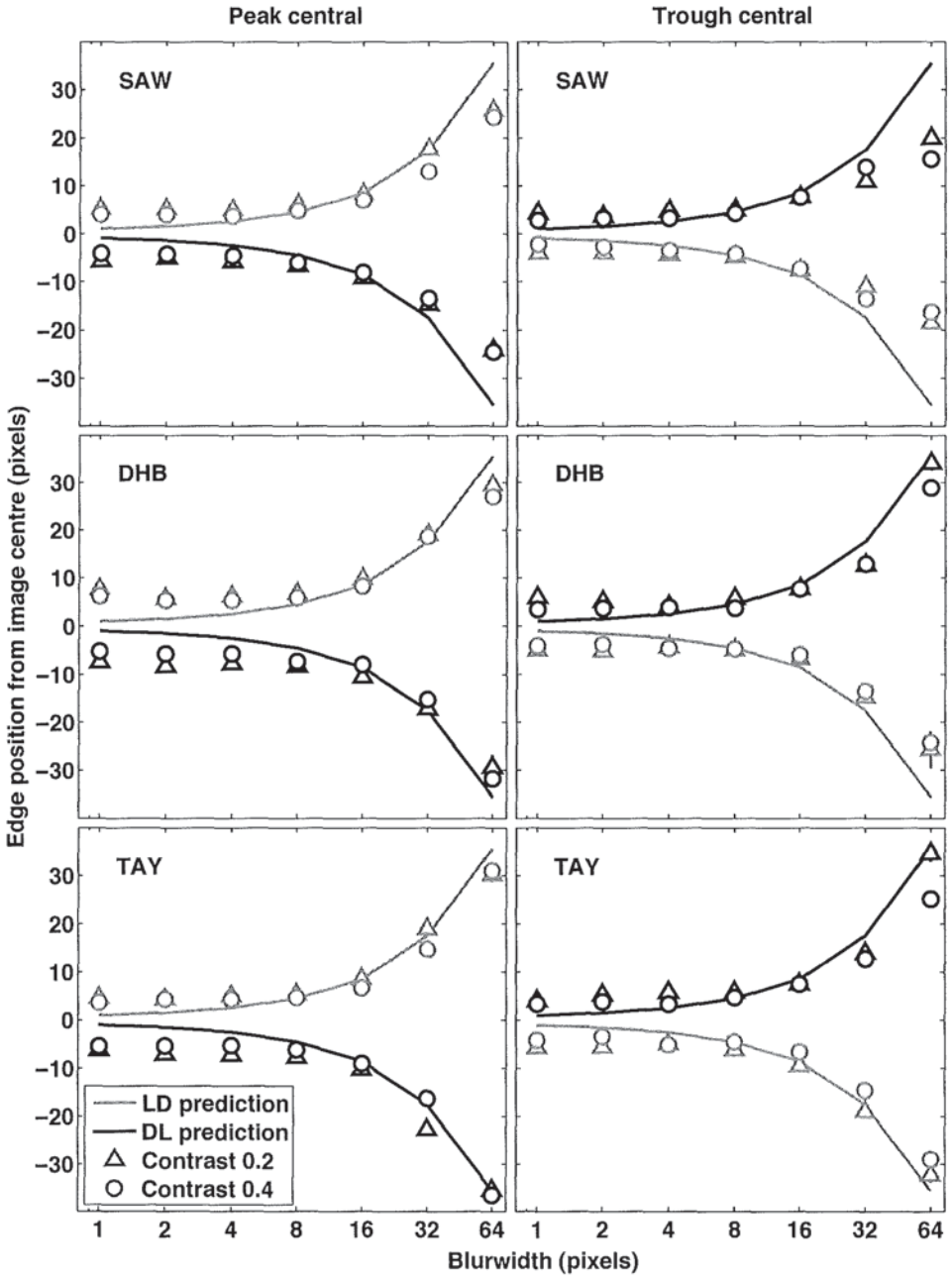

Figure 3.3: Mean of 9 trials (symbols) and 1D N3+ model predictions (lines) for positions of dark-to-light (DL) and light-to-dark (LD) edges in 1D blurred triangle-wave images. Left column is from peak-central waveforms and right column is from trough-central waveforms. Each row shows data from one observer. Errorbars (±1 standard error) are plotted behind symbols and are generally smaller than symbol size. 64 pixels = 60 arcminutes.

3.1.4 Discussion

All observers perceived edges in these stimuli, despite the absence of 1st derivative peaks and 2nd derivative zero-crossings, at a small or moderate scale. At a very broad scale, the 1st derivative approach does predict an edge centred on each 1st derivative plateau, due to the following explanation. Since convolution is distributive, a broad scale 1st derivative operator is equivalent to a fine scale 1st derivative operator, followed by blurring. If the 1st derivative of an experimental waveform, shown in figure 3.2, was blurred, it would become more like a sinusoid. Its upper plateau would become a shallow peak, whose maximum lies mid-way between the image centre and its left-hand border. Likewise the lower plateau would become a shallow trough, whose minimum lies mid-way between the image centre and its right-hand border. These peaks and troughs would therefore signal edges in a broad scale 1st derivative operator.

By a similar argument, a broad-scale 2nd derivative operator would produce zero-crossings at the same locations. For all experimental images used here, these predicted edges lie at ± 64 pixels. However, observers never marked edges at these locations.

Since the 1st and 2nd derivative approaches are unable to account for perceived edges at the marked locations, a different approach must be employed. The 3rd derivative of each experimental waveform does produce a clear peak and trough near the centre of each image (see figure 3.2), so an approach based on it may be fruitful. Third derivative filters are the basis of the N3+ model, so an implementation of it was applied to the experimental waveforms. Since the stimuli were 1-dimensional, a simplified version of the N3+ model was used here, where the circular filters were replaced with 1D versions. It's edge predictions are shown as lines in figure 3.2, where line shade denotes edge polarity (DL = black, LD = grey). Prediction match data well, although they are slightly too central at low blurwidths, and slightly too peripheral at high blurwidths. Nevertheless, these data clearly support the use of 3rd derivative operators - specifically that edges may be seen at peaks and troughs in the 3rd derivative, where there are relatively abrupt changes in the curvature of the luminance profile.

The perceived edges will henceforth be referred to as 'Mach edges', because they are analogous to Mach bands for reasons given below. Mach bands are illusory perceived bars where there is no peak in the luminance profile (Ratliff, 1965). They are seen in images where a luminance ramp meets a plateau; a configuration that has come to be known as a Mach ramp. An example of a Mach band image is shown in the left panel of figure 3.4 with its luminance profile and first three derivatives. Although the luminance profile is a linear ramp flanked by two plateaux, the reader may observe an apparent vertical light bar in the image, which is located at the junction of the upper plateau and the ramp. This is illusory, as there is no peak here in the luminance profile. It is a Mach band. Likewise, a dark Mach band may be observed at the junction of the lower plateau and the ramp. The right panel of figure 3.4 shows a high contrast example and corresponding waveforms of a blurwidth 32 experimental image. Note the similarity in shape of the experimental image's 1st derivative (and higher derivatives) and the Mach bands luminance profile (and higher derivatives). The experimental image luminance is the integral of a Mach ramp. Just as Mach bands show unequivocally that all bars are not peaks in the luminance profile, so the Mach edges show that not all visible edges arise from peaks in luminance gradient.

To return to the data shown in figure 3.2, the error-bars are small. Also, in every trial, both edges were marked. These observations suggest that Mach edges are stable and reliable phenomena.

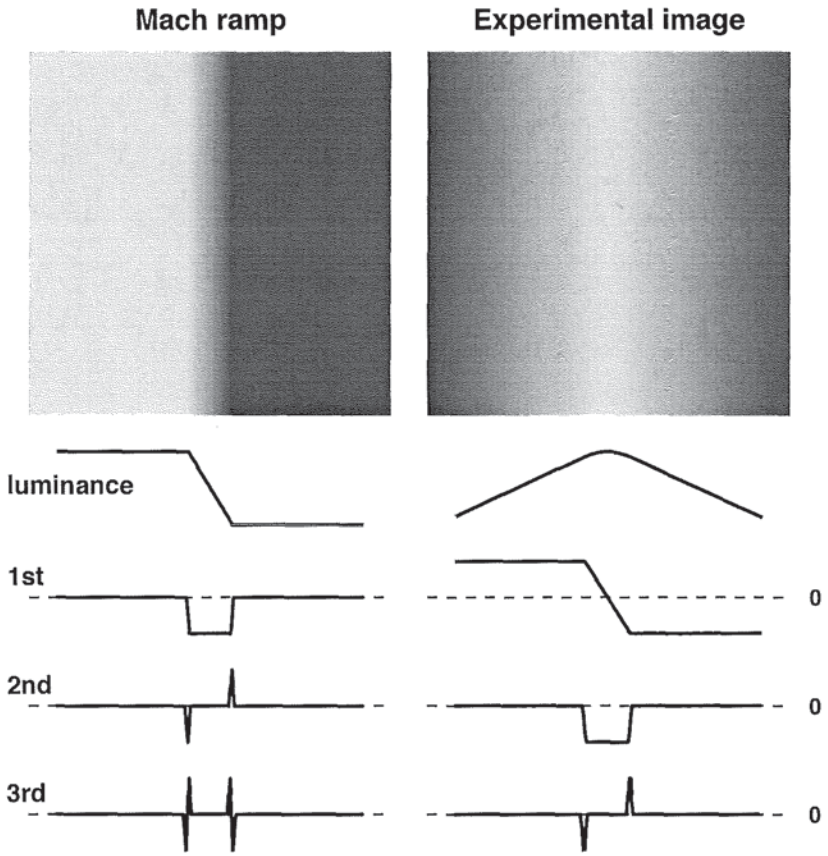

Figure 3.4: Mach ramp and experimental images with luminance waveforms and their first three derivatives.

The N3+ model (or indeed any linear model) must make identical predictions for the locations of the edges in both peak-central and trough-central conditions. Just the polarity of the prediction changes. Figure 3.3 shows that the two edges were perceived to be slightly further apart in the peak-central conditions than the trough-central conditions. This appears to be an example of the Helmholtz irradiation effect (Mather & Morgan, 1986), which is an illusory shift in edge position towards its dark side.

3.2 Experiment 1.2 - Mach edges, truncated

Figure 3.3 shows that there were small deviations of the model from the data. At low blurwidths model predictions are slightly too close to the image centre, and at high blurwidths the predictions are slightly too far from the image centre, particularly for SAW. The shift of the model away from the data at low blur-widths might be explained as follows.

The 3rd derivative model is 1-dimensional. It takes into account only (in this application) the horizontal contrast variation, as the experimental images were also 1-dimensional. Nevertheless, the images did have vertical, unchanging content. Although the observers were instructed to fixate on the images horizontal mid-line, they were free to use the full vertical extent of the image. The scale of the

smallest operator is thought to increase with retinal eccentricity (Croner & Kaplan, 1995; Poirier & Gurnsey, 2005). This may effectively increase the stimulus blur, emulating an increase in blurwidth. Furthermore, the scale-spce model output below (figure 3.5) shows that as scale increases, peak responses move outward. So if the observers were attending to the peripheral image content, the visual analysis might be pushed towards larger scales, which would push perceived edge positions away from the centre.

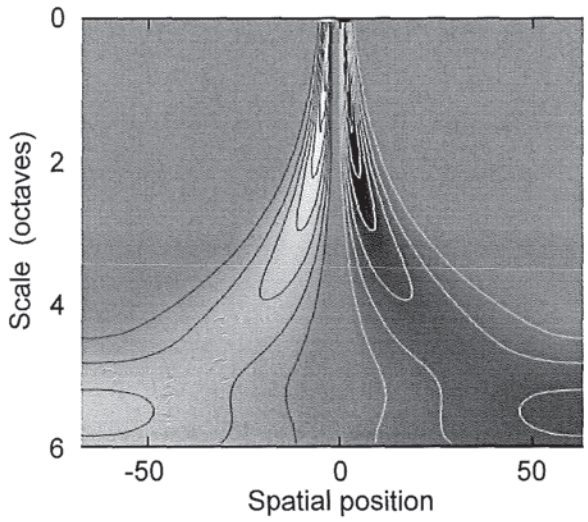

Figure 3.5: Scale-space map produced by the 3rd derivative model using an image from experiment 1.1. Note that the peaks move outwards as scale increases.

To test this explanation, the image height was decreased. This manipulation prevents vertical peripheral viewing (and therefore blurring) of the Mach edge image, and the truncated image contains less low frequency information. So it was anticipated that the output of the broader scale operators might be reduced, thus shifting the percept to back smaller scales, and causing the low blur-width edges to shift centrally.

3.2.1 Method

The method was the same as experiment 1 except for the following changes:

a. The image height was reduced to 8, 16 or 32 pixels (still 256 pixels wide).

b. Since there was little or no effect of contrast in experiment 1.1, only one contrast level was used here (0.4).

c. Only four blur-widths were used (1, 4, 16 and 64 pixels).

d. Marker-spots were 8 pixels above and below the image borders.

Examples of three trough central, blurwidth 16 experimental images are shown in figure 3.6. The three panels show image heights of 8, 16 and 32 pixels, on a mid-grey background. As in the previous experiment, the mid-grey background filled the entire screen.

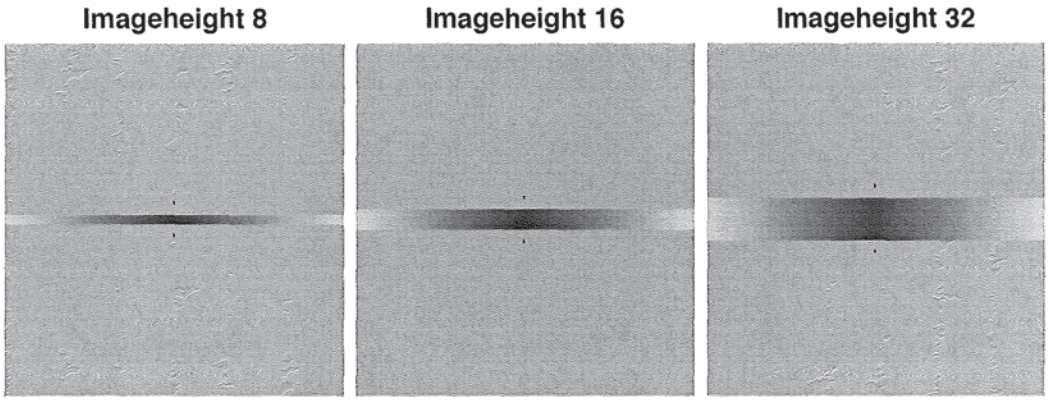

Figure 3.6: Examples of the three image heights used in experiment 1.2. Each panel shows a trough central, blurwidth 16 image of height 8, 16 or 32 pixels on a mid-grey background, with a central marker.

3.2.2 Results

The top row of figure 3.7 shows plots of perceived edge position against image blurwidth, averaged across all 3 observers. The contrast 0.4 data from experiment 1.1 are shown in the lower row, for comparison. Left column is from peak-central waveforms and right column is from trough-central waveforms. Each symbol represents the mean of 27 trials. The solid lines show predictions of the N3+ model. These predictions are identical in both rows, as the current version of the model is 1-dimensional, so cannot account for image height. The dotted lines in the upper plots show the predictions of a version of the N3+ model that has been extended to account for image height, and is described in the discussion. The shade of the symbols and lines denotes edge polarity (DL or LD). The similarity of the data in both rows of figure 3.7 shows that the pattern of data in this experiment is very similar to that of experiment 1, which used full height images. As blurwidth increases, so the marked edge positions spread symmetrically from about ± 6 pixels (for blurwidth 1) to about ± 25 pixels (for blurwidth 64). In figure 3.7, symbol shape denotes image height. For each blurwidth and edge polarity, the symbols for each height generally overlap. Error-bars ($\pm$1se) represent inter-observer variation (n=3) and are plotted behind symbols. Although there are more error-bars visible in the plots for this experiment than in the plots for experiment 1.1, they are still small. This supports the reliability of Mach edge position.

In order to examine the effect of image truncation in more detail, the difference between the data of experiment 1.1 and 1.2 is plotted in figure 3.8 (i.e. the data from the bottom row of figure 3.7 subtracted from the top row). The magnitude of the edge shift away from image centre, due to image truncation, is plotted as a function of blurwidth. The left panel shows the peak central data, where most of the image height 16 and 32 data are within 1 se of zero. However, half of the image height 8 data are above 1 se. The right panel shows the trough central data, where all of the image height 8 data and half of the image height 16 data are more than 1 se above zero. The direction of shifted edge position is away from image centre for the majority of data.

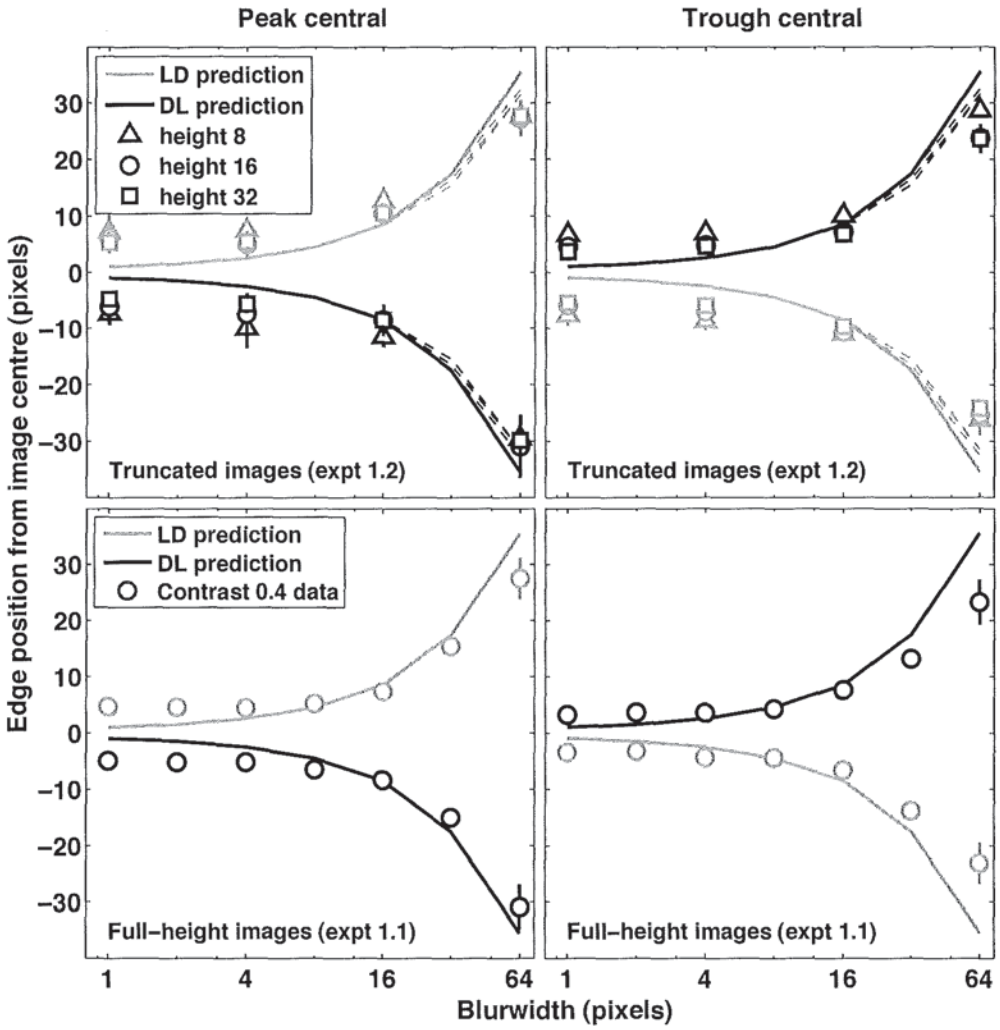

Figure 3.7: Mean across observers (symbols) and N3+ model predictions (lines) for positions of dark-to-light (DL) and light-to-dark (LD) edges in 1D blurred triangle-wave images. Left column is from peak-central waveforms and right column is from trough-central waveforms. Top row shows data from experiment 1.2, which used truncated images. For comparison, the bottom row shows data from experiment 1.1, which used full-size images. N3+ model predictions are plotted in both rows as solid lines. Dashed lines in the upper plots show predictions of the updated N3+ model. Errorbars ($\pm$1se) show inter-observer variation (n = 3).

3.2.3 Discussion

The marked edge positions are very similar in both experiments, which suggests that the effect of image truncation is small. This conclusion is further supported by the minimal effect of image height, which is illustrated by the high degree of overlap between the three symbols at each blurwidth and edge polarity. A closer examination of this effect (figure 3.8), suggests that truncation to a height of 8 pixels shifts perceived edge locations about 4 pixels away from image centre. The data clearly do not demonstrate the expected shift towards image centre. Therefore, the shift of the data away from model

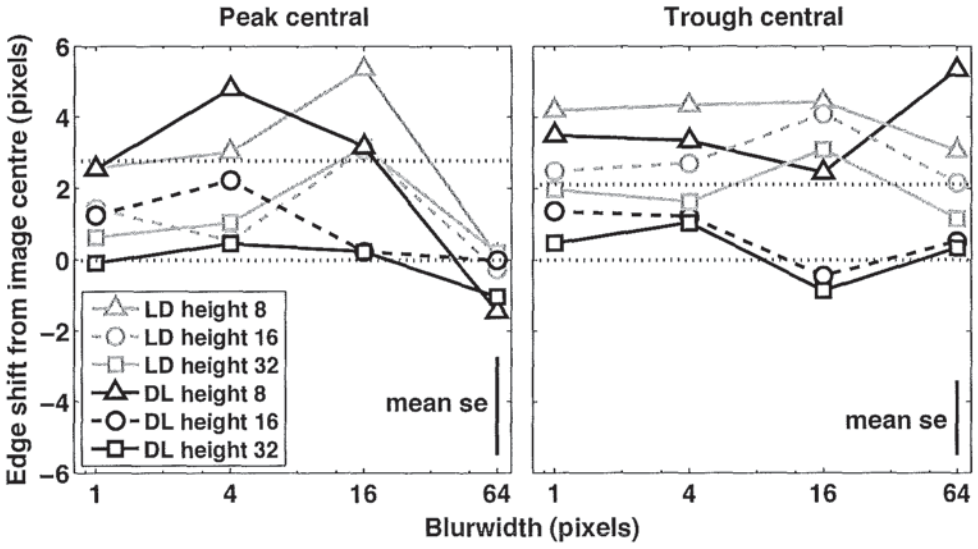

Figure 3.8: Effect of image truncation. Shift in perceived edge position away from image centre, averaged across 3 observers, plotted as a function of blurwidth. The mean se bars show 1 standard error and represent inter-observer variation (n = 3) and are calculated from the mean of the standard errors of all data points in the panel. The upper dotted line is 1 se above zero.

predictions at low blurwidths observed in experiment 1.1 is unlikely to be due to the explanation given above regarding relatively peripheral viewing.

Previous research has demonstrated that perceived edge locations are shifted away from their mid-luminance point towards their dark side (Georgeson & Freeman, 1997; Westhiemer, 2007; Watt & Morgan, 1983) by an amount proportional to their blur (Mather & Morgan, 1986; Morgan et al., 1984). Also, a sufficient shortening of edge length can reduce its percieved blur (Georgeson et al., 2007). Taken together, these findings may lead one to hypothesise that in the present experiment, image truncation would appear to sharpen the edge, which would then shift its perceived location towards its light side. However, most of the peak central data do not support this prediction, as the edge shift is generally towards the dark side of each edge (except for blurwidth 64). Additionally, the data do not show a trend of increasing shift as blurwidth increases. So this hypothesis does not account for the present data.

The perceived edge shift is similar in both image polarities, so is unlikely to be related to Mach edge luminance (which was higher in the peak central conditions than in the trough central conditions). It is not clear why the shift occurs.

Do the N3+ model predictions match the present data? Although the model is multi-scale and generates scale-space maps of variable height, its current version is inherently 1-dimensional in the spatial domain. So it makes identical predictions for images of any height. In order to account for the present data, it was updated to incorporate image truncation, by the method described below.

The filtering is performed in two stages in the model. The first filter computes the 1st derivative of the luminance profile and the resulting output is half-wave rectified and passed to the second filter. This

computes the 2nd derivative of its input, and then a second stage of half-wave rectification is applied. The combination of both filters obtains the (non-linear) 3rd derivative of the luminance profile. In order to account for image height, at both filtering stages, the derivative filters of every scale were attenuated by multiplication with

$$2\Phi\left(\frac{h}{2\sigma}\right) - 1 \tag{3.1}$$

where h is the image height (pixels), σ is the scale of the filter (pixels) and Φ is the standard normal integral. Equation 3.1 is based on the assumption that the filters have a 2-dimensional circular receptive field in space. Some cortical cells in cat (Freeman & Ohzawa, 1990) and primate (De Monasterio & Gouras, 1975) are known to have this configuration.

Where image truncation impinges on a filter's receptive field, the amplitude of its output is reduced proportionally. For a given image height, the net effect across filter scale is a smooth and progressively greater amount of attenuation as filter scale increases.

The updated N3+ model predictions are shown as thin dashed lines in figure 3.7. Separate lines have been plotted for each of the 3 image heights, but they overlap for almost all of their length. At blurwidth 64, two dashed lines can be distinguished; the most central pair are from an image height of 8 pixels, and the more peripheral pair are from an image height of 32 pixels. Overall, the updated N3+ model predictions are identical to those of the original version for all blurwidths below 32 pixels. But at blurwidth 64, the updated model provides a slightly better fit to the data than the original version of the model. This improved fit may indicate that the original model places slightly too much weight on the large scale filters.

To conclude, since image truncation did not move apparent edge location centrally, the shift of the data away from model predictions at low blurwidths observed in experiment 1.1 is unlikely to be due to the explanation given above regarding relatively peripheral viewing. An alternative explanation will be considered below, immediately after a control experiment.

3.2.4 Experiment 1.3 - A control for marker-spot position

Introduction

In experiment 1.1, the marker-spots were on the image, but in experiment 1.2 they were placed outside the image border, on the background. This difference might be relevant if the marker-spots affected Mach edge perception. They might have a masking effect that extends beyond their boundary, as a central fixation point has been shown to slightly increase contrast detection thresholds for patches of sine-wave gratings (Summers & Meese, 2009). Also, each marker-spot was a thin bar, and the idea that bars and edges may mutually inhibit each other has some support (Ratliff, 1984; Tolhurst, 1972b). This masking might affect Mach edge location in an unpredictable manner, as the marker was moved across each image by the observer. Furthermore, eye movements were not monitored during the experiments. So although observers were instructed to fixate at the mid-point of the two spots, they may have occasionally fixated on the upper or lower spot whilst aligning the marker with a Mach edge.

Method

In order to examine the effect of marker-spot position, a control experiment repeated the method of experiment 1.2, with the alteration that the marker was a single marker-spot placed on the image's

horizontal mid-line, and observers were instructed to fixate on it. This manipulation was expected to increase any masking effect. This might particularly affect the images of 8 pixels height, as the marker was 3 pixels high itself and would therefore obscure 3/8 of the length of each Mach edge.

Results

Data were collected from the same three observers that took part in the previous two experiments. There was little inter-observer variation, so data were averaged across observers, and plotted against blurwidth. These values are shown in the top row of figure 3.9 (symbols) with predictions from the updated N3+ model (lines). The marked edge positions generally follow the previously observed trend of a gradual shift away from the image centre as blurwidth increases, with minimal difference between

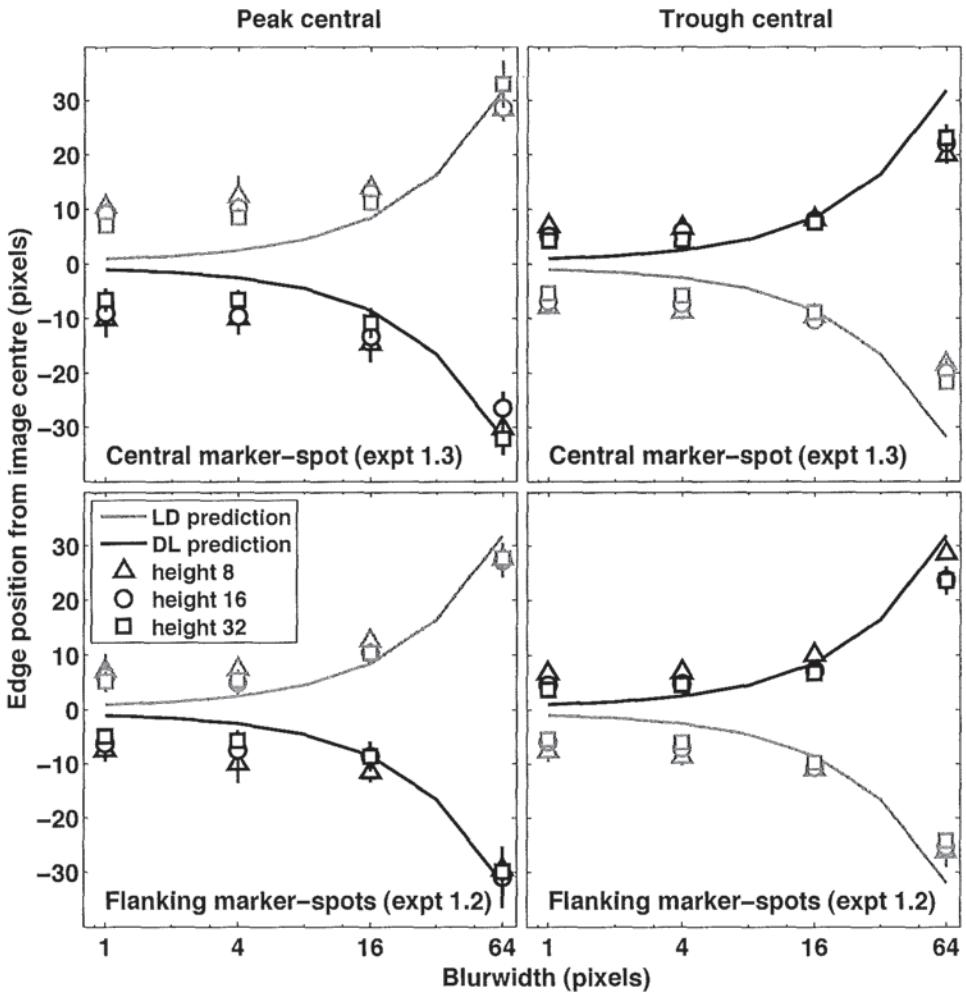

Figure 3.9: Mean across observers (symbols) and updated N3+ model predictions (lines) for positions of DL and LD edges in truncated 1D blurred triangle-wave images. Left column is from peak-central waveforms and right column is from trough-central waveforms. Top row shows data from experiment 1.3, which used a central marker-spot on the image. For comparison, the bottom row shows data from experiment 1.2, which used a pair of marker-spots on the background. Errorbars (±1se) show inter-observer variation (n = 3).

the data of different image heights. For comparison, the bottom row of figure 3.9 shows data from experiment 1.2, which used a pair of marker-spots on the background. The difference between the data of each row (experiment 1.3 data minus experiment 1.2 data) is plotted in figure 3.10 as a shift away from image centre. The right panel shows the trough central data, which demonstrates little or no

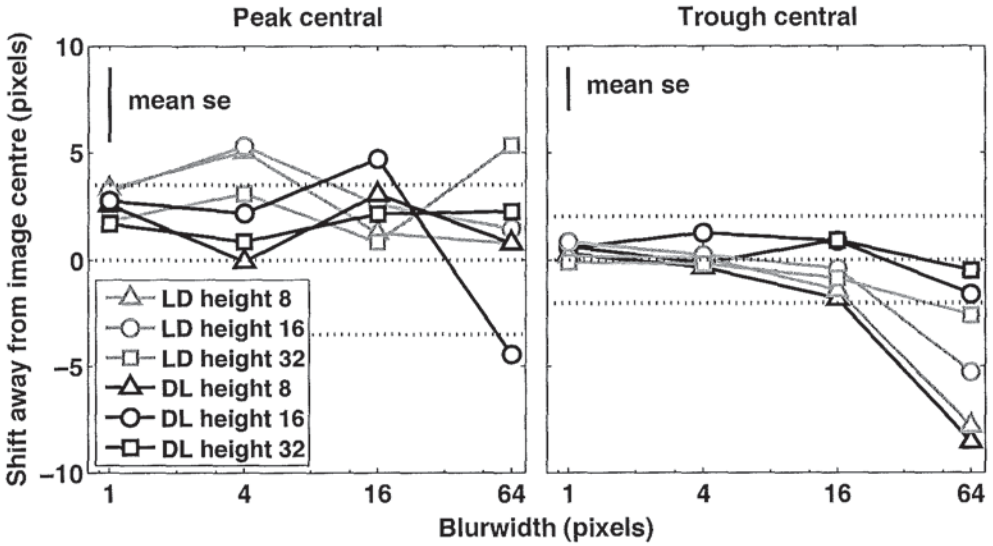

Figure 3.10: Effect of marker-spot configuration. Shift in edge position away from image centre when the marker-spot configuration was changed, plotted as a function of blurwidth. The marker was changed from a vertical pair of dots on the background to a single central dot on the image's horizontal mid-line. The mean se bars show 1 standard error and the upper and lower dotted lines are 1se from zero.

shift of edge position for blurwidths 1 to 16. For blurwidth 64, the edge positions are shifted towards the image centre, and the magnitude of the shift is large (about 7 pixels) for image height 8, and small (about 1.5 pixels) for image height 32. The left panel shows the peak central data, which are more noisy than the trough central data, but nevertheless appear to show a general shift of about 2.5 pixels away from the image centre for all blurwidths.

Discussion

The similarity of the data in the rows of figure 3.9 shows that the change in marker-spot configuration had only a small effect. When the effect was examined in more detail (figure 3.10), the peak central data show a small shift away from image centre, suggesting an effect of marker configuration. However, since most of the data lie within 1 mean se of zero, such a conclusion can only be tentative. An absence of shift is also evident in most of the trough central data. But for blurwidth 64, there is a shift towards image centre, which is greater than 1 mean se for the LD data of image heights 16 and 32, and for LD and DL data of image height 8. These 4 data points do hint at an effect of marker configuration, for this blurwidth, which is stronger as image height decreases. This shift is towards the dark side of the edge, which is also the direction observed in the peak central data. Nevertheless, since the vast majority of the data fall within 1se of zero, no overall effect of marker-spot configuration can be concluded. This suggests that the placement of the marker-spots onto the background in experiment 1.2 had little impact.

3.3 Add optical and neural blur to the model

An alternative explanation for the shift of the model away from the data at low blur-widths is that the model omits the eye's optical and neural blur. This will have maximum effect at high resolutions. The amount of blurring induced by the eye is often expressed as a 'point spread function' (PSF), which describes the amount that a point source of light is blurred across space. Campbell & Green (1965) used an ingenious technique to estimate it. They used a laser to project interference fringes directly on the eye's retina, which formed a sinusoidal grating pattern that was uncontaminated by the eye's optics. The contrast threshold for detection of the fringes was obtained at various spatial frequencies. The contrast thresholds for detection of external gratings of the same spatial frequencies were also measured for the same observers. The ratio of these threshold values to those obtained with interference fringes determined the modulation transfer function (MTF) of the eye. Although the MTF could then be subjected to an inverse Fourier transform to produce a PSF for the eye, this manipulation was not performed by Campbell & Green. However, it was performed in a subsequent study (Campbell & Gubisch, 1966), which used a 'double pass' method, described below. Functions were obtained for several pupil diameters. For a 3mm pupil, the PSF can be approximated by a Gaussian of $\sigma = 0.7$min (fit performed by eye with PSF in figure 10, Campbell & Gubisch (1966)).

This interference fringe method was compared to an alternative 'double pass' method by Williams et al. (1994), whereby a point of light is focussed on the retina, and the light that is reflected from the retina is used to compute the MTF. Williams *et al* found that the double-pass method produced MTF's that were similar to but slightly lower than those of the interferometric method. They were also slightly lower than those obtained by Campbell & Gubisch, and this was attributed to methodological improvements. The inverse Fourier transform of a lower MTF is a broader PSF, which can be conceptualised as a less sensitive eye scattering more light. The PSF obtained by Williams *et al* can be approximated by a Gaussian of $\sigma = 0.8$ to 1min (fit performed by eye with PSF in figure 6, Williams et al. (1994)).

The N3+ model was updated to include a Gaussian smoothing pre-filter. Model predictions where the pre-filter's $\sigma = 1.25$mins are plotted with data from experiment 1.1 in figure 3.11. It is now a strikingly better fit at low blur-widths. Model predictions were also obtained with other σ values between 0.75 and 2 (at 0.25 increments), but these all resulted in larger RMS errors between predictions and data. The selected value of 1.25mins may represent the combination of optical blur and the scale of smallest filter used by observers.

3.4 Experiment 1.4 - Resolving power of revised model is tested with step edges

3.4.1 Introduction and method

The N3+ model now incorporates a factor that could represent optical and/or neural blur, in the form of a Gaussian smoothing pre-filter. Since this blur is applied to all images, it reduces the model's ability to resolve fine detail. This loss of resolution may be excessive, as the amount of blur in the model ($\sigma = 1.25$) is greater than the amount of blur estimated from the PSFs of the above studies ($\sigma = 0.7$ to 1). The next experiment aims to check that a blur of 1.25mins is not excessive, by testing

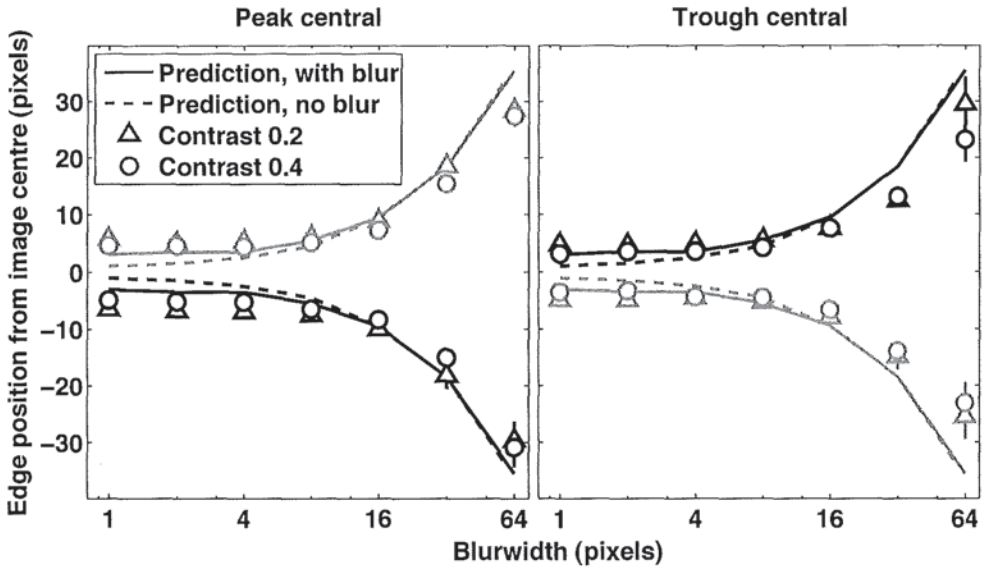

Figure 3.11: N3+ model predictions with and without Gaussian blur (σ = 1.25).

the resolving power of the model with experimental data that uses images containing sharp step edges.

Experiment 1.1 was rerun, but now with a different set of images and fewer sessions. The blurred triangle-wave images were replaced with images consisting of a vertical sharp-edged bar on a mid-grey background. Bar widths were determined by the edge separation data from experiment 1.1, averaged across observers. They ranged from 2 to 69 pixels (1.9 to 64.7 mins) and were either light or dark. All other details were as experiment 1.1. The first observer (SAW) completed 2 sessions whose data were almost identical. The other two observers completed only one session each.

3.4.2 Results and conclusion

Data were almost identical for both contrast levels and all 3 observers, so means are shown in figure 3.12. Data and model predictions still fit well, even at the smaller barwidths. This validates the addition of the blur, and suggests that 1.25 minutes is appropriate. There is again some evidence of the Helmholtz irradiation effect, in that the marked edges are further apart for light bars than for dark bars.

3.5 Summary and discussion of experiments 1.1 to 1.4

The majority of previous derivative-based approaches to edge detection have relied upon peaks in the 1st derivative of the image luminance profile (Canny, 1986) or zero-crossings in the 2nd derivative (Marr et al., 1979). The experimental results reported here strongly suggest that these approaches provide an incomplete account of edge detection, and that the 3rd derivative has an important role to play. The experimental luminance profiles were blurred triangle-waves that contained no 1st derivative peaks or 2nd derivative zero-crossings (except at a very large scale). But they did contain peaks and troughs in the 3rd derivative. Using a feature-marking paradigm, 3 observers marked the position and

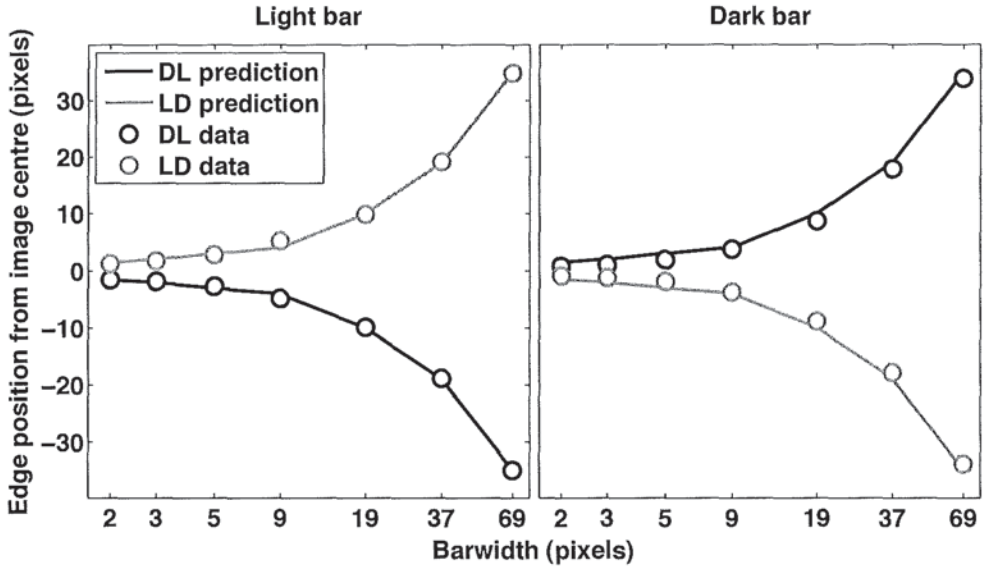

Figure 3.12: Perceived edge position, averaged across observer and contrast level (symbols) and N3+ model predictions with blur (lines) as a function of step edge bar width. Each symbol represents the mean of 24 trials. Error bars (±1se) are much smaller than symbol size (not shown). The model fit is good.

polarity of all perceived edges. They consistently marked an edge at, or close to, each 3rd derivative peak and trough.

Interestingly, these edge locations correspond to no striking change in the luminance profile at all (see figure 3.2). This is in contrast to a blurred step edge, whose location is clearly indicated by the maximum steepness of the luminance profile (figure 3.16). The edges marked by observers were named 'Mach edges' because they are analogous to Mach bands. Mach bands are apparent bars where there is no peak in the luminance profile, and Mach edges are apparent edges where there is no peak in the 1st derivative. The Mach edges seen here are a new phenomenon, and their existence demonstrates that not all visible edges arise from peaks in luminance gradient. They therefore pose a serious challenge to approaches based on 1st derivative peaks. They also show that not all edges are zero-crossings in the 2nd derivative, as a step change that merely meets zero will generate an edge percept (see figure 3.2).

Whilst previous researchers have suggested that the 3rd derivative informs blur perception (Georgeson et al., 2007; May & Georgeson, 2007a,b), this is the first report of a series of experiments that were designed to test its use for edge localisation. At first sight, the 3rd derivative may appear unpromising as an edge detector, as it generates not only a peak at blurred step edge, but also a pair of flanking troughs, thus signalling 2 spurious edges of opposite polarity to the central 'correct' edge (figures 1.11 and 3.16). However, the addition of 2 stages of half-wave rectification to the computation of the 3rd derivative is able to overcome the problem of spurious edges (described in chapter 1, section 1.10.2). The N3+ model is based on this approach and was shown here to provide a good fit to the data. At a very large scale, a 1st derivative model predicts edges at ±64 pixels for images of all blurwidths, but they were not marked by observers.

The N3+ model predictions were slightly narrower than the data at low blurwidths. A hypothesis was posited to account for the peripheral shift in marked edge positions, which was based on the

use of relatively peripheral larger scale filters. However, the lack of a central shift in perceived edge location when images were truncated (experiment 1.2) suggested that this hypothesis was unlikely to be correct. A better explanation for the small discrepancy between model and data at low blurwidths is that the model required a factor that could represent optical and/or neural blur. When this factor was added, as a Gaussian blurring pre-filter, the resulting model predictions for low blurwidths were shifted peripherally, and provided a strikingly better fit to the data.

In order to check that the addition of the Gaussian blur to the N3+ model had not unduly reduced its ability to resolve fine detail, the feature-marking method was applied to images of sharp-edged bars. The revised model was still able to provide a good fit to marked edge positions, even at barwidths as low as 2 pixels. This demonstrated that the amount of added blur ($\sigma = 1.25$mins) in the pre-filter was not excessive.

All of the results pose a challenge for the 1st and 2nd derivative approaches and support the use of the 3rd derivative, which is a crucial component of the non-linear, scale-space N3+ model. This is not the only model, however, that uses a bank of filters at a range of scales to derive feature locations. The local energy model (Morrone & Burr, 1988), which was described in chapter 1, employs pairs of odd- and even-symmetric operators at a range of scales to predict the polarity and position of edges and bars. Although it differs from the N3+ model in the way that filter outputs are combined, some of its filter profiles approximate the shape of a (4-lobed) 3rd derivative operator. So it may produce predictions that are similar to the N3+ model.

An implementation of the local energy model was applied to the experiment 1.1 waveforms. They were filtered with four pairs of even and odd-symmetric filters of quadrature phase. Each filter amplitude spectrum was defined as a Gaussian function of logarithmic frequency, given by

$$A(f) = exp\left(-\frac{(ln\,(f/p))^2}{2\,(q.ln(2))^2}\right) \tag{3.2}$$

where p is the peak spatial frequency and q is the half-bandwidth in octaves at 0.6 height.

Following Morrone & Burr (1988), the values of p were 1, 5.7, 27.4 and 65.2 cycles per image (equating to 0.25, 1.42, 6.85 and 16.3 cycles per degree here), and the corresponding half-bandwidths were 1.2, 0.9, 0.73 and 0.57 octaves. The even filters were constructed from component cosine-waves, multiplied by the amplitude given by equation 3.2 and an additional shallow Gaussian that is designed to reduce ringing. The resulting filter was therefore defined as

$$F(x) = \sum_{f=1}^{f=n} A(f)cos(2\pi fx).exp\left(\frac{-f^2}{2\sigma^2}\right) \tag{3.3}$$

where n is the nyquist frequency and σ is half of the nyquist frequency.

Each matched odd filter was generated by the same method except that cosine was replaced with sine in equation 3.3 . The profiles of all filters are shown in figure 3.13. Features were identified by peaks in the energy profile. The type of feature was determined by the amplitudes of the even- and odd-symmetric filter outputs. An edge was predicted where the odd-symmetric filter amplitude was greater than the even-symmetric filter amplitude at the location of the energy peak. A bar was predicted if the even-symmetric filter amplitude was greater than the odd-symmetric filter amplitude. Since there were 4 pairs of filters, the model generated separate predictions for each filter-pair. An edge prediction

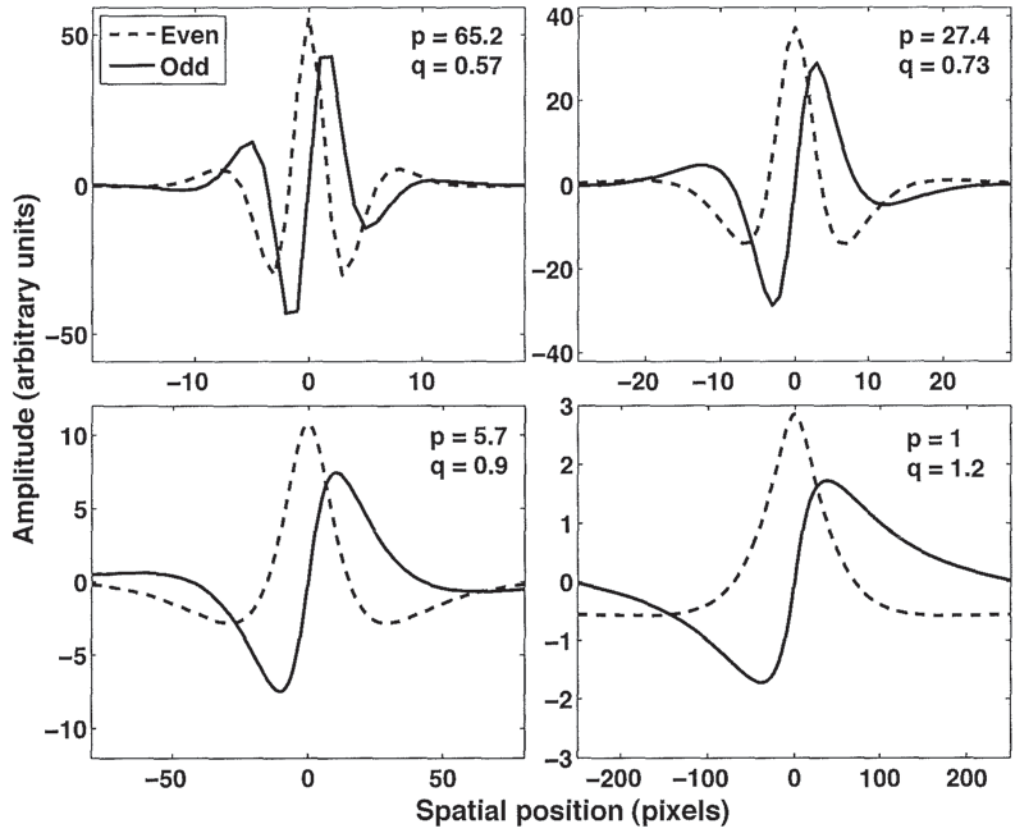

Figure 3.13: Energy model filters.

from one or more filter-pairs was assumed to indicate the presence of an edge. Where filter-pairs of different scale predicted a common edge, its predicted location was usually at the same location (to the nearest pixel). However, the largest filter-pair occasionally predicted a location that was shifted by a few pixels. Therefore, the rule adopted for combination across scale, was to use the predicted location provided by the smallest filter-pair that provided edge predictions.

This version of the energy model predicted a central light bar for all peak central waveforms, with flanking edges for all waveforms except the two smallest blurwidths. However, edges were resolved for all waveforms when two extra filters pairs were added that were scaled down versions of those shown in the top left panel of figure 3.13 (p = 130.4 and 220; q = 0.57 for both) and were maximally responsive at 32.6 and 55cpd. Where an edge was predicted by more than one filter-pair, its was at a common location.

A plot of the experiment 1 data with predicted edge locations is shown in figure 3.14. The energy model makes the same predictions as the 3rd derivative model for most blurwidths, despite its different approach. This high degree of overlap suggests a commonality between the models, which might be explained by the shape of the energy model filter profiles. The smaller odd filters (figure 3.13, top panels) resemble 3rd derivative operators (4 lobes).

In order to examine whether this 4-lobed shape was critical for model predictions, the blurwidth 8 image set was convolved with a set of energy filters whose peak spatial frequencies were the same

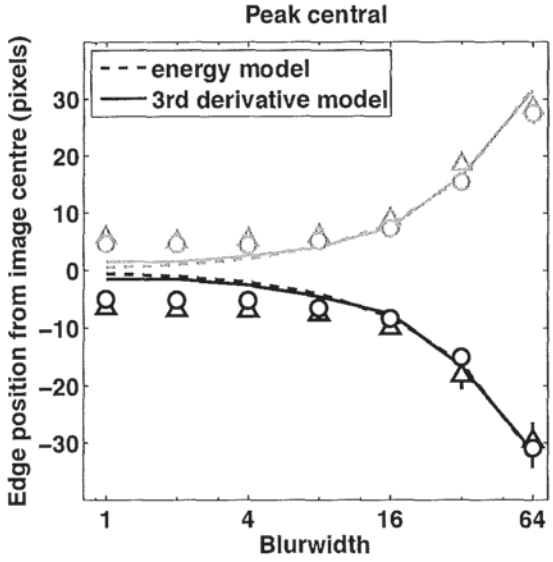

Figure 3.14: Mach edge data from experiment 1.1 with predictions from the energy and single-scale models.

as used above (0.25 to 55cpd), and whose half-bandwidth at 0.6 height ranged from 0.2 to 1.8 in 0.2 increments. For each peak spatial frequency, bandwidths were recorded that produced accurate predictions; i.e. a pair of edges whose polarity and location were close to the data, and no spurious features. The 16.3cpd filter produced a higher number of accurate predictions than any other. The lowest and highest recorded bandwidths produced filters that had 6 and 2 lobes respectively. Although this shows that a 4-lobed shape is not critical for the energy model, the mid-range recorded bandwidth was 4-lobed, which suggests that that shape may be optimal. Thus the success of the energy model does appear to be related to its use of filters whose shape is approximated by 3rd derivative operators.

The derivation of feature locations in the energy model is dependant on energy peaks that are generated by the combination of the outputs of odd and even filters. This combination necessarily loses information, which means that the model fails to predict features that observers can see. For example, it predicted no features at all in a sine-wave grating (Morrone & Burr, 1988) and only 2 of the 6 marked features in each period of a grating composed of a fundamental and 3rd harmonic (Georgeson & Freeman, 1997). It also incorrectly predicted no shift in feature position when phase was altered in the components of a blurred square-wave image (Hesse & Georgeson, 2005). So although it correctly predicted Mach edges here, its application to feature-detection appears to be somewhat limited.

To return to the derivative approach, is a full scale-space N3+ model required to account for the results, or will a more parsimonious, single-scale version suffice? In order to examine this question, a single-scale version was applied to the experimental data, using derivative operators of a fixed fine scale. This generated the predictions shown in figure 3.15, which are superimposed on the N3+ model predictions (without pre-filter) for comparison. Both models fit equally well at most blurwidths. The single-scale version fits the high blurwidth data slightly better than the N3+ model, but it cannot generate the scale (edge blur) predictions that the N3+ model provides. Nevertheless, the single-scale model is clearly an acceptable simplification of the N3+ model for Mach edge position.

The experimental images were specifically designed to separate the locations of the 1st and 3rd

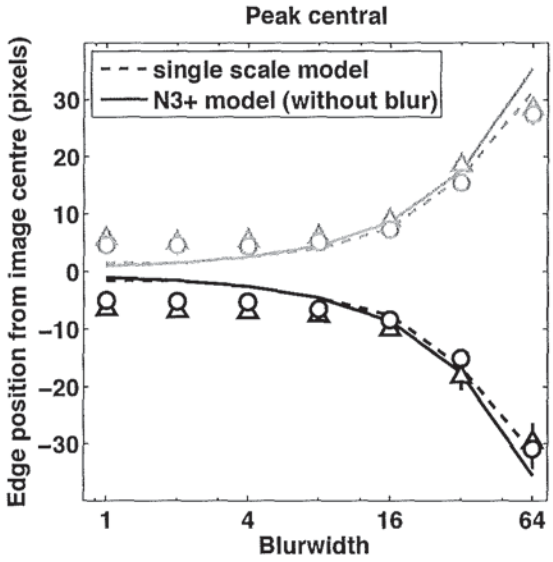

Figure 3.15: Feature-marking data from experiment 1.1 with predictions from the N3+ and single-scale models. Both models fit equally well at most blurwidths.

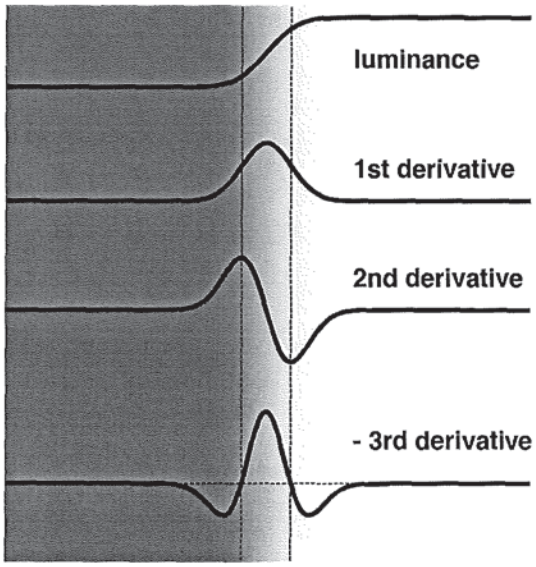

Figure 3.16: A blurred edge, its luminance profile and first 3 derivatives. The 3rd derivative is inverted and its central peak is narrower than the 1st derivative peak

derivative extrema. But although 3rd derivative extrema can exist where there is no 1st derivative peak, the reverse is not true. From basic calculus, all 1st derivative peaks must have a corresponding 3rd derivative extremum. So we cannot exclude the possibility that the 3rd derivative is the primary approach for the majority of edge localisation and identification in human vision. However the 3rd derivative of a blurred step edge provides no additional information above that provided by the 1st derivative, concerning edge position, polarity or blur. So why should vision have evolved to use the 3rd

derivative? One answer proposed by Georgeson et al. (2007) is that it may provide higher resolution. This is illustrated in figure 3.16, which shows that the -3rd derivative peak is narrower than the 1st derivative peak. So we would expect that a 3rd derivative model will resolve two neighbouring same-polarity edges better than a 1st derivative model. This resolving power is not addressed in this thesis, but is a viable area for future research.

3.6 A critical test of the scale-space model using blurred triangle-waves with added linear ramp

3.6.1 Introduction

Experiment 1 demonstrated that Mach edges were reliably perceived at the locations of 3rd derivative peaks in blurred triangle-wave stimuli, despite the absence of 1st derivative peaks and 2nd derivative zero-crossings. This supported the use of 3rd derivative filters to detect edges in this image set, and the data were well-fitted by a non-linear scale-space 3rd derivative (N3+) model and a single-scale simplification of it. Excitingly, Mach edges are a new phenomenon that have not been observed previously. However, as with any newly-discovered phenomenon, its ability to generalise to other image sets is currently unknown. In order to test its generalisability, Mach edge perception was examined with a new set of images.

A critical property of any image that tests the 3rd derivative approach against the 1st and 2nd derivative approaches is that peaks in the 3rd derivative of its luminance waveform should not coincide with peaks in the 1st derivative or zero-crossings in the 2nd derivative. This was achieved with the blurred triangle-wave images used in experiment 1. So a simple manipulation of those images was considered. The addition of a constant to a waveform does not alter its lower derivatives because the derivative of a constant is zero. So a (modest) fixed increase or decrease in luminance is applied to the blurred triangle-waves of experiment 1 would not be expected to diminish Mach edge perception. However such a manipulation is unlikely to be informative in practice, as the visual system is likely to adapt to the new luminance levels. Furthermore, the effect of contrast was shown to be very small in experiment 1.

But the addition of various amounts of constant to the 1st derivative *would* yield an interesting image set, because although the 3rd derivative is unaltered, the predictions of the non-linear 3rd derivative models (both scale-space and single-scale versions) are dependant upon the magnitude of the constant. This is illustrated in figure 3.17, where various amounts of positive vertical shift have been added to the 1st derivative of the blurred triangle-waves used in experiment 1. Since the 1st derivative is the gradient of the luminance profile, its vertical shift is equivalent to the addition of a linear ramp to the luminance profile.

Figure 3.17 shows that the ramp has no effect on the second or higher derivatives. But the N3+ and single-scale models predict two edges for ramp gradients of less than 1 (figure 3.17, 2 left plots) and one edge for steeper added ramp gradients (3 right plots). This is because, as the added ramp is increased, the 1st derivative becomes entirely positive. Hence the LD channel is silenced, because its first rectifier makes it responsive only to negative gradients (LD edges).

The gradient offset ± 1 waveform deserves particular mention (figure 3.17, central plot), as here the

image is $\frac{1}{2}$ cycle of a blurred trapeziodal-wave, which is expected to cause Mach bands (Ratliff, 1965). The bands may appear to consist of two adjacent edges of opposite polarity, yet the model predicts only one edge here. Therefore this image is expected to strenuously test the model.

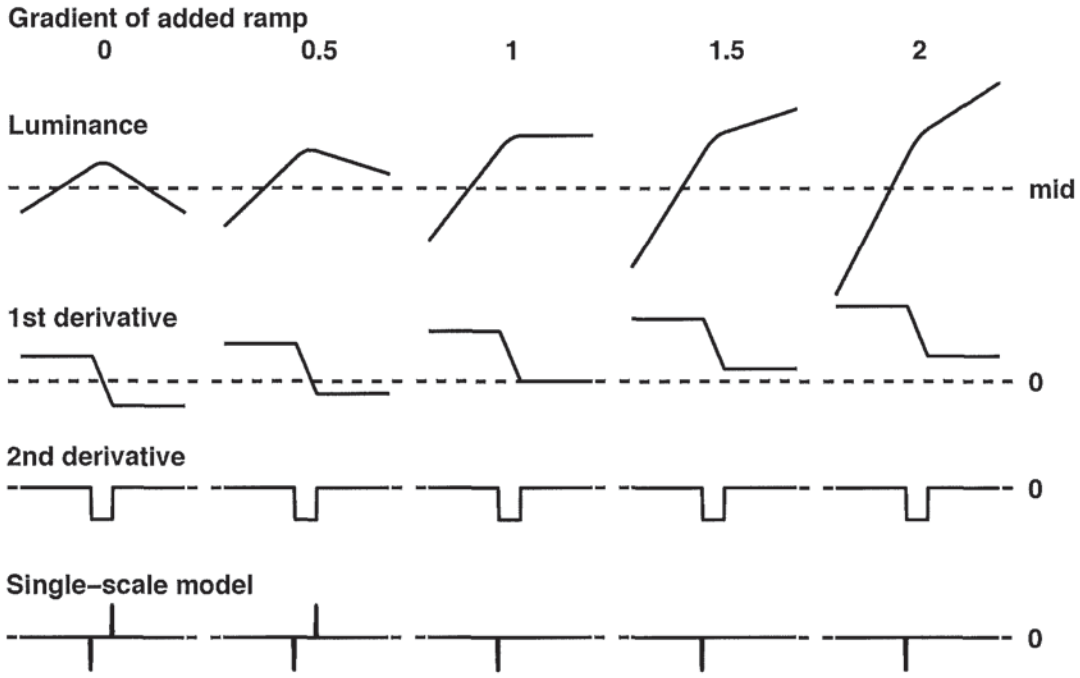

Figure 3.17: Experiment 2 luminance waveforms, their first 2 derivatives and the single-scale non-linear 3rd derivative model output. Waveforms are derived from a blurred triangle-wave (top left) by adding a luminance ramp of various gradients. Only positive added gradients are shown here.

3.6.2 Stimuli

The basic stimulus was one period of a triangle-wave. It was blurred by a rectangular function 8 pixels wide. The gradients of its rise and fall were defined as +1 and -1. A linear luminance ramp was added to form the stimulus luminance profile. The ramp gradient was -2 to +2, in 0.5 increments. These 'peak central' luminance profiles were all inverted in opposite phase ('trough central') conditions. The vertical 1-D image was 256 by 256 pixels. It subtended 4.27 degrees at the viewing distance of 123.4cm. Thus 1 pixel subtended 1 minute. The image was surrounded by a full-screen mid-grey of luminance $40.7cd/m^2$. The image mid-level luminance was set to equal the background. The addition of the luminance ramp necessarily increases the contrast of the waveform as its gradient increases. So the Michelson contrast ranged from 0.25 (gradient 0) to 0.8 (gradient ± 2).

3.7 Experiment 2.1 - Mach edge detection is tested with a Yes-No paradigm

3.7.1 Method

A single interval yes-no procedure was adopted. Images were presented for 300ms, temporally flanked by a mid-grey screens with a central fixation marker. The observer indicated whether 1 or 2 edges were seen near the centre of the image; he had unlimited time to respond. The observers response initiated the next trial. The inter-trial interval was at least 1 second. Each of the 18 conditions (9 images, 2 phases) was presented 10 times in a randomised order within one session, which took about 5 minutes to complete. Four psychophysically experienced observers each completed 7 sessions, but the first session was discarded as practice. Data totalling 60 trials were collected per condition. Only 2 of the 4 observers (SAW and MAG) were aware of the purpose of the experiment.

3.7.2 Results

The Proportion of '2 edges' responses for each condition is shown in figure 3.18. The top 4 rows show the data from each observer. The average across observers is shown in the bottom row. Left and right columns represent the peak central and trough central phases respectively. Thick dashed lines show single-scale non-linear 3rd derivative model predictions. The single-scale model was used for simplicity and because it provided almost the same predictions as the scale-space model for the Mach edge image set in experiment 1. Error bars ($\pm 1se$), are calculated from the expected variance in binomial sampling, and are given by

$$se = \sqrt{p.\left(\frac{1-p}{n}\right)} \tag{3.4}$$

where p is the proportion of '2 edges' responses, and n is the number of trials per datum (60 here). Two edges were seen reliably with added gradients up to 0.5. This matched model predictions closely. With steeper added ramps, reports of two edges fell away as the model predicts, but more gradually than predicted. The model *cannot* predict 2 edges for offsets ± 1 to ± 2, because here the luminance gradient for each image is entirely of one sign (positive or negative, depending on the sign of the added ramp). To take the example of a positive added ramp, the first half-wave rectifier in the model blocks any response in the LD channel, as the 1st derivative filter that precedes it produces only negative values. So only the DL channel can respond. In order to provide a better fit to the data, the model was revised to allow a more gradual transition between the predictions of one edge and two edges, as described in the next section.

In order to test for a polarity effect, Pearsons X^2 statistic was applied in a comparison between the peak central and trough central data, separately for each observer. It identified a significant effect for observers MAG, DHB and RJS ($p < 0.01$). However, no general inference can be drawn, as the direction of the effect varies across observers, and the data averaged across observers indicate little or no polarity effect.

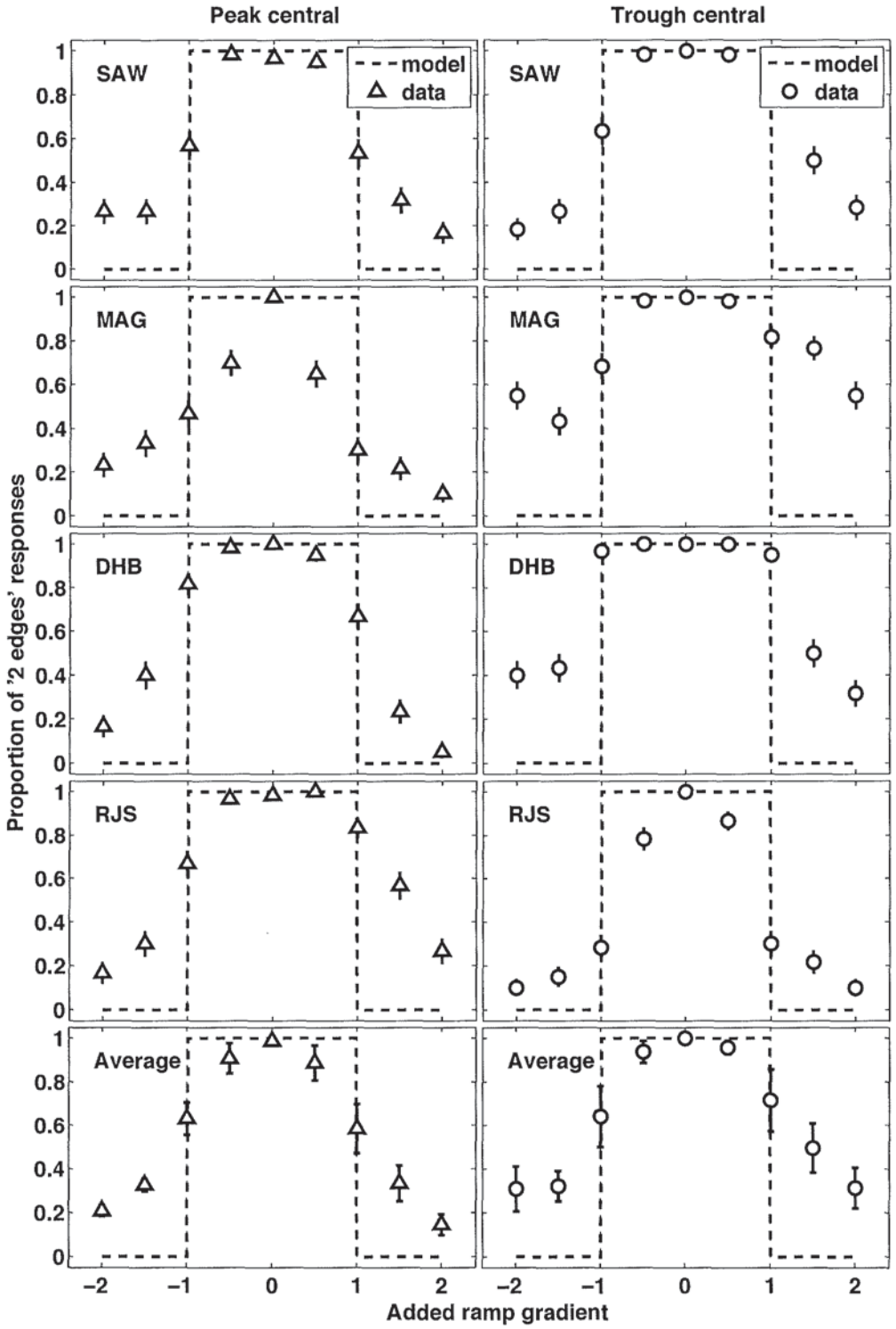

Figure 3.18: Proportion of '2 edges' responses (symbols) and single-scale model predictions (dashed lines) for 'peak central' and 'trough central' phases (left and right columns) of a blurwidth 8 triangle-wave with added ramp. Top 4 rows show data from each observer; error bars here ($\pm 1se$) are calculated from the expected variance in binomial sampling. The mean across observers is shown in the bottom row; error bars here are calculated from the inter-observer variation.

3.8 Model refinement - addition of a pre-filter and a noisy decision process

The stages in the N3+ model of derivative computation and half-wave rectification are suggestive of a cortical locus for its implementation due to two lines of evidence. Firstly, the low spontaneous rate of discharge of V1 neurons (Movshon et al., 1978; Williams & Shapley, 2007) implies half-wave rectification, as the neurons cannot lower their response rate below zero. Secondly, Gaussian derivatives models have been found to accurately describe the spatial and temporal properties of primate simple cells (Young et al., 2001; Young & Lesperance, 2001). However the visual signal does not arrive at the visual cortex without any prior processing. It has already passed through the retina, which has filtering properties of its own. Retinal ganglion cells are known to have an antagonistic centre-surround receptive field profile (De Monasterio & Gouras, 1975), which is typically described by a Difference-of-Gaussians (DoG) model (Enroth-Cugell et al., 1983; Enroth-Cugell & Robson, 1966; Rodieck & Stone, 1965). The receptive field profile of a DoG filter is shown in figure 3.19 (left). It is composed of a

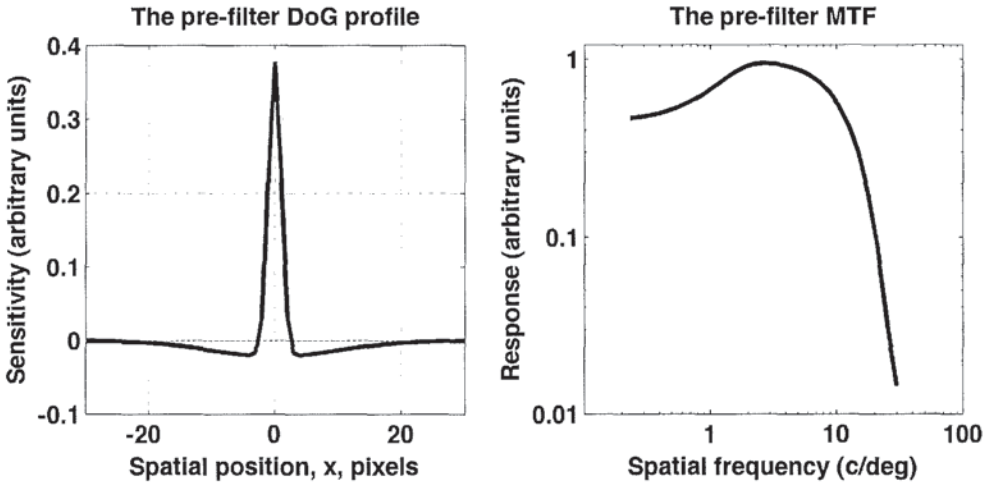

Figure 3.19: The pre-filter receptive field profile and its MTF

small central upright Gaussian added to a broader inverted Gaussian of low amplitude. Its modulation transfer function (MTF) is shown in the right panel of figure 3.19, which illustrates a mild band-pass filter. The addition of a DoG pre-filter to the single-scale model may provide predictions that more accurately represent the effect of human visual processing, and provide a better fit to the present experimental data.

A DoG filter can be characterised by 3 parameters; the spread of the excitatory small central Gaussian (s_c), the spread of the inhibitory large surrounding Gaussian (s_s), and the weight of the large Gaussian relative to the small Gaussian (k). These values determine the filter's shape and response characteristics. A DoG pre-filter was added to the single-scale model using the median parameter values specified by Croner & Kaplan (1995) for primate central parvocellular ganglion cells (s_c = 1min, s_s = 6mins and $k = 0.5$).

The filter's effect is to accentuate points of high curvature. For the experimental images with +1 added ramp gradient this filter introduces a small, central negative gradient (figure 3.20), that generates a peak in the LD edge channel, where there previously was none, so predicting two edges instead of one.

By symmetry, two edges are also predicted for images with -1 added ramp gradient. However, the filter parameter values used here did not produce this essential small reversal in gradient for experimental images of added ramp gradients of ±1.5 or ±2. Therefore the model still predicted only one edge for these images. But two edges were predicted when s_s was sufficiently increased, as described below.

The current model does not account for edge probabilities other than 0 or 1. In order to allow for intermediate probabilities, a noisy decision process was appended to the model. It assumed independent detection of the 2 edges in 2 separate channels, but with the same noise level and criterion applied in each channel. It had the form:

$$p = \Phi\left(\frac{a_1 - c}{\sigma_n}\right).\Phi\left(\frac{a_2 - c}{\sigma_n}\right) \tag{3.5}$$

where a_1 and a_2 are the response amplitudes of the two opposite polarity edges, σ_n is a the standard deviation of the noise, Φ is the standard normal integral, which converts a z-score to a probability, and c is the observer's criterion given by:

$$c = -z(F).\sigma$$

where $z(F)$ is the z-score of the false alarm rate. Thus p represents the probability that both edge response amplitudes exceed the decision criterion c, and led to the '2-edge' response.

The addition of the pre-filter and noisy decision process has added 5 parameters to the model; s_c, s_s and k from the pre-filter and σ_n and F from the noisy decision process. Initial model fits to the data averaged across observers allowed all 5 parameters to be free, and was implemented in Matlab (using 'fmins') by minimising the RMS error between the output of the model and the data averaged across observer. The model produced values for s_c and k (1 and 0.42 respectively) close those expected for central P cells (Croner & Kaplan, 1995). However, the fitted value for s_s (60mins) was was implausibly high. When s_s was fixed at the more plausible value of 6mins, the resulting filter was unable to account for 2 edges at added ramp gradients over ±0.5. However, when s_s was fixed at 10mins, the model

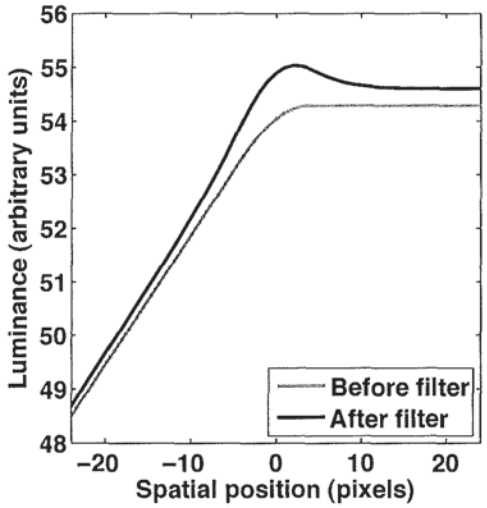

Figure 3.20: Central region of an experimental waveform (added ramp gradient = 1.5) before and after an initial version of the pre-filter. The 'after filter' curve has been scaled by an arbitrary amount and contains a small negative gradient that is absent in the 'before filter' curve.

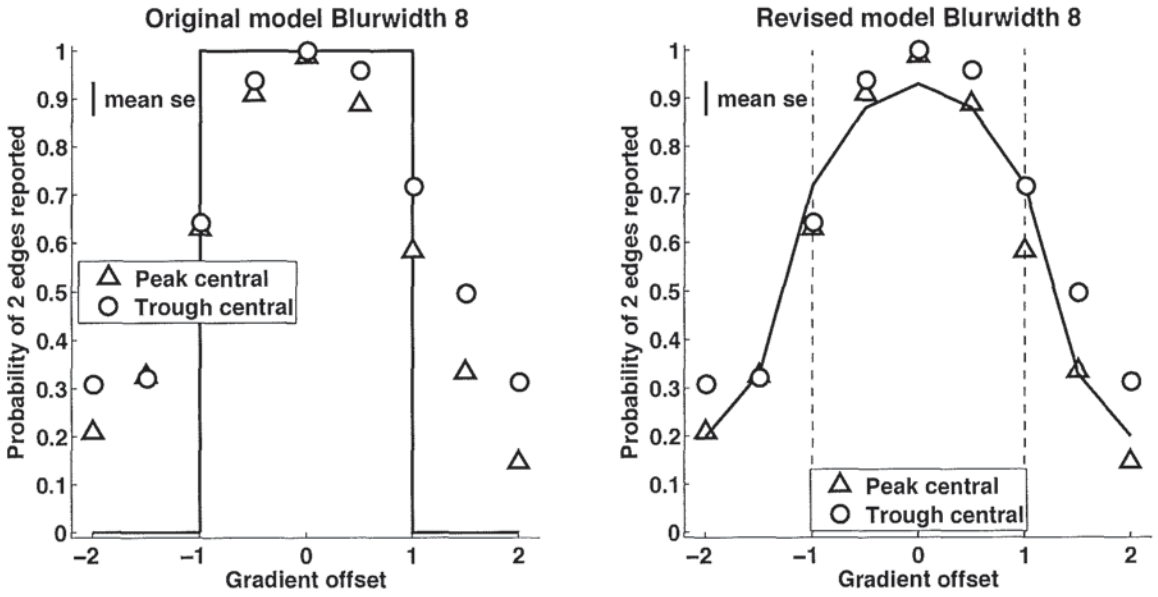

Figure 3.21: Proportion of '2 edges' responses averaged across 4 observers (symbols) and model fits (lines) plotted against gradient of added luminance ramp. Left and right plots show identical data with fits of the original and revised models respectively. The mean se bar shows the inter-observer variation.

returned 2 edges at added ramp gradients of ± 1 and ± 1.5. The gradient ± 2 data were then fitted by fixing the false alarm rate at 0.2. The best-compromise model parameters are shown in table 3.1. Although the value of s_s (10) is higher than the median reported by Croner & Kaplan (1995), it is within their specified interquartile range. Therefore, it is still biologically plausible.

The addition of the pre-filter and noisy decision process to the model resulted in the revised fit shown in the right plot of figure 3.21. The revised model produces a strikingly better fit to the data than the original model.

Parameter	Value
Pre-filter centre standard deviation (s_c) mins	(1)
Pre-filter surround standard deviation (s_s) mins	(10)
Pre-filter surround/centre weight (k)	(0.55)
False Alarm probability (F)	(0.20)
Noise (σ_n)	1.30×10^{-5}

Table 3.1: Fixed and free parameters in the single-scale model with pre-filter and noisy decision process. Brackets indicate fixed values.

3.9 Experiment 2.2 - Mach edge detection at a finer scale

3.9.1 Introduction and method

Experiment 2.1 used images whose luminance profiles consisted of a single blurred triangle-wave (blurwidth 8 pixels) with added linear luminance ramps of various gradients. Observers consistently reported that 2 edges were seen near the image centre for added ramp gradients of up to ± 0.5, but at higher gradients, responses fell gradually towards '1 edge' . These results were well predicted by a revised nonlinear 3rd derivative model. However, only one blurwidth was tested. So it is unknown whether the results generalise to other blurwidths. Therefore experiment 2.1 was repeated with the same subjects, but the waveform blurwidth was reduced from 8 to 2. One subject (SAW) also completed 6 sessions at blurwidth 32.

3.9.2 Results

The proportion of '2 edges' responses for each condition is shown in figure 3.22. Data from blurwidth 8 are also shown for comparison. The pattern of results is clearly similar to the blurwidth 8 data. Two edges were reliably perceived at added ramp gradients within ± 1, and this fell away to 1 edge as the ramp gradients increased.

Pearsons X^2 statistic (with Bonferroni correction) was applied in a comparison between the blurwidth 2 and blurwidth 8 data, separately for each panel. A significant effects is annotated in 1 panel. However, no corresponding inference can be drawn, as the effect was not significant for the other 3 observers, and the 'average' plots indicate little or no blurwidth effect. Also, since all of the blurwidth 2 data were collected after the blurwidth 8 data, any comparison between the respective data may be subject to order effects.

One observer (SAW) also completed 6 sessions at blurwidth 32. This data is shown in figure 3.23, in addition to the data from the other two blurwidths. There does appear to be a trend of lower response probability with increasing blurwidth. Pearsons X^2 statistic (with Bonferroni correction) was significant for pair-wise comparisons between peak central, blurwidths 2 & 8 ($X^2 = 52.29, d.f. = 9, p < 0.001$), peak central, blurwidths 2 & 32 ($X^2 = 118.66, d.f. = 9, p < 0.001$), and trough central blurwidths 2 & 32 ($X^2 = 31.32, d.f. = 9, p < 0.00026$).

3.9.3 Modelling

The single-scale model with pre-filter and noisy decision process was fitted to the blurwidth 2 data averaged across observer and pooled across image polarity, with all 5 parameters fixed at the best-compromise values that were used for the blurwidth 8 data. The model fit and data are shown in figure 3.24, with the data and fit for blurwidth 8, for comparison. The model correctly captures the data trend and provides a good fit to the data, with no free parameters. Despite the fact that all 5 model parameters were fixed, the blurwidth 2 fit is higher than the blurwidth 8 fit. This difference arises because a reduction in blurwidth causes the amplitude of the 3rd derivative of the luminance profile to increase. So the amplitude of the outputs of both edge channels is correspondingly increased, which results in a higher probability of '2 edges seen' responses, providing that all other factors are constant.

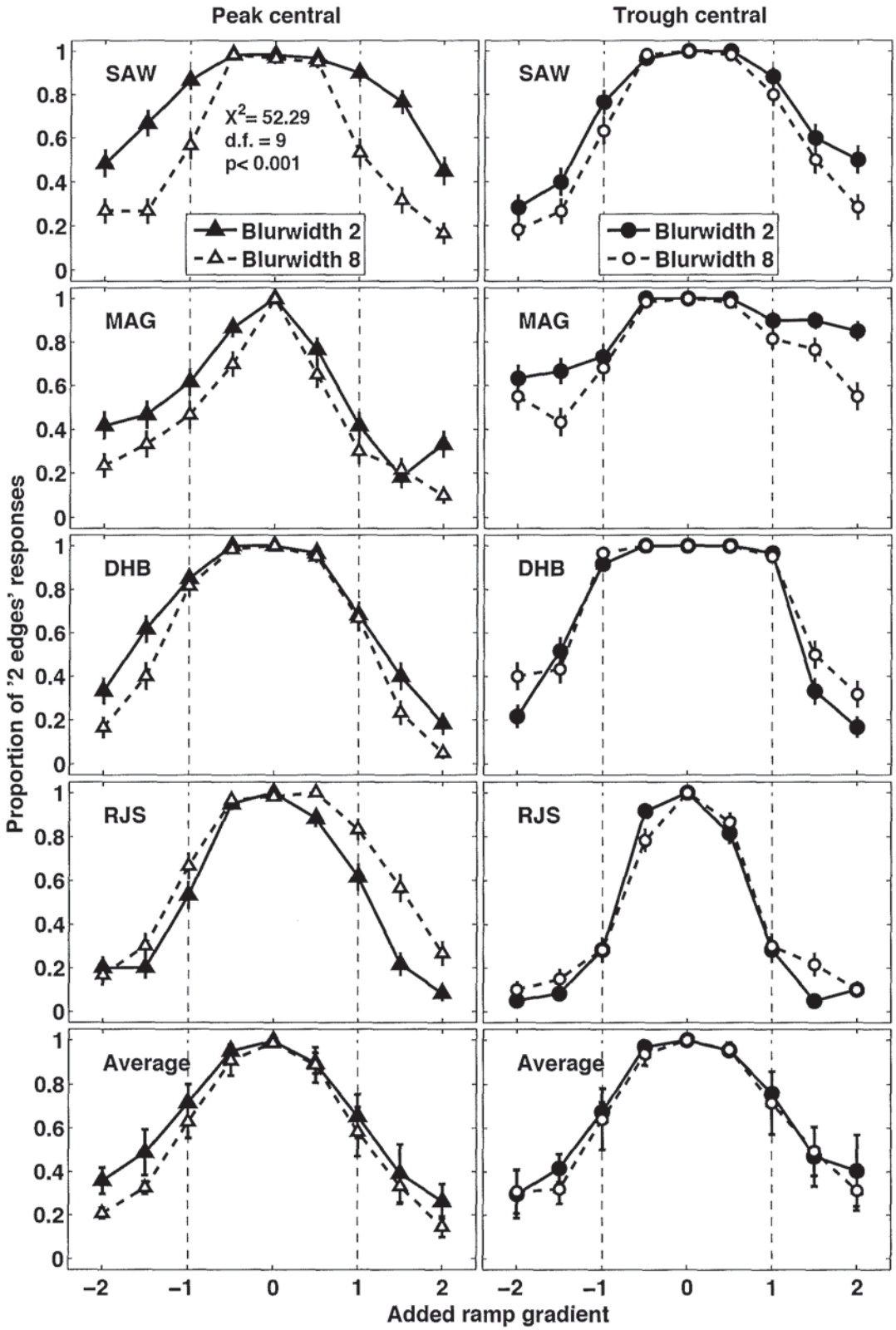

Figure 3.22: Effect of blurwidth - Proportion of '2 edges' responses for 'peak central' and 'trough central' phases (left and right columns) of a blurwidth 2 (filled symbols) or blurwidth 8 (open symbols) images. Top 4 rows show data from each observer. The mean across observers is shown in the bottom row. Error bars are as figure 3.18. Panels are annotated with Pearsons X^2 where significant.

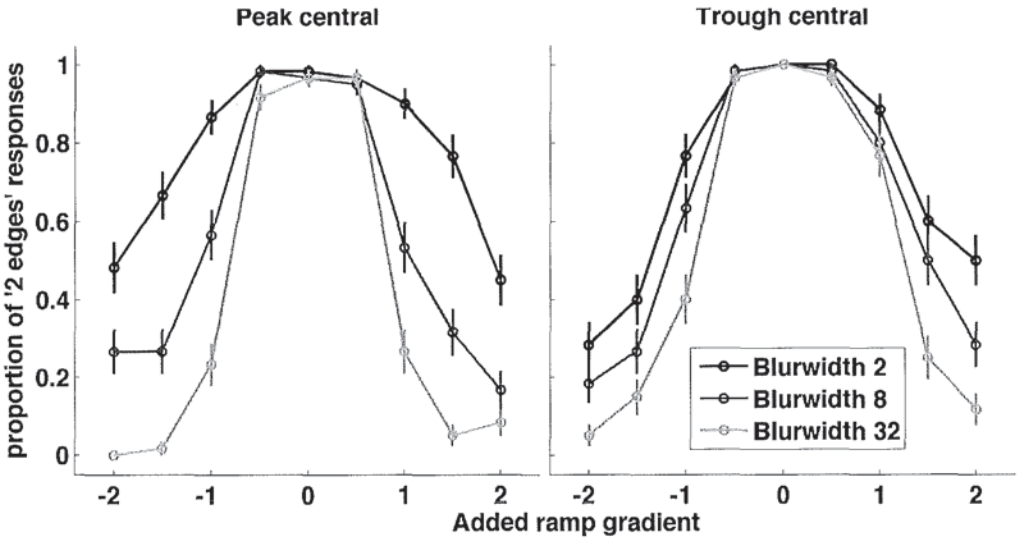

Figure 3.23: Proportion of '2 edges' responses as a function of gradient of added ramp, for observer SAW. Error bars ($\pm 1sd$) are calculated from the expected variance in binomial sampling. For ramp gradients beyond ± 0.5 an increase in blurwidth leads to a reduction in the proportion of '2 edges' responses.

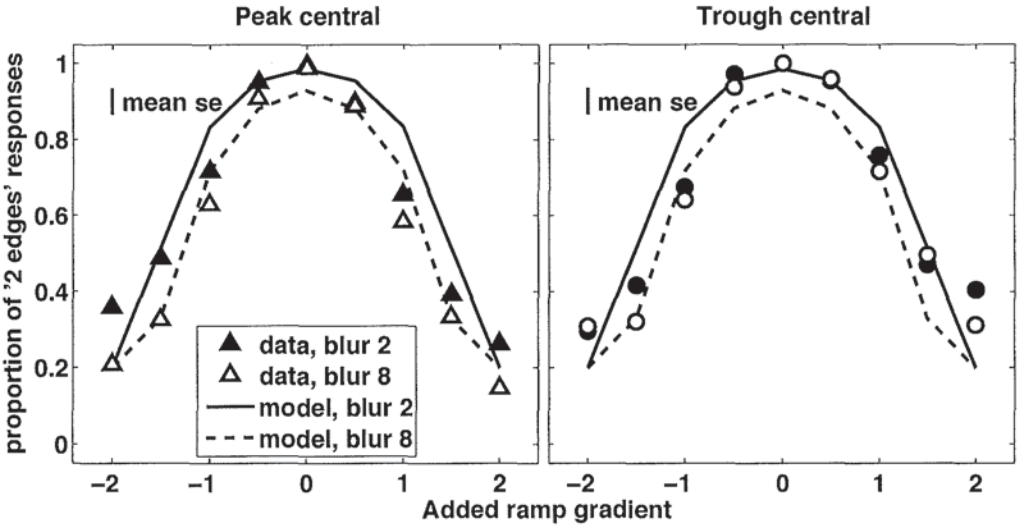

Figure 3.24: Effect of image blurwidth on model. Data averaged across observer from figure 3.22, reploted here with revised model fits where all 5 parameters are fixed at the values shown in table 3.1.

This explanation does not apply to the predictions for the added ramp gradient ± 2 images because they are equal to the false alarm rate.

Figure 3.24 shows that the model fits correctly capture the small increase in responses observed in some of the data when blurwidth was reduced. However, the RMS error between model and data is worse (0.10) for the blurwidth 2 images than the blurwidth 8 images (0.08). Although this difference

may reflect a noisier blurwidth 2 data set, it may alternatively suggest that the constraint of a common value for all 5 parameters is too restrictive, and a strikingly better fit would be obtained if greater freedom was permitted.

A second model fit was performed to the blurwidth 2 data, but with all parameters free. This resulted in a reduced RMS error (0.045), but the fitted value for the spread of the pre-filter's inhibitory surround (s_s) was implausibly high (49.3). This parameter was fixed at the more plausible value of 10, as before, and the modelling repeated with the other 4 parameters free. The resulting fit is shown in figure 3.25 (solid line) alongside the previous fit where all 5 parameters were fixed (dashed line).

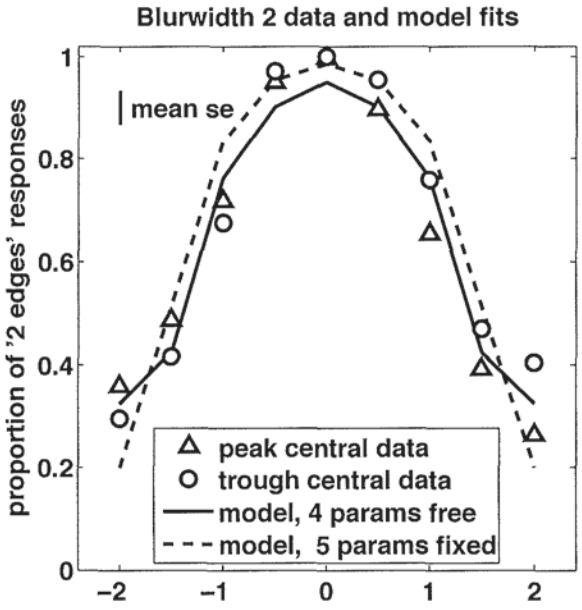

Figure 3.25: Competing model fits to the blurwidth 2 data. Blurwidth 2 data averaged across observer from figure 3.22, reploted here with revised model fits where all 5 parameters are fixed at the values shown in table 3.1 (dashed lines) or where 1 parameter was fixed (solid line). The mean se bar shows inter-observer variation (n=4).

As expected, the increase in the number of free parameters from zero to 4 has improved the model's fit to the data. The RMS error has fallen from 0.10 to 0.055. However, figure 3.25 shows that the improvement is not particularly striking, as the model now provides a worse fit to the data of added ramp gradient 0 and -0.5, and no improvement is apparent at -1.5 and $+0.5$. The parameters values for both fits are shown in table 3.2, where brackets denote fixed values. The second column shows the values for the fit where 1 parameter was fixed. When compared to the first column (fit 1, where all parameters were fixed) the fitted values are identical for s_c and similar for k. The false alarm rate has increased from 0.20 to 0.32, and again specifies the fitted value for the images of added ramp gradient ± 2. Fit 1 and 2 are optimised for the data from blurwidth 8 and 2 respectively. Since these two fits are not very dissimilar, a third model fit was performed to the data of both blurwidths simultaneously. The values of s_c and s_s were fixed at 1 and 10 respectively, and the other 3 parameters were free (but constrained to be the same for both blurwidths). This resulted in the parameters shown in the 3rd column of table 3.2 and the fit and data are shown in figure 3.26, as solid lines and symbols respectively. Although common parameter values were used in both fits shown, the fits differ across panels because

Parameter	Blurwidth 2		Blurwidths 2 & 8
	Fit 1 (all fixed)	Fit 2 (1 fixed)	Fit 3 (2 fixed)
Centre spread(s_c)	(1)	1	(1)
Surround spread (s_s)	(10)	(10)	(10)
Surround/centre weight (k)	(0.55)	0.48	0.49
False Alarm probability (F)	(0.20)	0.32	0.30
Noise (σ_n)	(1.30×10^{-5})	1.76×10^{-5}	1.56×10^{-5}
RMS error	0.10	0.055	0.072

Table 3.2: Parameter values and RMS errors for 3 fits of the revised single-scale model to the blurwidth 2 data (fits 1 & 2) or the data of both blurwidths (fit 3). In fit 1, all parameters were fixed at the values derived from the blurwidth 8 data. Brackets denote fixed values.

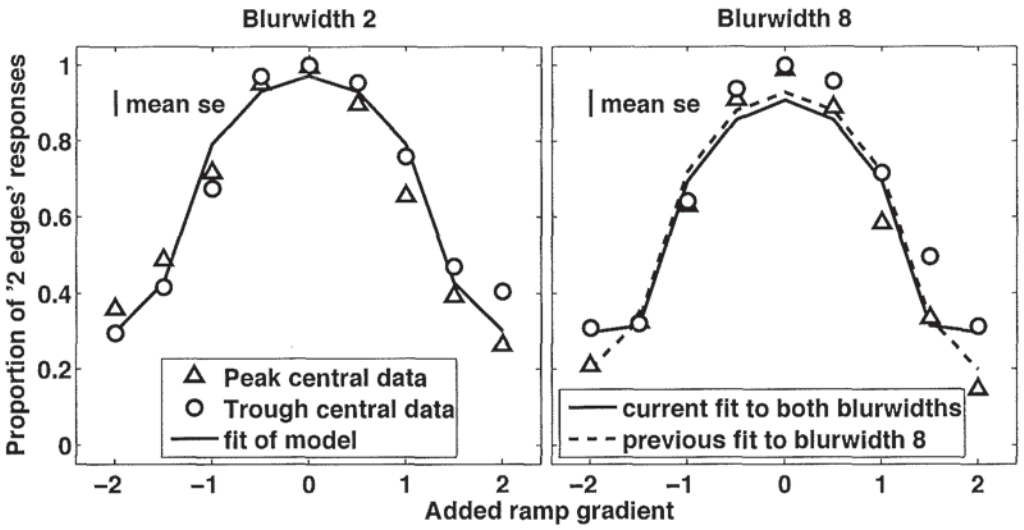

Figure 3.26: Revised single-scale model fitted simultaneously to the blurwidth 2 and 8 data (solid lines). The 2 fixed and 3 fitted parameters values are shown in table 3.2. The dashed line shows a previous fit to the blurwidth 8 data (fit 1) for comparison.

the model was applied to a different set of images in each panel. The fit to the blurwidth 2 data is very good, but the fit to the blurwidth 8 data is slightly too high at ± 2 and slightly too low for added ramp gradients 0 and ± 0.5. However, the fit is only slightly worse than the best previous fit to the blurwidth 8 data alone (fit 1) which is shown as a dashed line; the RMS error has increased from 0.077 to 0.072.

3.9.4 Conclusions

The similarity of the blurwidth 8 and 2 data indicates that the results of experiment 2.1 are not specific to that bluwidth, and that the effects are robust. Furthermore, SAW's blurwidth 32 data also shows the same pattern of gradually decreasing responses with increases of the absolute value of the added ramp gradient.

Since the blurwidth 2 data are so similar to the blurwidth 8 data, it is unsurprising that the model still provides a good fit, even with all 5 parameters fixed. However a modest alteration in 3 of the model parameters improved the fit to the blurwidth 2 data, without unduly worsening the fit to the blurwidth 8 data. The current version of the revised single-scale model provides a good fit to all of the data. Thus it may provide an accurate description of several stages of edge detection in early vision.

The pattern of results across blurwidth for SAW (figure 3.23) demonstrates a blurwidth effect, but this effect was not apparent in the data averaged across observers. It may be the case that SAW was better able to discriminate between the different image sets (blurwidths) due to prior exposure to them while piloting these experiments.

3.10 Experiment 2.3 - Revised model predictions are tested with short duration stimuli

3.10.1 Introduction and method

The addition of the pre-filter to the model produced revised predictions that matched data well. The pre-filter has a Difference of Gaussian receptive field shape that was inspired by known retinal ganglion cell physiology. It consists of a central peak, flanked by two regions of an inhibitory surround. An examination of the spatial and temporal properties of macaque parvocellular and magnocellular lateral geniculate nucleus cells has suggested that the inhibitory surround is less sensitive to higher temporal frequencies than the centre (Barlow, 1958; Derrington & Lennie, 1984). Furthermore, psychophysical work has found that the low frequency cut in the CSF is rather mild at short durations (25-100 ms) compared with long durations (500-1000 ms) (Georgeson & Georgeson, 1987; Legge, 1978b). So the use of short duration stimuli may reduce the effect of the surround, so that the ganglion cells now become more like simple low-pass blurring filters. When applied to the present experimental waveforms, they would no longer be able to accentuate edges. So a reduction in stimulus duration is predicted to reduce the proportion of '2 edges' responses to images of added gradient above ± 0.5. The Yes-No method of experiment 2.2 was repeated here, except that stimulus duration was reduced from 300ms to 50ms.

3.10.2 Results

The Proportion of '2 edges' responses for each condition is shown in figure 3.27, as a function of added ramp gradient, and is plotted alongside the 300ms data from experiment 2.2. Left and right columns show data for each image polarity and the top 4 rows show data from each observer. Error bars ($\pm$1sd) are calculated from the expected variance in binomial sampling. The data averaged across observers are shown in the bottom row. Error-bars here ($\pm$1se) represent inter-observer variation ($n = 4$). The

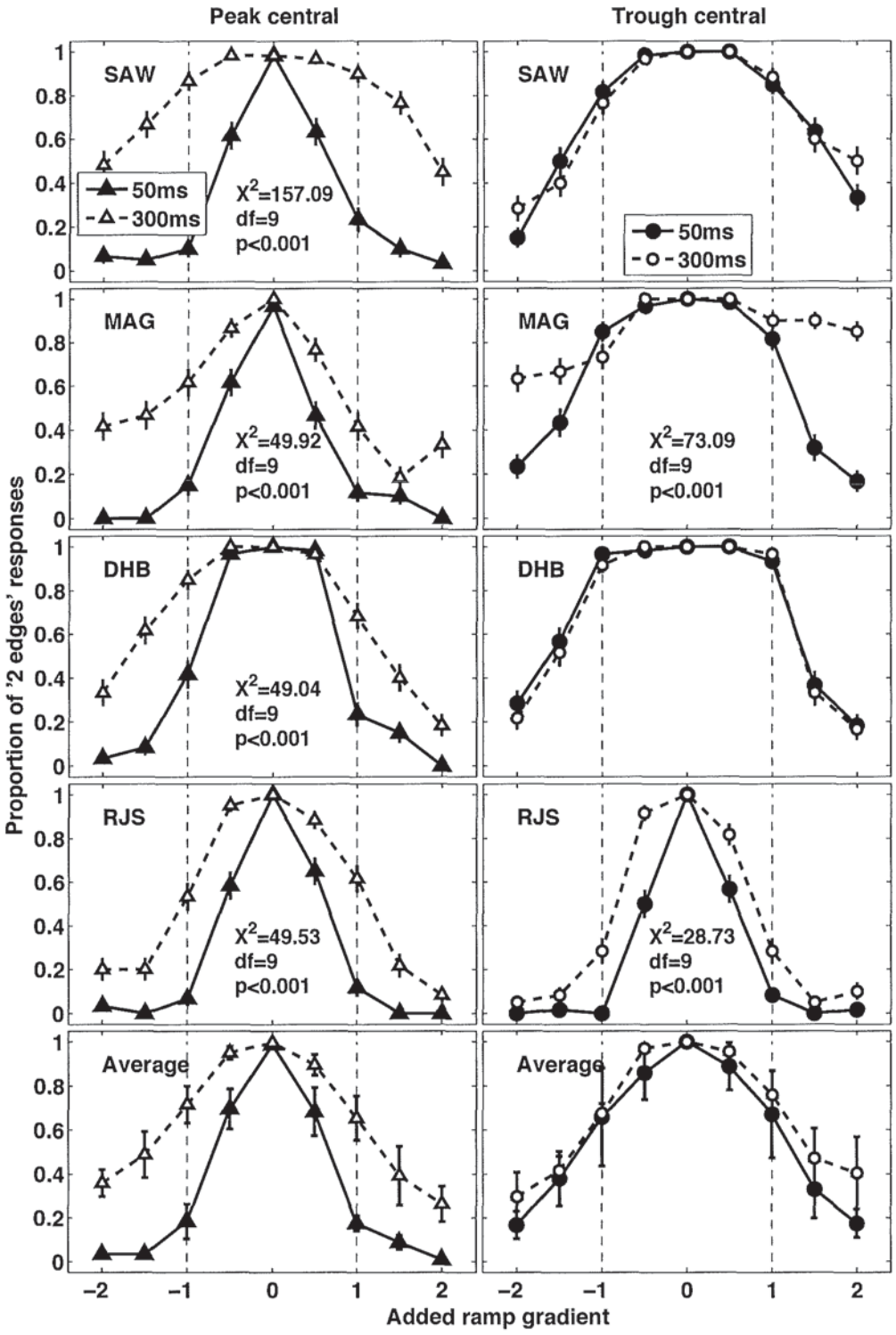

Figure 3.27: Effect of stimulus duration (blurwidth 2). Proportion of '2 edges' responses for 'peak central' and 'trough central' phases (left and right columns) when stimulus duration was 50ms (filled symbols) or 300ms (open symbols). Top 4 rows show data from each observer and means across observers are shown in the bottom row. Error bars are as figure 3.18. Panels for individual observers are annotated with Pearsons X^2 statistic where the stimulus duration effect is significant.

proportion of '2 edges' responses is lower in the peak central images than the trough central images, for all added ramp gradients except zero.

Panels for individual observers are annotated with Pearsons X^2 statistic where the stimulus duration effect is significant (Bonferroni correction applied). For all observers, the 50ms peak central stimuli evoked a significantly lower '2 edges' response than the 300ms stimuli, as predicted. Unexpectedly, the trough central stimuli show a weaker (but still significant) duration effect for observers MAG and RJS, and barely any duration effect at all for observers SAW and DHB.

A comparison across the columns of figure 3.27 shows the effect of image polarity. The average data (bottom row) show little or no polarity effect for the 300ms duration data. But at 50ms, the responses are lower for peak central than trough central images. This pattern is particularly striking for observer SAW. A Pearsons X^2 statistic (with Bonferroni correction) was calculated for each pairwise comparison of peak central and trough central data. This showed a significant polarity effect for the 50ms data for SAW, MAG and DHB, and the 300ms data for MAG ($X^2 = 28.73$ to 157.09, $df = 9$, $p < 0.001$).

3.10.3 Discussion

The results show that a reduction in stimulus duration from 300ms to 50ms causes a significant reduction in '2 edges seen' responses in the peak central images for all observers, and in the trough central images for 2 observers. This would appear to support the prediction that the use of short duration stimuli would reduce the effect of the pre-filter's inhibitory surround. But if this explanation is correct, then the trough central data should be very similar to the peak central data. However, the data averaged across observers (bottom row of figure 3.27) shows a dissimilarity between the two image polarities, as the effect of stimulus duration is large in the peak central images but very small in the trough central images. In the latter case the overlapping error-bars suggest that no difference can be concluded between the 50ms and 300ms data. To conclude, the experimental data cannot be considered to support the above prediction, without further qualification. So this casts doubt on a simple interpretation of the pre-filter's role in Mach edge detection.

The overall pattern of data suggests an interaction between image polarity and stimulus duration. For 2 of the 4 observers, a significant effect of polarity was present in the 50ms data but absent in the 300ms data. This is consistent with a rapid gain control mechanism, which is explained as follows. Gain control is a process whereby a cell's sensitivity is kept within it's operating range despite a much wider range of retinal illumination. The time course of this process is known to be rapid (Enroth-Cugell & Shapley, 1973a; Saito & Fukada, 1986; Shapley, 1986), so can be assumed to be complete by 300ms, but may still be ongoing at 50ms. The luminance at the image centre (close to the Mach edges) was higher in the peak central images (46 to $53cd/m^2$) than in the trough central images (22 to $31cd/m^2$), and this would lead to a local contrast that was lower for the peak central than trough central images. The 300ms stimuli in the present experiment would be subject to the full effect of gain control, which would be expected to dramatically reduce the difference in effective local contrast between peak central and trough central images, and thus lead to the observed absence of a polarity effect. The 50ms stimuli, however, would not be so strongly affected by gain control, and therefore the polarity effect would be expected to survive.

As a further test of role of the pre-filter, it was hypothesised that the model would fit the 50ms data by applying the best-compromise parameters obtained from the 300ms fit, but with the weight of the

pre-filter's inhibitory surround (k) reduced to zero or a low value. This reduction is intended to mimic the attenuation of the pre-filter's inhibitory surround which is expected with short duration stimuli. Figure 3.27 (bottom row) shows that for the 50ms duration, the proportion of '2 edges' responses is lower in the peak central images that the trough central images. Therefore the model was fitted separately for each image polarity. All of the model parameters except k were fixed at the values used previously that provided the best fit to the blurwidth 2, 300ms duration data (see table 3.2, fit 2). k was free. Model fits are shown in figure 3.28 (solid lines) with data averaged across observer (symbols). The error-bars (±1se) represent inter-observer variation (n=4). Dashed lines show the fit when k was optimised for the 300ms, blurwidth 2 data. When k was free, the model returned its value as 0.11 for both peak central and trough central images. This value is much reduced from that obtained with the 300ms duration data (0.48). This would appear to support the hypothesis that the model would fit the 50ms data by applying the parameters obtained from the 300ms fit, but with the weight of the pre-filter's inhibitory surround reduced to zero or a low value. However, the resulting model fits shown in figure 3.28 (solid lines) are strikingly too high for the peak central data of added ramp gradients ±1 to ±2, and the trough central data of ±2, so this model version cannot be described as a good fit to the data. Therefore, the hypothesis is not supported.

The model is shown in figure 3.28 to provide identical predictions for the peak central and trough central image sets for blurwidth 2, when the same parameters are used for both image polarities. This was also the case for the blurwidth 8 images, and reflects the symmetry of each image with its opposite polarity. Whilst the 300ms duration data of both blurwidths show this symmetry, the present 50ms data do not. This suggests that the model lacks a factor that can account for the asymmetry.

A more striking deficiency of the current model is that the fixed false alarm rate of 0.32 prevents the fits from falling below 0.32, even though the data approach zero at added ramp gradients beyond ±1 for the peak central data. So a further fit was performed which allowed the false alarm rate to be free. The other 4 parameters were fixed at the the values fitted to the combined blurwidth 2 and 8 data

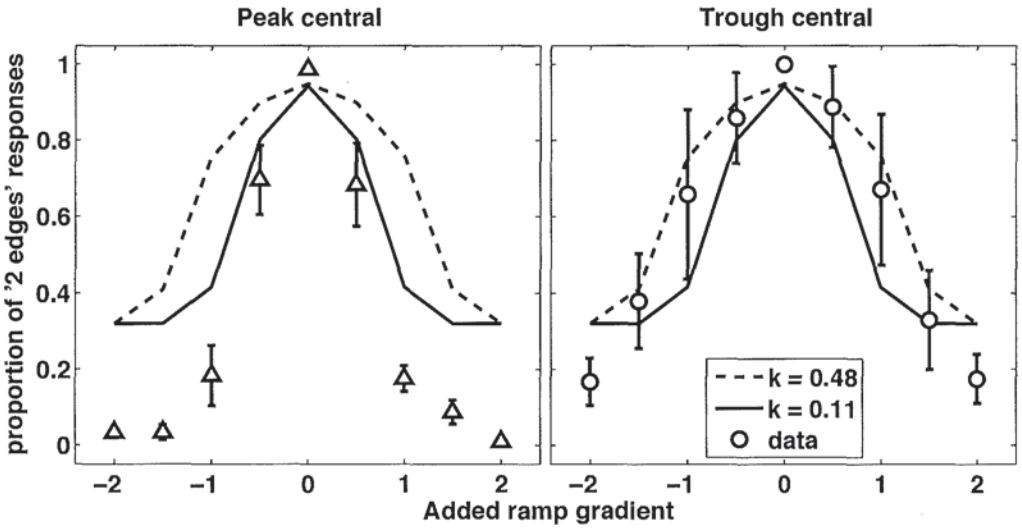

Figure 3.28: Effect of pre-filter's inhibitory surround weight (k) on model predictions (lines) for blurwith 2 images, with duration 50ms data averaged across observers (symbols). The error-bars (±1se) represent inter-observer variation (n=4).

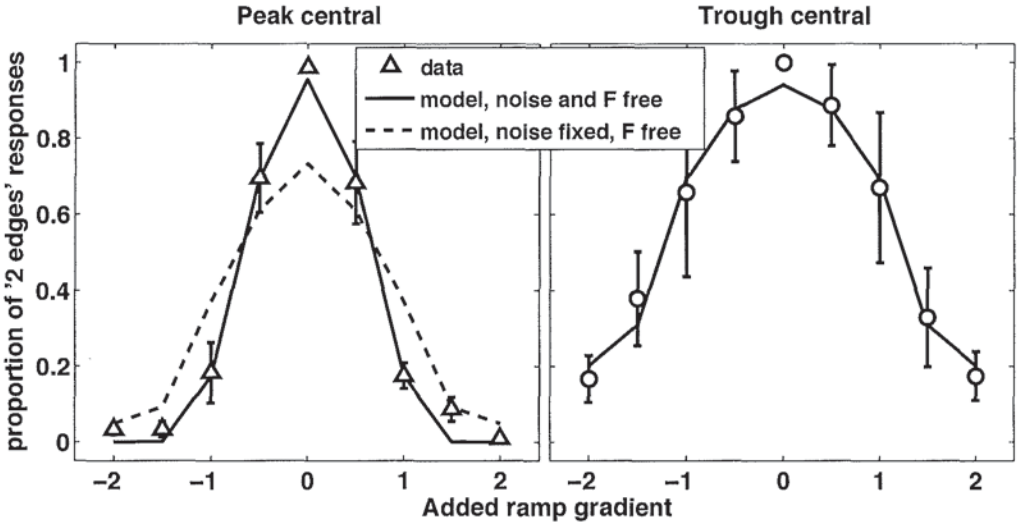

Figure 3.29: Blurwidth 2, duration 50ms data averaged across observers (symbols) and 2 model fits. Solid lines show the fit where noise and the false alarm rate were free and these parameter values are specified in table 3.3. Dashed lines show the fit where only the false alarm rate was free, and is superimposed on the solid line in the right panel. The error-bars (±1se) represent inter-observer variation (n=4).

sets (fit 3, table 3.2). But the resulting fit to the peak central data was poor, as shown in figure 3.29 (dashed line). So a further fit was performed that allowed noise and the false alarm rate to be free, and the other 3 parameters were fixed as before.

The resulting parameter values are shown in table 3.3, and the data and model fits are shown as solid lines in figure 3.29. Dashed lines show the fit where only the false alarm rate was free. For the trough central polarity the dashed line is beneath the solid line because when noise was free, the model returned a value of 1.56×10^{-5}, which is its fixed value in the other fit. The error-bars (±1se) represent inter-observer variation (n=4). The model now provides a strikingly better fit to the data when compared to figure 3.28, and correctly captures the effect of image polarity. However this property of the model appears to be derived largely from the different false alarm rates for the two polarities, as table 3.3 shows that the difference in noise (the only other free parameter) is relatively small (a factor of 2). But it seems implausible that different false alarm rates should apply to images of opposite polarity, as they were randomly interleaved within the same experiment. Therefore the model

Parameter	Peak central	Trough central
Centre spread(s_c)	(1)	(1)
Surround spread (s_s)	(10)	(10)
Surround/centre weight (k)	(0.49)	(0.49)
False Alarm probability (F)	0	0.20
Noise (σ_n)	0.73×10^{-5}	1.56×10^{-5}
RMS error	0.03	0.04

Table 3.3: Parameter values and RMS errors for fits of the revised single-scale model to the blurwidth 2, 50ms duration data, plotted as solid lines in figure 3.29. Brackets denote fixed parameter values.

currently appears to be unable to account for the data of both polarities in a harmonious manner, and may require revision. However, it is not clear how to perform a revision which accounts for the polarity difference in the current data and its lack in the 300ms duration data.

3.11 Experiment 3 - Mach edge location is tested with a feature-marking paradigm

3.11.1 Introduction

Experiments 2.1 to 2.4 examined the detection of edges in blurred triangle-wave images with added luminance ramps. Observers reported the presence or absence of a 2nd edge. But the revised single-scale model does not just predict the presence or absence of edges. It is also able to return spatial location predictions. The aim of this experiment is to test how well the model's spatial location predictions match the marked edge locations produced by observers. It can be expected that the model will provide a very good fit to the marked edge locations when the added ramp gradient is zero, because that was found to be true in experiment 1, which examined this condition. But as the added luminance ramp steepens, the revised single-scale model will be shown to predict a progressive shift in edge locations and a systematic increase in the asymmetry of edge strengths. This experiment will determine whether these changes are reported by observers and whether they can be correctly captured by the model.

3.11.2 Method

The image set was the same as the blurwidth 8 images used in experiment 2.1. Blurwidth 2 images were not used here, as the Yes-No method had revealed little or no difference between results for blurwidth 2 and blurwidth 8 at 300ms duration. Furthermore, the position data for blurwidth 8 in experiment 1.1 were within $\pm$ 1.2 pixels of the blurwidth 2 data for each observer (see figure 3.3). The feature-marking method that was used in experiments 1.1 to 1.4 was applied here. But now, observers marked the position and polarity of all perceived bars in addition to edges, in order to characterise stimulus perception in more detail. Stimulus duration was 300ms. There were 18 conditions (9 ramp gradients, 2 polarities, 1 blurwidth). Each session consisted of 6 repetitions of each condition in a randomised order and took about 25mins to complete. The same four observers who took part in experiment 2 were used here, as they had prior exposure to this image set. They each performed 5 sessions, but the first session was discarded as practice.

3.11.3 Results

Tables were generated for each observer's responses in a given condition and any clear finger errors were corrected by eye. This was a straight-forward procedure, as there was never more than one occurrence of each type of feature in any one trial, and so this enabled the reallocation of a datum to a new feature type. For example, a 'DL edge' response at a location which was within a range of locations marked as a 'LD edge' on most other trials was corrected to 'LD edge'. Data were similar for all 4 observers, so averages and model predictions are shown in figure 3.30 as symbols and lines respectively.

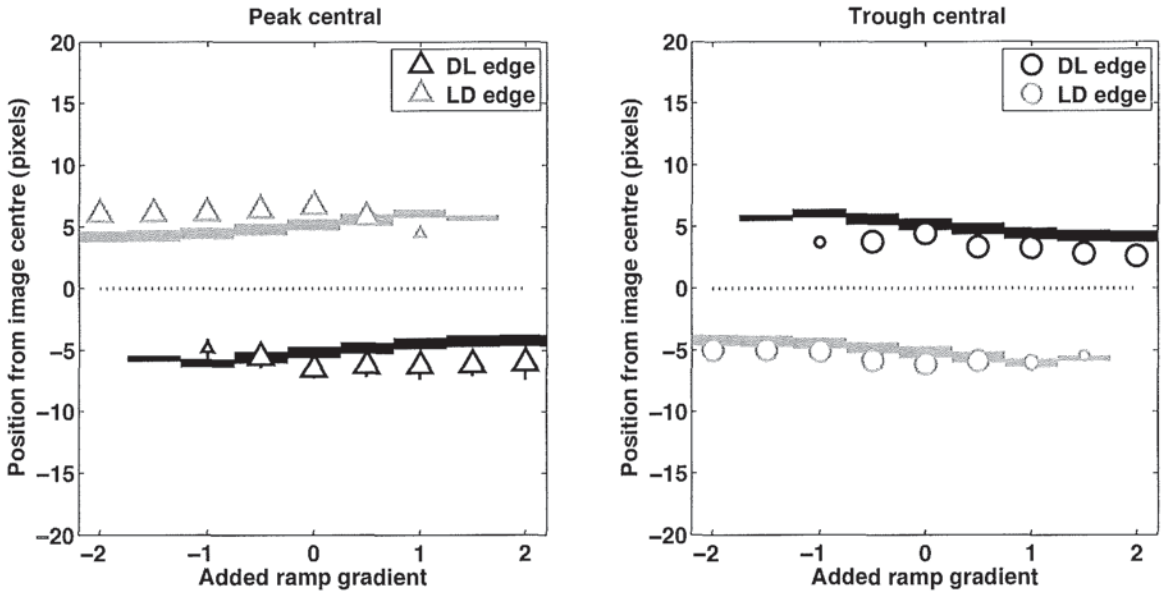

Figure 3.30: Feature-marking data averaged across observer (symbols) and revised single-scale model predictions (lines) for peak central (left) and trough central (right) blurwidth 8 images. Each symbol represents up to 96 responses, and symbol size (and line width) represents the probability of response (4 bins). Symbols and lines are absent if the probability fell below 0.25. Errorbars (±1se) are plotted behind symbols and represent inter-observer variation (n=4).

Position from image centre is negative for positions left of centre and positive for positions right of centre. Each symbol represents up to 96 responses, and symbol size (and line width) represents the probability of response (quantized into 4 bins). Symbols and lines are absent if the probability fell below 0.25. Errorbars (±1se) are plotted behind symbols and represent inter-observer variation (n=4). Lines show predictions of the revised single-scale model (with pre-filter and noisy decision process), with its 5 parameters fixed at the values used in experiment 2.1 (and specified in table 3.1).

All observers marked a central bar with 2 edges for added ramp gradients up to ±0.5. For gradients $\geq \pm 1$, perception of 2 edges smoothly reduced to 1 edge (smaller symbols denote fewer markings). The marked edge positions show a small shift towards image centre as the added ramp gradient increased or decreased from zero. But this trend was more apparent in the trough central data than the peak central data. Where an edge was consistently marked at added ramp gradient ±2, its position was about 1.5 pixels closer to the image centre than when the added ramp gradient was zero, for the trough central images. The peak central data show a smaller shift of only 0.5 pixels.

The predictions of the revised single-scale model (with pre-filter and noisy decision process) are shown as lines in figure 3.30. The model correctly predicts the polarity and location of the Mach edges. The RMS error between data and model was only 1.4 pixels for the peak central images and 1.1 pixels for the trough central images. The progressive shift in edge position with changes in added ramp gradient is also correctly captured by the model, especially for the trough central DL edge positions. Furthermore, the model correctly predicts the reduced probability of Mach edge perception where the data show this with smaller symbols. However the model departs slightly from the data, as it predicts

that the most peripheral edge positions should be located at added ramp gradients of -1 for the DL edge and +1 for the LD edge, rather than at zero for both edges. But this departure is small.

In the majority of trials the observers marked either a single edge, or a pair of opposite polarity edges and a central bar of the appropriate polarity (e.g a dark bar flanked by a LD edge on the left and a DL edge on the right). However, on a small number of trials this pattern was not followed. Observer RJS marked a pair of edges without a central bar in 5 trials, and observer MAG marked a bar and just one edge in 9 trials. Given that there were a total of 216 trials per observer, this is a tiny proportion, which may indicate uncorrected finger-errors (not corrected because they could not be clearly classified as such), or noise in the observers' visual system. However it may hint at the intriguing possibility that a bar and its flanking edges may not be inexorably linked.

Although the observers were responding only to the feature positions, their data also contain information regarding the frequency of edge perception, as the absence of an edge datum indicates that the edge was not perceived on that trial. The data were therefore transformed into the same form as the Yes-No data (experiment 2.1) by obtaining, for each condition, the proportion of times that 2 edges were perceived. These data are shown in figure 3.31, superimposed upon the corresponding Yes-No data for comparison.

The top 4 rows show data from each observer. The mean across observers is shown in the bottom row. Error bars are derived from the expected variance in binomial sampling for the top 4 row and are generally longer for feature-marking than Yes-No data, due to the lower number of trials per datum for each observer (feature-marking, 24; Yes-No, 60). The error-bars in the bottom row show inter-observer variation ($\pm$1se).

The results averaged across observers (bottom row) show rather little difference between the outcome of the two methods. All of the panels that show data for each observer and stimulus polarity also show little or no effect of method, except in two cases. These two are endorsed with a significant Pearson's X^2 statistic (with Bonferroni correction), but in both cases the data from the opposite polarity did not reach significance.

3.11.4 Discussion

The plots of the edge-position data of this feature-marking experiment (figure 3.30) have small error-bars, which indicate that Mach edge position is stable across observers. The large size of most of the symbols shows that the Mach edges are reliably reported. This was observed previously (experiments 1.1 to 1.3) when the added ramp gradient was zero, but is now seen to extend to non-zero gradients. However, a pair of Mach edges was not perceived in all images used here. The smaller symbols for some data in figure 3.30 show a reduced frequency of marking, and suggests that these edges were perceptually weaker. These data were from added ramp gradients of below -0.5 for the DL edge and above +0.5 for the LD edge, and included the images of added ramp gradient $\pm$1, which were blurred versions of the right or left half of the classic Mach band image. To take the example of a +1 added ramp gradient, peak central image, a DL edge was frequently marked at about 6 pixels left of image centre and a LD edge was marked at about 5 pixels right of image centre, but much less frequently. In the vast majority of trials, where 2 edges were marked, a central bar was also marked. In the present example, it was always a light bar. This bar appears to be an example of the well-known Mach band,

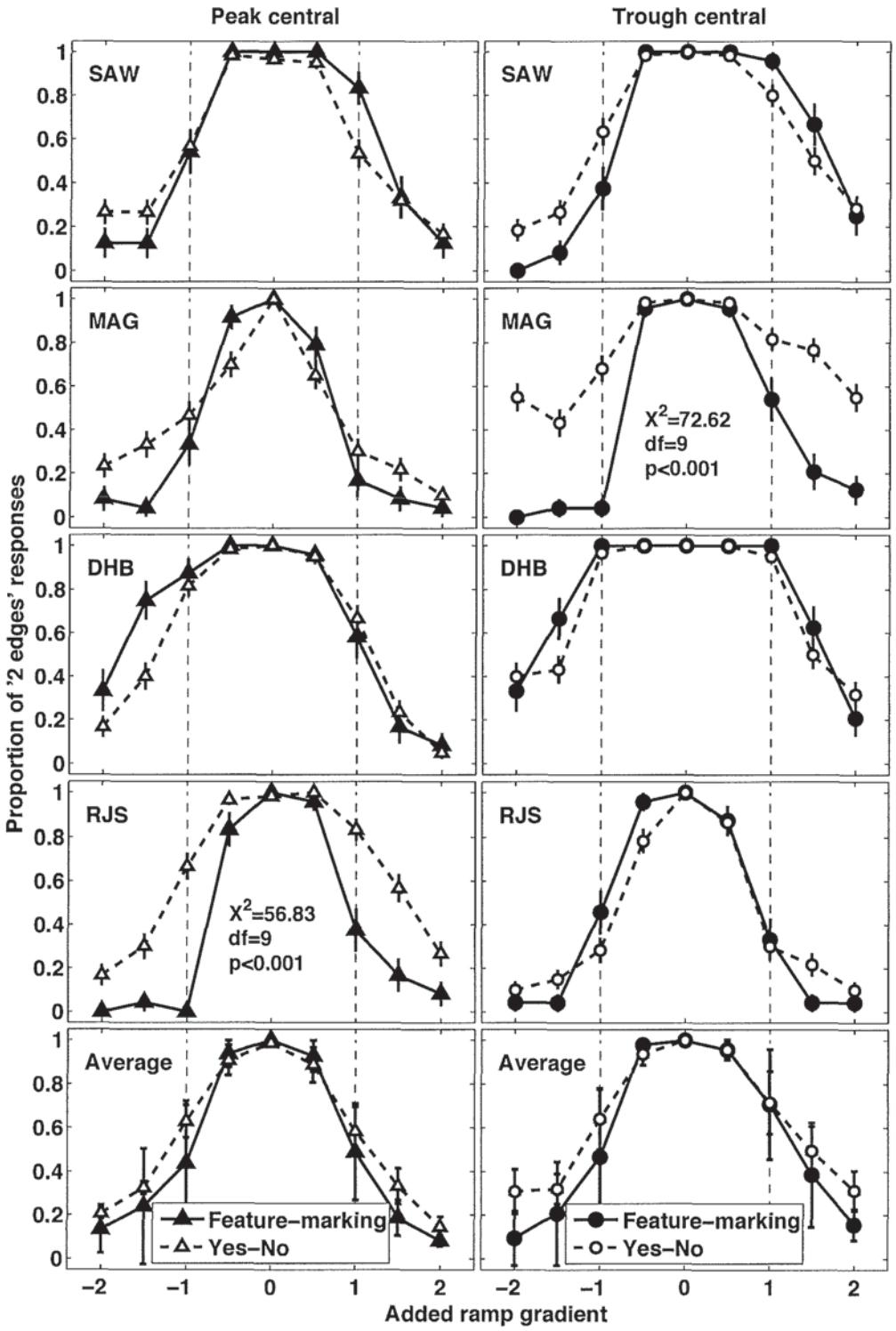

Figure 3.31: Effect of method. Proportion of times 2 edges were perceived when Yes-No and feature-marking methods were used with the same image set. Top 4 rows show data from each observer. The mean across observers is shown in the bottom row. Error bars are as figure 3.18, and are generally longer for feature-marking than Yes-No data, due to the lower number of trials per datum for each observer (feature-marking, 24; Yes-No, 60). A significant difference between methods is indicated on 2 panels by a Pearson's X^2 statistic (Bonferroni correction applied).

and the current experimental results now suggest the novel conclusion that Mach edges are the edges of Mach bands. The results also suggest that Mach bands are perceived when the plateau of a classic Mach band luminance profile is replaced with a ramp; an observation first made by Ernst Mach himself (Ratliff, 1965). Mach band perception will be explored further in the next chapter.

The bottom row of figure 3.31 shows the high degree of similarity between the present feature-marking data and the Yes-No data from experiments 2.1 to 2.3. The fact that two different methods produce such similar results, with the same image set, supports the conclusion that both are examining the same perceptual process. This similarity across method also improves the strength of any subsequent conclusions.

The feature-marking data were well-fitted by predictions of the revised single-scale model, with its 5 parameters fixed at the same values used in the Yes-No experiment which provided a good fit to the blurwidth 8, 300ms data (table 3.1). The predictions have a certain symmetry. The left panel of figure 3.30 shows the peak central data and predictions. The grey lines represent the LD edge predictions. If they are flipped left-right and then inverted they become the DL predictions (black lines). Similar symmetry is observed in the peak central data. The trough central data, whilst preserving the same general symmetry, show a shift of the DL data towards image centre. Thus, where a pair of edges is seen in an image, they are closer together in the trough central conditions than the peak central conditions. This appears to be an example of the Helmholtz irradiation effect, which was observed in the experiment 1.1 data. It is now shown to extend to the non-zero added ramp gradients used here. Most of the trough central data show a small shift towards the left, when compared to the symmetrical model predictions, which suggests a left bias in responses. But this directional bias is not present in the peak central data, so its cause remains puzzling.

The addition of a pre-filter and a noisy decision process to the single-scale model was required to account for the data when an luminance ramp was added to the blurred triangle-wave image of blurwidth 8. When there was no added ramp (gradient = 0), the revised model correctly predicted the polarity and position of the Mach edges and their very high frequency of occurrence.

However, it might be the case that the addition of the pre-filter and noise would shift predicted edge locations away from the data for the Mach edge images of other blurwidths, which were used in experiment 1.1 (range 1 to 64 pixels). The revised model was applied to the full set of images used in experiment 1.1. with its 5 parameters fixed as before. It predicted edge locations that were ≤ 1 pixel from the predictions of the model without the pre-filter and noise, and probabilities of occurrence that were 1 for all images except for the blurwidth 64 image, which was 0.96. Although observers marked 2 edges in 100% of the images in experiment 1, rather than the 96% predicted here, that experiment employed a method that generated the expectation of at least one edge in every trial. So the 4% difference may represent an increased false alarm rate. In conclusion, the addition of the pre-filter and noise did not worsen the model's prediction for the Mach edge images used in experiment 1.1.

To summarise, the revised single-scale model, with pre-filter and noise, provides a good fit to the Mach edge data of experiment 1.1 and the present experimental data. This supports the model, and the non-linear 3rd derivative approach in general, and suggests that the model correctly captures some aspects of human visual processing. However, the failure of the model to harmoniously account for the peak central, blurwidth 2, 50ms duration data suggests that the current model is incomplete. A nonlinear response to luminance in the early phase of light adaptation could be one source of extra complexity in this (Scott-Samuel & Georgeson, 1999).

3.11.5 The data are fitted with a 1st derivative model

The effect of the pre-filter is to accentuate changes in curvature of the luminance profile (figure 3.20). Crucially, it caused a small reversal of gradient in the luminance profiles of the experimental images of added ramp gradient ± 1 and ± 1.5. But since these filtered profiles now contain gradients of both signs, it now seems possible that a 1st derivative model would be able to generate edge predictions that also match the experimental data. In other words, the issue is whether the 3rd derivative filters in the revised single-scale model can be replaced by 1st derivative filters.

The blurwidth 8 waveforms were passed though a 1st derivative version of the model, with the same 5 parameter values used previously (table 3.1). The resulting predictions for the proportion of '2 edges seen' responses were all at 1 for added ramp gradients up to ± 2, where they fell to 0.2. The high probability of response arose because the amplitudes of the 1st derivative filter responses were much higher than those of 3rd derivative filters. So the filter gains were reduced to a suitable amplitude, by adding an 'amplitude multiplication factor' free parameter to the fitting program. The best fit was for an amplitude multiplication factor of 0.034. The resulting feature-marking predictions of this 1st derivative model, with pre-filter and noisy decision process, are shown in figure 3.32, with the experimental data. The model fit appears to be good. Although it does not capture the subtle changes

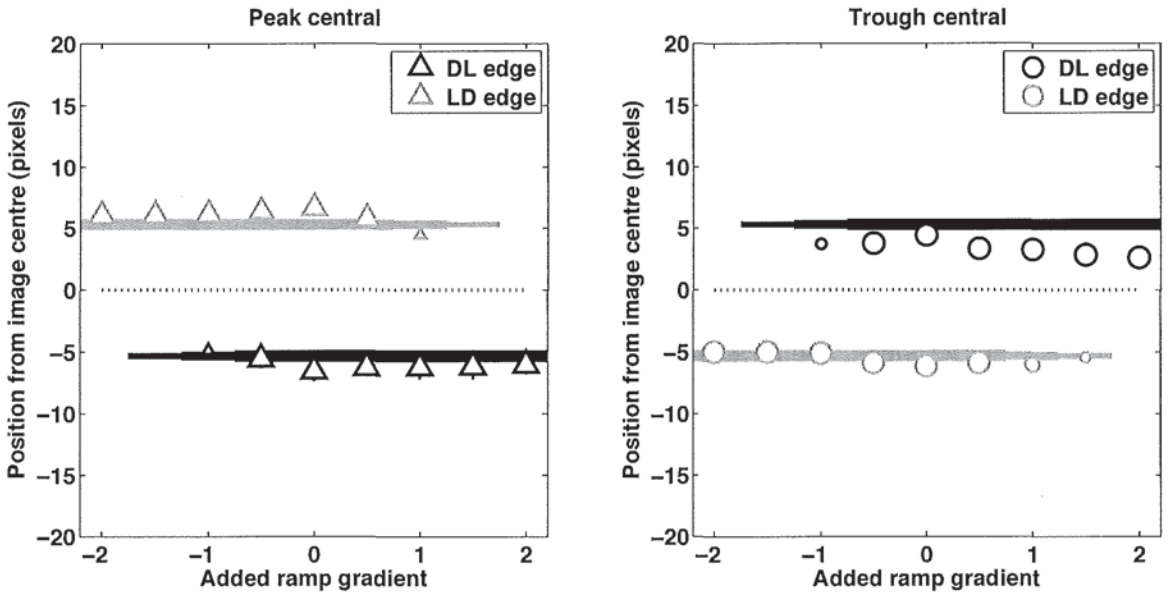

Figure 3.32: Feature-marking data averaged across observer (symbols) and 1st derivative model predictions (lines) for peak central (left) and trough central (right) blurwidth 8 images. Other details are as figure 3.30.

in position across added ramp gradients, a comparison with the 3rd derivative model predictions (figure 3.30), suggests that it fits equally well. The RMS errors between model and data are now 0.81 and 1.30 for the peak central and trough central polarities respectively, whereas for the 3rd derivative model these values were 1.42 and 1.10 respectively. The subtle changes in edge position observed in the data may merely represent random fluctuations. The Yes-No data with fits of both models are shown in figure 3.33. The 1st derivative model (RMS error = 0.065) fits the central data slightly better than the 3rd derivative version (RMS error = 0.077). Thus it appears to perform equally well or slightly better

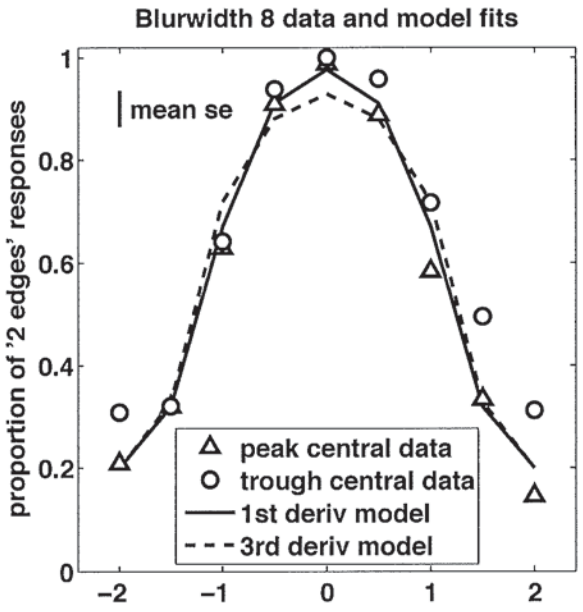

Figure 3.33: Blurwidth 8 Yes-No data and model fits obtained from the revised single-scale 3rd derivative model and its 1st derivative version.

that its 3rd derivative counterpart with both feature-marking and Yes-No data.

The simplification of the model may be taken a step further. The pre-filter is shaped approximately as a 2nd derivative operator, as it has 3 lobes (figure 3.19). This means that when it is followed by a 1st derivative operator, it may be able to replace the 3rd derivative filters in the earlier version of the single-scale model that had no pre-filter or noisy decision process. If so, then the Mach edge data of experiment 1.1 would be well-fitted by the predictions of a model that consists of a pre-filter followed by a 1st derivative operator. These predictions were generated with the same pre-filter parameter values as used previously (table 3.1). The Matlab 'gradient' operator was then used to obtain the 1st derivative. The resulting predictions of this 'pre-filter + 1st derivative' model for the Mach edge image set used in experiment 1.1 are shown in both panels of figure 3.34 (solid lines) with the data averaged across observer (symbols). The predictions of the single-scale 3rd derivative model, without blur are shown as dashed lines. Left and right panels show the 3rd derivative model without and with the pre-filter respectively (pre-filter parameters are as used previously - see table 3.3). The pre-filter + 1st derivative model clearly fits the data well, and provides a better fit at the low blurwidths than the 3rd derivative model. The fit to the trough central data is similar (not shown). Thus the 3rd derivative approach is no longer the sole dominant account for the results of experiment 1.1, as the addition of the pre-filter now means that a 1st derivative filter performs equally well (or even slightly better). The pre-filter + 1st derivative model has much in common with the 3rd derivative model. The pre-filter approximates a 2nd derivative filter, and when followed by a 1st derivative filter, the model effectively computes an approximation to the 3rd derivative.

A comparison across the panels of figure 3.34 shows that the addition of the pre-filter to the single-scale 3rd derivative model causes a very small improvement in its fit to the data. Thus there are now 3 models that can account for the Mach edge data of experiment 1.1. Firstly, there is the N3+ model, which cannot account for the experiment 2 data with added luminance ramps. Secondly, there is the

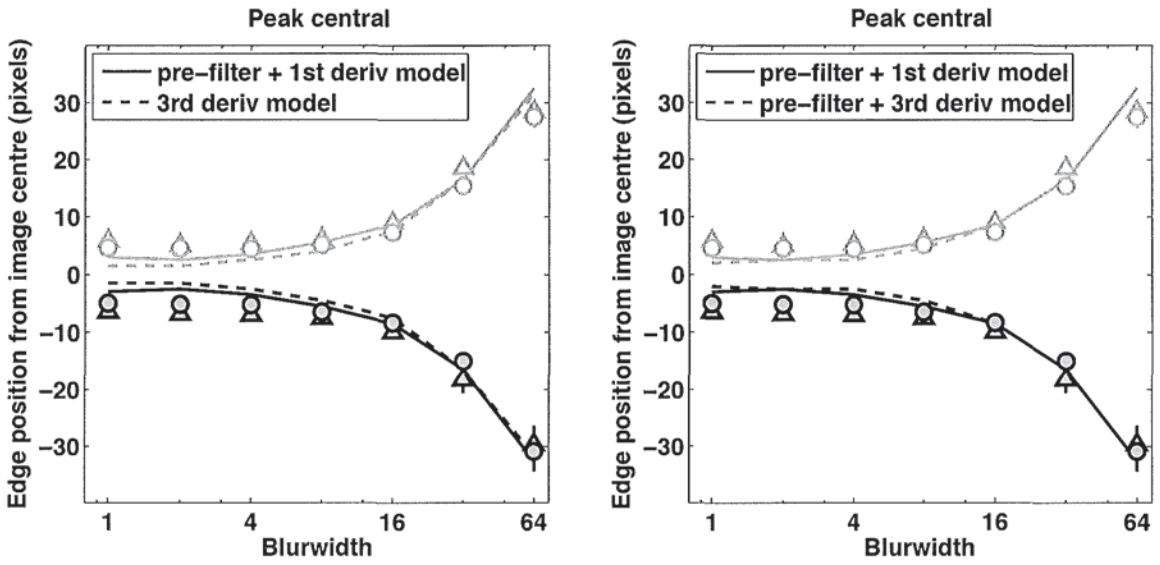

Figure 3.34: Peak central Mach edge data from experiment 1.1 averaged across 3 observers, with predictions of the single-scale 3rd derivative model (wthout blur) and the 'pre-filter + 1st derivative' model. Left and right panels show the 3rd derivative model without and with the pre-filter respectively.

N3+ model with pre-filter, which can account for the data of experiment 1.1 and 2. Thirdly, there is the pre-filter + 1st derivative model, which can also account for the data of experiment 1.1 and 2. The next experiment aims to distinguish between the latter two models, by testing the role of the pre-filter in Mach edge perception.

3.12 Experiment 4.1 - The pre-filter is tested at a high spatial frequency and short duration

3.12.1 Introduction

In experiment 1.1, observers consistently marked the presence and position of Mach edges in blurred triangle-wave images, and the data were predicted by a 3rd derivative model, but not a 1st derivative model. This was taken as support for the 3rd derivative approach. In experiment 2.1, luminance ramps were added to previous stimuli, and the data required the addition of a pre-filter to the model. However, the addition of the pre-filter now means that the 1st derivative approach can account for the experiment 1.1 data, as the pre-filter + 1st derivative model fitted as well as the single-scale 3rd derivative model. Therefore, it can no longer be assumed that experiment 1.1 distinguishes between 1st and 3rd derivative approaches. However, this conclusion rests on the assumption that the pre-filter is used by observers to perceive Mach edges. The aim is to now test whether observers do use such a pre-filter to perceive Mach edges, by manipulating the stimulus in a way that tests for the existence of its inhibitory surround.

With the parameter values used in the recent previous experiments (table 3.1), the pre-filter's MTF

(figure 3.19) shows that it attenuates spatial frequencies below about 4 cycles per degree (cpd). This property is due to the filter's inhibitory surround; if it was absent, the MTF would remain at a high level at low spatial frequencies (Enroth-Cugell et al., 1983).

If the filter is used with stimuli above 4 cpd, the impact of the inhibitory surround should be greatly lessened. This effect should also be enhanced by reducing the stimulus duration, as the inhibitory surround of both magnocellular and parvocellular retinal ganglion cells is relatively less sensitive to higher temporal frequencies, when compared to the cells' central region (Derrington & Lennie, 1984). Furthermore, psychophysical work has found that the low frequency cut in the CSF is rather mild at short durations (25-100 ms) compared with long durations (500-1000 ms) (Georgeson & Georgeson, 1987; Legge, 1978b; Nachmias, 1967)

If these manipulations cause a large effect, then the weight of the inhibitory surround would be reduced to zero. The pre-filter would then become a simple blurring function, and would no longer act like a 2nd derivative operator, as it now has 1, not 3 lobes. Therefore, if observers need the pre-filter to detect Mach edges, then the edges should be less visible or absent at high spatial frequency and brief duration. This experiment aims to test this prediction.

3.12.2 Method

A single interval detection method was used with a high frequency version of the Mach edge images used in experiment 1.1. One set of stimuli were 16 cycles of a 6cpd blurred triangle-wave (512 pixels wide), generated as in experiment 1.1. Waveforms had a blurwidth of 1,2,4 or 8 pixels and were of 2 polarities (peak central or trough central). These waveforms were used to form 1-dimensional grating images. They were also used to generate second type of image by shifting their Fourier phase spectrum by 90 degrees, and leaving the amplitude spectrum unaltered. This produced blurred square-wave waveforms, whose amplitude spectrum declined at $1/f^2$, where f is spatial frequency[2]. The 90 degree phase shift also caused an apparent lateral spatial shift, such that the light and dark bars of the square-wave images did not align with those of the triangle-wave images. In order to restore this alignment, the begining 1/4 wavelength of each blurred square-wave was wrapped to its end. A high contrast trough central example of each type of experimental image is shown in figure 3.35. The left panel shows a triangle-wave image, which appears to contain a thin light (or dark) bar of high contrast at the centre of each wide light (or dark) bar. These thin bars are absent in the square-wave image that is shown in the right panel. A single cycle of the blurwith 1 and blurwidth 8 waveforms is shown in figure 3.36. These blurwidths form the extremes of the range of blurwidths used.

The images subtended 2.67 × 2.67 degrees at the viewing distance of 383cm. They were surrounded by a full screen of of mid-grey luminance and viewed in a dark room. The Michelson contrast of the triangle-wave images was 0.4, while that of the square-wave images was 0.79, 0.79, 0.84 and 0.93 for blurwidths 1, 2, 4 and 8 respectively. The RMS contrast (defined as the ratio of the standard deviation of the luminance profile to mean luminance (Burr et al., 1989)), was the same for both triangle-wave and square-wave images of any given blurwidth.

Prior to data collection, each observer was shown a pair of 3cpd stimuli (blurwidth 8), and it was confirmed that a pair of edges was visible in the centre of each bar in the triangle-wave image but not

[2]The amplitude spectrum of a square-wave that has not been blurred declines at $1/f$.

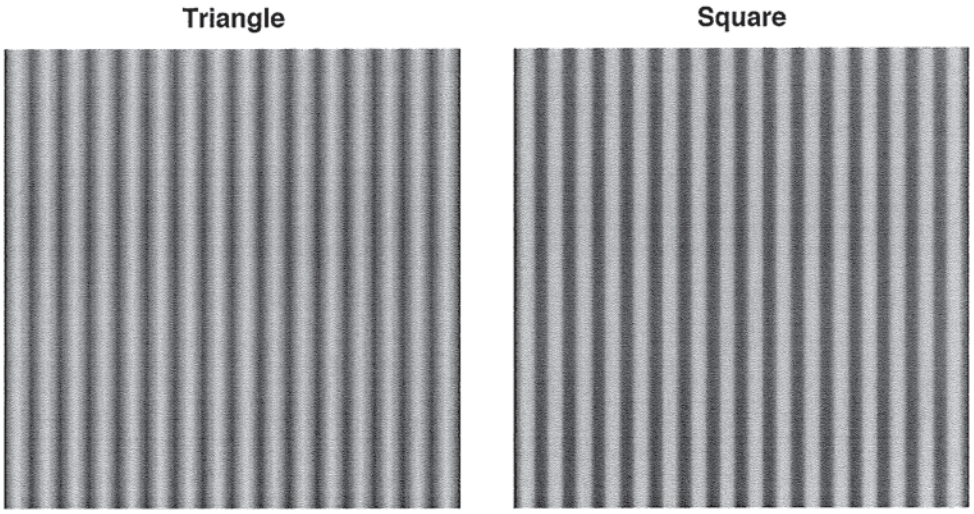

Figure 3.35: High contrast examples of each type of the trough central experimental images, blurwidth 4.

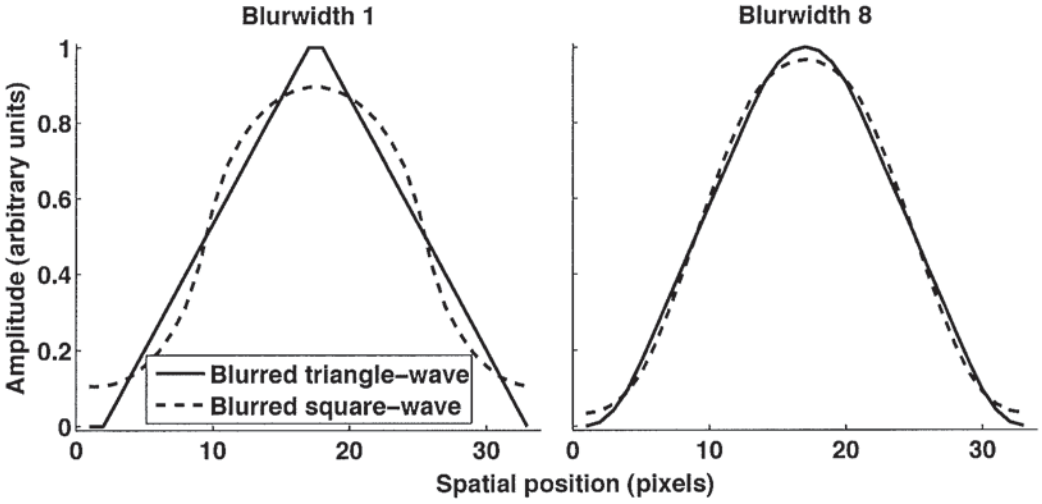

Figure 3.36: One cycle of the blurwidth 1 and blurwidth 8 experimental waveforms. These blurwidths form the extremes of the range of blurwidths used.

in the square-wave image. Each trial consisted of a single stimulus presentation of 50ms, followed by a full-screen of mid-grey with a central small fixation dot (4×4 pixels $= 1.25 \times 1.25$mins). The observer had unlimited time to indicate the presence or absence of a central pair of edges in one or more bars of the stimuli. No feedback was given about the correctness of response because this experiment aims to examine the observers' subjective experience of Mach edges, rather than their ability to distinguish between the triangle-wave and square-wave images per se, which may be based on unspecified image properties (such as the edge blur of the wide bars in the image). The inter-stimulus interval was at least 1sec. Each of the 4 blurwidths, 2 image-types and 2 polarities were shown 15 times in a randomised block. Two observers each completed 9 blocks. The first block from each subject was discarded as practice.

3.12.3 Results

One observer (SEW) was unable to discriminate between the triangle-wave and square-wave images at 6cpd, as all images appeared to contain the central pair of edges. So the viewing distance was reduced to 255cm for that subject (images were 4cpd). For each observer, the proportion of 'edge pair seen' responses for the peak central images was very similar to that of the trough central images, so data were averaged across both image polarities. The resulting data were converted to z-scores and then used to calculate the discriminability (d') of the triangle-wave images from the square-wave images. The calculation was performed by the application of the standard formula (Green & Swets, 1966)

$$d' = z(H) - z(F) \tag{3.6}$$

where H is the proportion of 'edge pair seen' responses to the triangle-wave images (Hits) and F is the proportion of 'edge pair seen' responses to the square-wave images (False alarms). This analysis has the advantage that it factors out observer bias, as the value obtained for d' is independent of the location of the observer's criterion in decision space.

Z-scores of the proportion of 'edge pair seen' responses for each condition, averaged across stimulus polarity are shown as filled symbols in figure 3.37. Separate panels are for each observer. Since each image cycle extended over 1/8 of the number of pixels previously used (experiments 1 to 3) each level of blurwidth shown here represents 8 times the amount of previous blurring. For example, an experiment 1 waveform of blurwidth 16 had the same shape as 1 cycle of a blurwidth 2 waveform used here. Each filled symbol represents 240 trials and their error bars are derived from the standard deviation in binomial sampling. Both observers have a higher proportion of 'edge pair seen' responses for the triangle-wave than the square-wave images. The responses to the square-wave images were higher for

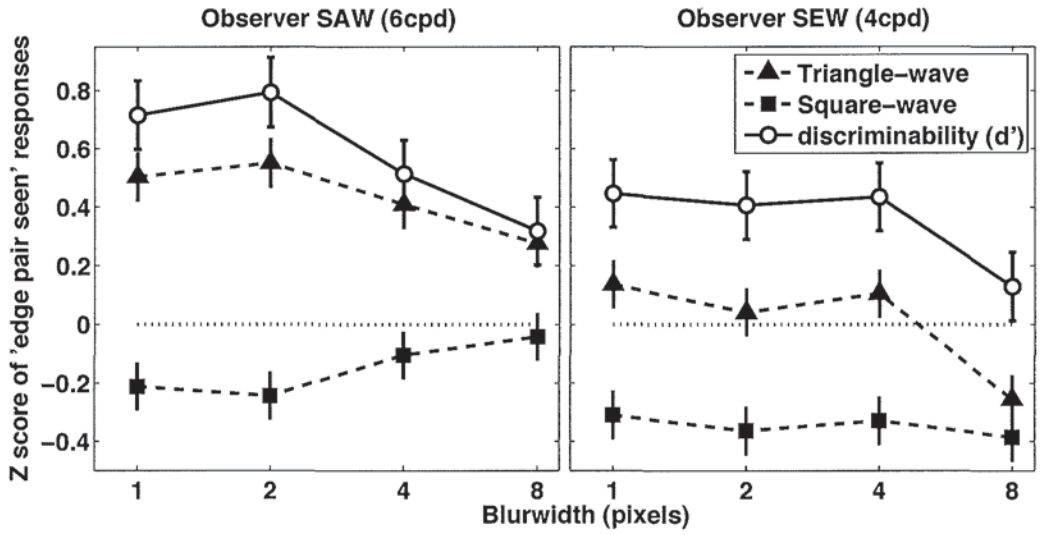

Figure 3.37: Z-scores of 'edge pair seen' responses (filled symbols) and discriminability of image type (open symbols) as a function of blurwidth, averaged across stimulus polarity. Each filled symbol represents 240 trials and their error bars (±1sd) are derived from the expected variance in binomial sampling. Each open symbol represents 480 trials. Error bars here are calculated differently (see text).

SAW than SEW, which suggests a lower criterion in decision space.

The difference between the z-scores of the triangle-wave and square-wave responses, is the discriminability (d'), which is represented by open symbols in figure 3.37. If two statistically independent variables are added or subtracted their variances add. Hence, the d' error-bars were calculated from

$$\sigma = \sqrt{\sigma_t^2 + \sigma_s^2} \tag{3.7}$$

where σ_t and σ_s are the expected standard deviation for the triangle-wave and square-wave z-score data respectively.

The d' data are all positive, which shows that both observers were able to discriminate between the triangle-wave and square-wave images. Both observers show a general trend of decreasing discriminability as blurwidth increases, and the reduction in d' from blurwidth 1 to 8 is similar for both observers (about 0.4 z-units). For each blurwidth, observer SAW's d' values are higher than SEW's, which demonstrates higher discrimination performance.

3.12.4 Experiment 4.2 - A control for Michelson contrast

Observer SAW reported that a change in contrast was observed between some images. In order to test the hypothesis that the discrimination between the triangle-wave and square-wave images was cued by the lower Michelson contrast of the square-wave images, this observer repeated experiment 4.1 with images that were all 0.4 contrast. Z-scores averaged across polarity are shown in figure 3.38 as solid symbols. These results are similar to those of experiment 4.1 (Pearsons's $X^2 = 0.59$, n.s.), which suggests that there was no artefactual effect of contrast.

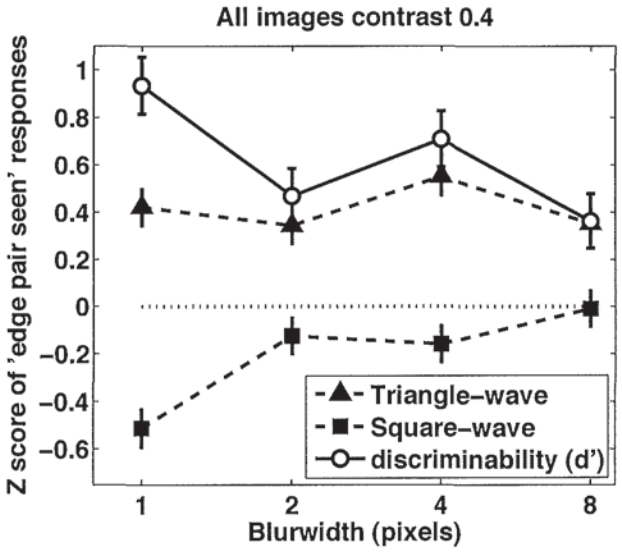

Figure 3.38: Z-scores of 'edge pair seen' responses (closed symbols) and image type discriminability (open symbols) as a function of blurwidth, for images whose Michelson contrast was always 0.4. The observer was SAW. Data are averages across stimulus polarity. Error bars are as figure 3.37.

3.12.5 Discussion

The blurred triangle-wave images used here were high frequency versions of the Mach edge stimuli used in experiment 1.1. The results show that observers were able to perceive a near-central pair of opposite polarity edges in one or more cycles of the image. Their near-central location suggests that they are Mach edges, as there are no peaks in the 1st derivative here (see experiment 1.1 method), and there are 2 extrema in the 3rd derivative at these approximate locations.

The triangle-wave data in figure 3.37 show that Mach edges were reported less frequently as blurwidth increased. This is consistent with a reduction in the amplitude of the 3rd derivative of the luminance waveform as blurwidth increases. The 3rd derivative amplitude is a measure of the rate of change of the curvature of the luminance profile, which can be conceptualised as the degree of sharpness at luminance profile 'corners'. The apex of the blurwidth 1 triangle-wave shown in figure 3.36 has a relatively sharp 'corner' at the junction of each ramp and plateau, whereas the blurwidth 8 waveform (right panel) has a much smoother transition. Thus the amplitude of the 3rd derivative is lower for the blurwidth 8 waveform than the blurwidth 1 waveform, and this would be expected to lead to a lower frequency of Mach edge reports for the blurwidth 8 images.

Mach edges were reported with a lower frequency for the square-wave images than the triangle-wave images. Yet inspection of the blurwidth 4, square-wave image in figure 3.35 shows that Mach edges were absent (other blurwidths were similar). Therefore, the observers' positive responses to the square-wave images at 50 ms duration are likely to represent false alarms in a noisy decision process. This conclusion is supported by the horizontal pattern of the square-wave responses, particularly evident in SEW's data, because the false alarm rate can be expected to be constant across blurwidth, due to the random interleaving of all images between trials. The false alarm rate reflects a bias in the responses, which would contaminate the results if a plot of percent '2 edges seen' had been produced. However, the transformation into z-scores has allowed this bias to be factored out, by way of the calculation of discriminability (d'), using equation 3.6. It appears that SEW has a constant false alarm rate, which suggests that her criterion is fixed relative to the non-signal distribution in decision space. SAW appears to have a 'symmetrical criterion', centred between the non-signal and signal sensory distributions; as d' decreases the 2 distributions move closer together, and the false alarm rate increases.

The pattern of d' values for both observers show a trend of decreasing discriminability as blurwidth increases. This is consistent with increasing similarity between the shape of the triangle-wave and square-wave profiles as blurwidth increases (see figure 3.36). The generally lower d' values for SEW compared to SAW, shows poorer discriminability performance. This may be due to a lower amount of previous familiarity with the stimuli, as SAW had previously undertaken a short pilot study. In order to determine whether SEW's poorer discriminability performance was due to lack of practice, she performed 3 additional experimental sessions at 6cpd. In keeping with her prior attempt at this spatial frequency, the results indicated no discrimination between image types (not shown), which suggests that her 4cpd data correctly reflect visual performance rather than a lack of practice.

The pre-filter was added to the single-scale 3rd derivative model in order to account for the prevalence of perceived edges in blurred triangle-wave images with added luminance gradients above ± 0.5. The revised model provided a good fit to that experiment's data, and the Mach edge data of experiment 1.1. However, both data sets were also well-fitted by a 'pre-filter + 1st derivative' model. Thus the addition of the pre-filter means that experiment 1.1 can no longer be considered to distinguish between

the 1st and 3rd derivative approaches, because competing models based on each derivative both fit that experiment's data. However, the present results will be shown to cast doubt on the pre-filter + 1st derivative model, due to the following explanation.

The present experimental results suggest that Mach edges were reliably reported in the blurred triangle-wave images, which are high-spatial-frequency and short duration versions of the Mach edge stimuli that were used in experiment 1.1. But the use of high-frequency stimuli is expected to reduce the effect of the pre-filter's inhibitory surround, for reasons stated in the introduction to this experiment (section 3.12.1). If the influence of the inhibitory surround is reduced to zero, it no longer accentuates points of high curvature in the luminance profile. Figure 3.39 shows one cycle of the blurwidth 4 luminance waveforms with predictions of the pre-filter + 1st derivative model and the single-scale 3rd derivative model with pre-filter, when the weight of the pre-filter's inhibitory surround is 0.49. Both model predictions have been scaled by different arbitrary values. The model predictions are identical when the weight of the pre-filter's inhibitory surround is reduced to zero, which suggests that the use of high frequency stimuli (of any duration) is sufficient to eliminate any effect of the pre-filter's inhibitory surround. In light of this, the use of short duration stimuli was not strictly necessary, but it should further ensure that the inhibitory surround was ineffective.

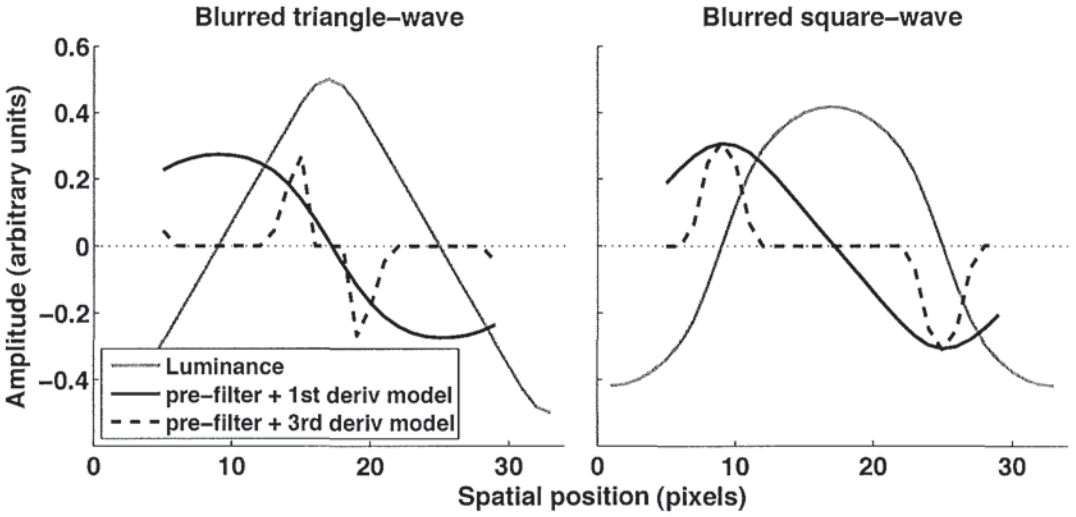

Figure 3.39: One cycle of the blurwidth 4 luminance waveforms with predictions of the pre-filter + 1st derivative model and the single-scale 3rd derivative model with pre-filter, when the weight of the pre-filter's inhibitory surround is 0.49.

The right panel of figure 3.39 shows that both models predict two edges for each cycle of the square-wave, each centred at the mid-point of each luminance ramp, and which correspond to the edges of each bar illustrated in the example experimental image (figure 3.35, right panel). Both models predict no additional edges for this stimulus. The left panel of figure 3.39 shows one cycle of the triangle-wave luminance profile and model predictions. The single-scale 3rd derivative model with pre-filter produces clear extrema that, in keeping with the present results, correspond to a pair of opposite-polarity edges close to the centre. However, the pre-filter + 1st derivative model produces no peaks for this waveform, which means that it predicts no Mach edges. Yet this prediction is contradicted by the experimental results. To conclude, the present results suggest that a pre-filter is not essential to Mach

edge perception, because Mach edges are still perceived when its inhibitory surround is effectively switched off. This clearly undermines the pre-filter + 1st derivative model. The omission of the inhibitory surround is almost equivalent to the omission of the pre-filter completely because its centre has a very small size ($s_c = 1\text{min}$), and so has only a small blurring effect.

The Mach edge data of experiment 1.1 were fitted equally well by two competing models; the single-scale 3rd derivative model with pre-filter, and the pre-filter + 1st derivative model. However, the latter model has been undermined by the present experimental results, which suggests the single-scale 3rd derivative model should be the favoured model. Therefore, the Mach edge results of experiment 1.1 can once again be regarded as support for the the 3rd derivative approach and not for the 1st derivative (plus pre-filter) approach.

The addition of the pre-filter to the 3rd derivative model was required to account for the added luminance ramp data of experiment 2. Yet the pre-filter was not required when fitting the experiment 1.1 data. So the question arises whether an alternative explanation, which does not invoke a pre-filter, can account for the experiment 2 data. One possibility is that Mach edges can be detected and located by first detecting a bar (the Mach band), and then inferring the location of its two edges from the bar's scale (blur). The perception of Mach bands will be addressed next.

3.13 Chapter summary

A new phenomenon was introduced, named 'Mach edges', which are perceived edges where there are no peaks in the 1st derivative or zero-crossings in the 2nd derivative. Their existence and perceived positions were nicely predicted by a non-linear 3rd derivative model.

The model was tested further with a new set of Mach edge stimuli with a luminance ramp. Modelling of the results required the addition of a pre-filter, whose properties were based on known ganglion cell physiology. Unfortunately, this addition introduced the pre-filter +1st derivative as a viable alternative to the pre-filter + 3rd derivative when modelling the data from both experiments.

High spatial frequency stimuli of short duration were then used to attenuate and by-pass the putative pre-filter's inhibitory surround. Mach edges were still perceived, and they were predicted equally well by the 3rd derivative model, and the pre-filter + 3rd derivative model, but not by the 1st derivative model, nor the pre-filter + 1st derivative model. The conclusion reached was that the 3rd derivative model defeated the 1st derivative model as an account of Mach edge perception.

However, the possibility was raised that Mach edges are detected indirectly, via the Mach band of which they are a component. The next chapter addresses the perception of Mach bands.

Chapter 4

Mach bands

4.1 Introduction

It seems intuitively obvious that we will perceive lightness where the luminance is high and darkness where the luminance is low. But this is not always the case. In the 1860's Ernst Mach noticed that illusory light (or dark) regions could be perceived in the absence of increases (or decreases) in luminance (Ratliff, 1965). Although initially observed on rotating discs, Mach found that these 'Mach bands' (as they were subsequently known) also existed in static images at the junction of a luminance ramp and plateau, as shown in figure 4.1. This luminance profile is sometimes known as a 'Mach ramp'. Mach bands are a strong and reliable illusion, and are not restricted to vision. Experiments with shearing forces, vibrations, and heat stimuli on the skin and with taste stimuli on the tongue also show Mach band effects (von Bekesy, 1967). It is important to understand such ubiquitous phenomena, and to draw appropriate conclusions about feature detection in human vision. Crucially, their existence shows that not all perceived bars arise from peaks in the luminance profile.

While attempts to measure the width, positions and brightness of Mach bands have shown substantial variability (Ratliff, 1965), some general characteristics have been determined as follows (Lowry & DePalma, 1961; Watrasiewicz, 1966; Fiorentini, 1972). An increase in the gradient of the Mach ramp, or an increase in the luminance of the light plateau causes an increase in the Mach bands' strength; the light band appears lighter and the dark band appears darker. These manipulations also narrow the apparent width of the light band, but the dark band width remains largely unaltered. Several researches have measured the subjective brightness across Mach bands by matching apparent brightness to an adjacent thin bar (von Bekesy, 1968; Charman & Watrasiewicz, 1964; Thomas, 1966; Lowry & DePalma, 1961; Watrasiewicz, 1966). They found an asymmetry between the dark and light bands. The light band appears brighter and narrower than the dark band and was centred on the light plateau-ramp junction. The wider and less intense dark band was centred on the dark-plateau-ramp junction or more towards the dark plateau.

However an asymmetry in the opposite direction was reported by Whiteside & Davidson (1971) who found that a linear Mach ramp caused a more pronounced dark than light band. The bands appeared more symmetrical if the luminance ramp was exponential rather than linear, which suggested a logarithmic transformation of the image prior to a stage that generated the bands. Whiteside &

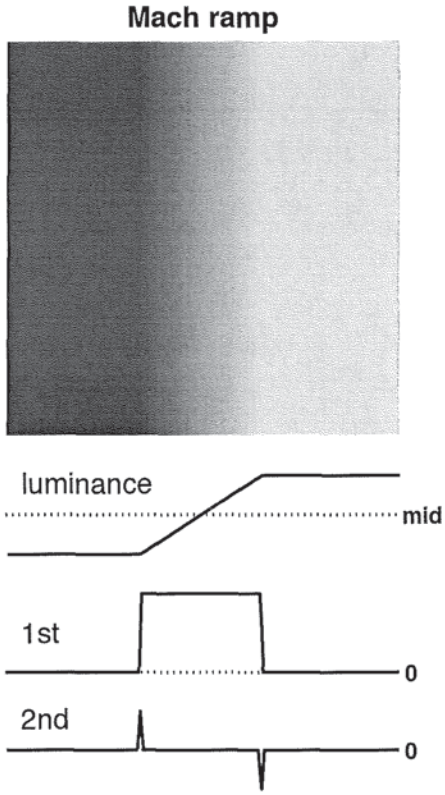

Figure 4.1: A Mach band stimulus, its luminance profile and first two derivatives. The reader may observe a dark bar at the 2nd derivative peak and the light bar at its trough. These are entirely illusory, as there are no corresponding troughs or peaks in the luminance profile.

Davidson stated that other researchers had reported results that conflicted with theirs: light bands appeared more prominent than dark bands when a matching method had been used. But Whiteside & Davidson resolved the apparent inconsistency by pointing out that both the stimulus and the matching thin bar would be equally affected by the logarithmic transformation. The atypical dark-light asymmetry was also reported by Gur & Syrkin (1993), who found that a lower contrast Mach ramp was required to detect its dark Mach band than its light Mach band. But this conflicting result is likely to be due to the different method employed, which was detection rather than matching.

Mach attributed Mach bands to a process of lateral inhibition, whereby the response of each retinal cell is inhibited (divided) by by its neighbours by an amount inversely proportional to the distance between cells (Ratliff, 1965). This was approximated by the sum of the luminance and its (negative) 2nd derivative. Mach's proposal was a striking anticipation of the later physiological discoveries of lateral inhibition in the retina, described next.

The lateral inhibition theory has received much physiological support. Ratliff & Hartline (1959) examined the response of a single ommatidium of the Limulus crab eye at various locations across step edge and Mach band images. For both image types, the cells response profile was of a Mach band form, with under- and overshoots on the dark and light sides (respectively). But these overshoots were abolished when the neighbouring ommatidia were covered, thus providing evidence for the role of lateral

inhibition in Mach band perception. Enroth-Cugell & Robson (1966) found a similar response pattern to step edges in cat X cells, and deduced a linear lateral inhibition process, from this finding and from responses to sine-wave gratings. In primates, the centre-surround mutually-antagonistic receptive field structure of retinal ganglion cells is well established (De Monasterio & Gouras, 1975; Croner & Kaplan, 1995; Chichilnisky & Kalmar, 2002).

Ratliff (1965) examined Mach's model and 5 others that were all based on lateral inhibition. One model was non-linear, as the response was divided by the inhibitory component, and the other 4 were linear, as the inhibitory component was subtractive. Ratliff found that despite their mathematical differences, all models provided an adequate fit to experimental data. So he concluded that they all could be considered as approximations of a more generic lateral inhibition model.

4.1.1 Mach bands at a step edge

An implicit assumption of lateral inhibition models is that perceived lightness or brightness is relatively unaffected by visual processing beyond the retina. The models assume a simple direct mapping from the retinal response to final percept. It is through this assumption that the models explain Mach bands, but its presence leads to the problematic prediction that Mach bands should be strongest at a step edge. A 1-dimensional slice across a retinal ganglion cell's receptive field profile yields a 3-lobed shape that approximates a 2nd derivative operator. Even though the 2nd derivative of a step edge is infinitely high, optical blurring of the retinal image transforms the step edge to one of finite width, and finite derivative amplitudes that are very large. However, experimental evidence for Mach bands at step edges has been limited.

von Bekesy (1968) used flickering step edge stimuli interleaved with a red marker bar to obtain Mach band position and brightness. He found that the light Mach band was stronger than the dark Mach band, and attributed this difference to the effect of a compressive transducer. He concluded that the transducer must be located before the filter that generates the Mach bands (an approximation to a 2nd derivative filter, but with a dc component), because if it was located after the filter, the light and dark Mach bands would have the same strength. However, a compressive transducer has the opposite effect to that intended by Bekesy. It *decreases* the luminance waveform's curvature at the light Mach band location, so reducing its 2nd derivative response. Thus it would reduce the strength of the light Mach band, not the dark one. This invalidates Bekesy's conclusion, and suggests that the transducer is accelerative, not compressive. In a study that employed a 2AFC procedure, rather than the more common single interval, Henning et al. (2000) found that a light, but not a dark, Mach band was seen adjacent to a step edge.

Most researchers, however, have failed to observe both Mach bands at a step edge (Fiorentini, 1972; Ratliff, 1984). For example, Ross et al. (1981) found that Mach bands were not perceived for luminance ramp widths below 4mins. This absence has prompted several theories, which attempt to simultaneously account for this phenomenon and the presence of Mach bands at a Mach ramp.

4.1.2 Mach band theories

This section outlines some competing theories that account for Mach bands, which are:

1. Inhibition of bars by edges (Tolhurst, 1972b).
2. MIRAGE (Watt & Morgan, 1985).
3. Local energy (Morrone & Burr, 1988).

1. Inhibition of bars by edges

An interesting explanation for the absence of Mach bands at step edges was provided by Tolhurst (1972b). He discovered that prior adaptation to an edge increased detection thresholds for both polarities of edge, but more so for same polarity than opposite polarity edges. To explain this finding, he proposed the existence of odd-symmetric 'edge detector' mechanisms, which were separate from even-symmetric 'bar detectors'. It might be though that his results could be explained purely by the action of bar detectors (without recourse to edge detectors at all), because adaptation to a DL edge would affect dark bar detectors that are located to the left of the edge and light bar detectors that are located to the right of the edge. However, this account can be rejected because Tolhurst instructed his observers to move their eyes around the image during adaptation. This movement would affect dark bar detectors on both sides of the retinal field, not just those on the left side, when adapting to a DL edge.

An implication of this eye movement is that the adaptation was cortical rather then retinal, because each retinal location received an input that flipped between low (dark side of edge) and high (light side of edge) at random temporal intervals. Therefore the eye-movement instruction (contrary to previous accounts of lateral inhibition) shifted the locus of investigation from the retina to the cortex. Tolhust supposed that a step edge would maximally stimulate an edge detector, and non-maximally stimulate adjacent bar detectors. This might be expected to cause the perception of Mach bands at all step edges, but Tolhurst also proposed that the edge and bar mechanisms are mutually inhibitory over a limited spatial extent. This explained a number of brightness illusions, including Mach bands. A luminance ramp can be considered to be a blurred edge, centred at its mid point. If the Mach ramp is wide, then this central edge is relatively far from the Mach bands and does not inhibit them. But if the ramp is narrow, then the central edge is close enough to inhibit the Mach bands.

This account inspired Ratliff (1984) to test the further prediction that the addition of a narrow biphasic bar (a light bar adjacent to a dark bar) at the centre of the Mach ramp would attenuate the Mach bands, as it is a strong edge stimulus. By varying the bar's contrast he hoped to control the amount of inhibition that it exerted on the Mach bands. An increase of contrast did indeed decrease the Mach bands' widths, ultimately to zero.

This Tolhurst-Ratliff account has been criticised by Pessoa (1996) for two reasons. Firstly, it relies on multiple scales. Only a broad scale odd-symmetric filter would respond maximally to a Mach ramp and signal an edge, but the Mach bands themselves are generated by fine scale even-symmetric filters centred at the ramp-plateau junctions. The use of multiple scales for both types of filters would generate spurious features at incorrect locations. For example, as the scale of an even-symmetric filter increases, it would provide peak positions (Mach band locations) that drift outwards from the ramp-

plateau junction. While this objection does not rule out multiple scales *per se*, it does highlight that the combination across scale requires specific rules.

Pessoa's second criticism of the Tolhurst-Ratliff account is that it is inconsistent with the experimental data of Ratliff et al. (1983), who found that attenuation of Mach bands was equally affective with biphasic or single bars placed on the luminance plateau near the Mach band. Pessoa argued that while the biphasic bars would strongly stimulate the odd-symmetric filters that attenuate the Mach bands, the single bars would stimulate the even-symmetric filters and so "should not attenuate Mach bands". However, this objection fails to recognise that Ratliff's bars were sharp-edged, and therefore consisted of two neighbouring edges of opposite polarity, one of which would be closer to the Mach band than the centre of the bar.

2. MIRAGE

The MIRAGE model (Watt & Morgan, 1985) was described in detail in Chapter 1 (section 1.8.4), so a summary follows. The image is filtered with a bank of 2nd derivative filters at four scales. The resulting output is split into positive and negative portions, and then combined across scale. Then two types of 'primitive' are obtained: a region of inactivity, and a zero-bounded response. Then parsing rules are applied to obtain feature position and type. A region of activity flanked by two regions of inactivity or two regions of activity indicates a bar. A region of activity abutting only one region of inactivity indicates the boundary of an edge.

A Mach ramp causes a region of activity centred on each ramp-plateau junction, both of which are flanked by inactivity. Therefore, each signals a bar, which is seen as a Mach band (figure 4.2, left). As the ramp width is reduced, the regions of activity converge. When they meet, there is no longer a region of inactivity between them, but they are still each abutting a region of inactivity on their less

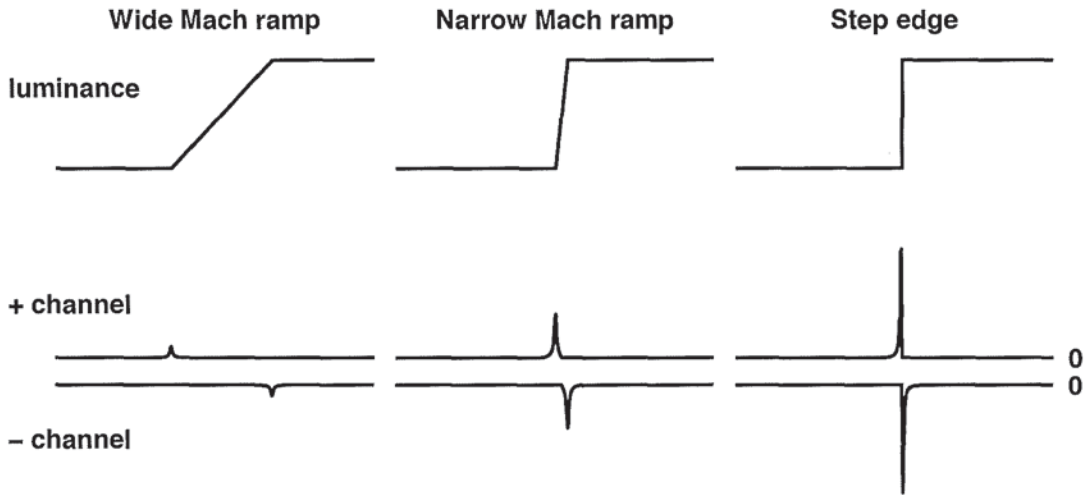

Figure 4.2: Output of an implementation of MIRAGE, showing the regions of activity in the positive and negative channels. The magnitude of each channel's output (area within each peak or trough) increases as ramp width decreases.

central side, so the model outputs a single blurred edge. Thus the absence of Mach bands is predicted for narrow ramps. As the ramp width reduces to zero, the waveform becomes a step edge, and the two peaks of activity converge. Each edge border is located at the 'centroid' of the region of activity, which is a weighted mean of the response. So the model still predicts a blurred edge but its blur is now reduced because the two centroids are closer to each other. The 'strength' of the predicted Mach bands is supplied by the area within each region of activity (termed 'mass' by Watt & Morgan (1985)). As the ramp width decreases, the mass increases.

The critical narrow ramp width that dictates the appearance and disappearance of Mach bands has been predicted by using MIRAGE to be 5mins (Watt & Morgan, 1985; Watt, 1988). This is in good agreement with the experimental findings of 4mins (Ross et al., 1989), and about 6mins (Ross et al., 1981).

However, MIRAGE's parsing rules may exclude visible features. It signals no central feature on a Mach ramp, in the presence of Mach bands. Yet, providing the ramp's gradient is not too low, the ramp is clearly visible in addition to the Mach bands (Kingdom & Moulden, 1992). Therefore, at a broad scale, the ramp may appear as a blurred edge. For example, the DL edge in figure 4.1 may appear to the reader to contain a central edge at a very broad scale (which is more apparent when viewed from a large distance). Experiment 5.3 will examine whether this central edge is, in fact, perceived.

The MIRAGE implementation illustrated in figure 4.2 shows that it predicts that Mach bands should increase in strength as the ramp width decreases. Yet their apparent strength to observers reduces to zero at a ramp width of 4mins (Ross et al., 1989), and Pessoa (1996) reports that Mach bands become weaker as ramp width decreases.

Furthermore, Kingdom & Moulden (1992) point out its limitations when applied to another brightness illusion: the Chevreul illusion (figure 1.9). MIRAGE's predictions for the chevreul illusion vary according to the spatial scale of the stimulus (Kingdom & Moulden, 1992); at one scale it predicts a series of edges, and at another a series of bars, but never both.

A variation of MIRAGE, called MIDAAS uses similar filters and parsing rules, but the combination across scales is shifted forward a stage, to follow the application of the parsing rules (Kingdom & Moulden, 1992). The resulting predictions capture many brightness effects including Mach bands. However, it was not intended as a model of feature detection, but merely as a preliminary exploration of a brightness perception model. Despite these limitations it is interesting that MIRAGE and MIDAAS accomplish so much with filters of only one type, and which are 2nd derivative operators.

3. Local energy model

The local energy model (Morrone & Burr, 1988) was described in detail in Chapter 1, so a summary follows. It uses matched pairs of even- and odd-symmetric filters to obtain a measure of the energy of the input image. Energy peaks denote features, whose identity (edge or bar) and polarity are determined by the filter type that has maximum amplitude at the energy peak. If the even and odd filters have similar amplitudes, then the energy peak is modelled as both an edge and bar which are perceived simultaneously or as a single ambiguous feature that flips between edge and bar. A Mach ramp image results in energy peaks at the middle of the ramp (a peak in the odd-symmetric channel, so signaling an edge) and at the junctions of ramp and plateau (peaks in the even-symmetric channel, so signaling bars). At a step edge, the model produces a single energy peak, centred on the edge, which

signals an edge without flanking bars.

This model has provided excellent fits to quantitative Mach band data (Ross et al., 1989) and a modified version of the Chevreul illusion (Morrone et al., 1994). It also provided a good fit to the Mach edge data from experiment 1 (Chapter 3), although that may be largely due to the use of filters that approximate 3rd derivative operators. However, it failed to predict the majority of features seen by observers in grating stimuli composed of a fundamental and its 3rd harmonic (Georgeson & Freeman, 1997) and in images composed of (partially blurred) square-waves with various amounts of applied phase shift (Hesse & Georgeson, 2005). More fundamentally, it predicts no features at all for a sine-wave grating, despite its stripy appearance (Morrone & Burr, 1988). Thus, the energy model appears to have limited application to feature detection.

4.2 Interim summary and way forward

The Mach band phenomenon has been attributed to a spatial lateral inhibition process, whereby the response of the visual system at each point in space is reduced by an amount that depends on the luminance in a small surrounding region. This process can be modelled by using a bank of filters that are 2nd derivative operators at a range of scales. Models based on this approach have generally provided a good fit to Mach band data. Manipulations that affect the retinal image 2nd derivative amplitude (such as a change in Mach ramp width or contrast) also affect the brightness and/or width of the Mach bands. However one manipulation that has received little attention is a change of curvature at the ramp-plateau junction. Since the 2nd derivative is a measure of curvature, this manipulation should affect Mach band perception, even when the ramp width and contrast remain unaltered. The next experiment aims to test this prediction.

4.3 Mach band detection using a generalised Gaussian image set

4.3.1 Introduction

Almost all of the previous experiments in this thesis examined the perception of Mach edges, which are perceived edges where there is no peak in the 1st derivative of the luminance profile. They are so named because they are analogous to the classic Mach bands, which are perceived bars where there is no luminance peak. We now turn our attention to Mach bands. The present experiment will examine the characteristics of Mach band perception with a novel set of images that contain various amounts of curvature at the ramp-plateau junctions. The amount of curvature can be expressed as 2nd derivative amplitude, so the main experimental aim is to test the prediction that Mach band strength is related to 2nd derivative amplitude.

4.3.2 Stimuli

A novel set of stimuli was generated from a family of waveforms that vary from the Mach ramp through to the Gaussian integral, and beyond, under the control of a single parameter. The starting point for

stimulus generation was a set of 'generalised Gaussian' curves, which were defined as follows.
A Gaussian of unit area, centred at 0, is given by

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

where σ is standard deviation. This Gaussian can be considered to be one of a family of 'generalised Gaussians', where the exponent 2 is replaced with another real number, n

$$\frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-|x|^n}{2\sigma^n}\right) \tag{4.1}$$

Each generalised Gaussian was integrated and scaled to a common amplitude to form the luminance profile of a 1-D blurred vertical edge. Image sets of differing blurs were obtained by setting σ to 3, 6 or 12mins. The levels of n were determined from a pilot study as 1, 1.5, 2, 2.5, 3, 4 or 5. Increases (or decreases) of n sharpen (or blur) the upper and lower 'corners' of the waveform, as shown in figure 4.3 (right). Thus the family of waveforms ranged from a slightly blurred Mach ramp (n=5), through a Gaussian edge (n=2), to a very blurred edge (n=1). The 1st derivatives (figure 4.3, left) ranged from a function that was one cycle of a slightly blurred square-wave (n=5), through a more blurred square-wave (n=2), to a discontinuous function that rose gradually to a central sharp peak and then fell similarly (n=1).

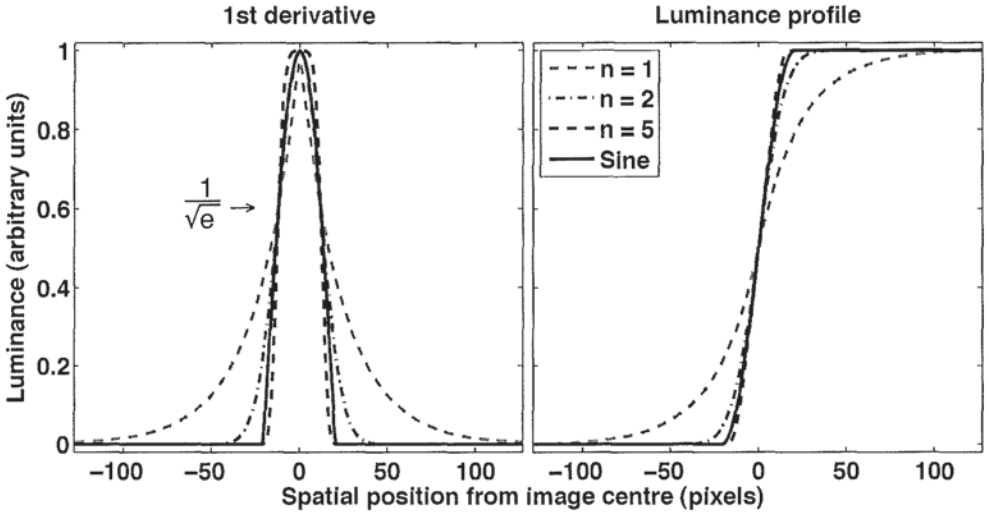

Figure 4.3: Stimuli luminance profiles (right panel) and their 1st derivatives (left panel) for various levels of exponent (n). Some intervening levels have been omitted for clarity. The 1st derivatives functions cross at $1/\sqrt{e}$. Only polarity DL is shown here.

A sine edge does not readily belong to a family of related curves, but it was included in the present study because previous scale-space modelling had suggested that Mach bands would be observed on it, and that no Mach bands would be observed on a Gaussian edge that was superficially similar (Georgeson, 2006). Thus it may help to delimit Mach band perception. The sine edge was generated by integration of one cycle of a cosine function. The frequency of the cosine function was set to ensure that it passed through $(\sigma, 1/\sqrt{e})$. This point forms the common crossing point of all the generalised Gaussian curves (see figure 4.3 (left)), because when $x = \sigma$ is substituted into the formula for a Gaussian of unit height,

which is

$$\exp\left(\frac{-x^2}{2\sigma^2}\right)$$

the resulting formula is

$$\exp\left(\frac{-x^2}{2x^2}\right)$$

which simplifies to

$$\exp\left(-\frac{1}{2}\right)$$

The waveforms were inverted to form opposite polarity conditions (LD edges). One set of images consisted of the 16 images (8 waveforms, 2 polarities) for a single value of σ. The images subtended 4.26 degrees at the viewing distance of 123.4cm. Thus 1 pixel subtended 1 minute. The images were viewed at a Michelson contrast of 0.4 and were surrounded by a full-screen mid grey of luminance $40.7cd/m^2$. High contrast examples of two experimental images are shown in figure 4.4.

Perception of these images was examined with two methods; a Yes-No method, to obtain detection data, and a feature-marking method, to obtain position and width data.

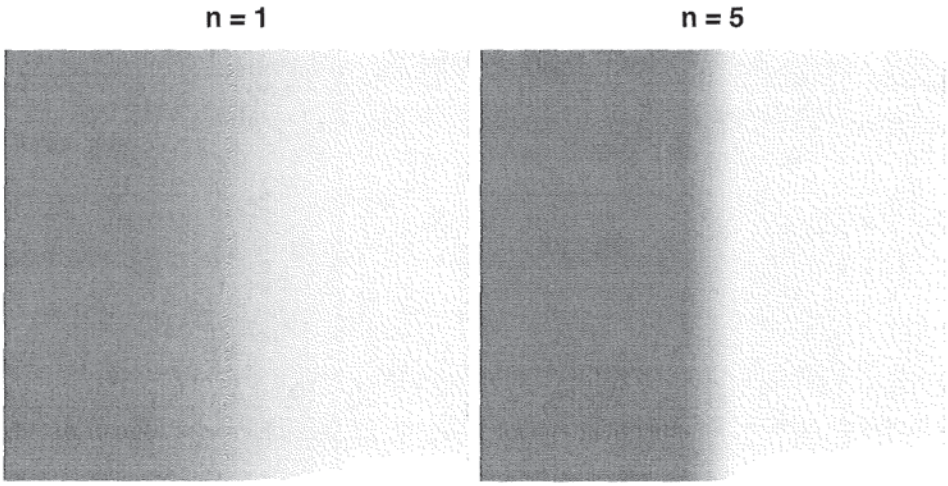

Figure 4.4: High contrast examples of two DL experimental images, generated with exponent (n) values of 1 (left) and 5 (right). σ=12mins. The reader may observe Mach bands in the right panel that are absent in the left panel.

4.4 Experiment 5.1 - Yes-No method applied to Mach band images

4.4.1 Method

The single-interval Yes-No method of experiment 2.1 was applied here, except that the task was to indicate the presence or absence of Mach bands. Observers were instructed to respond positively if they could see one or both Mach bands in each image. A classic Mach band image was shown to observers

prior to data collection, as training. All 16 images from each of the 3 image sets were shown 5 times in a randomised order within one experimental session, which took about 10 minutes to complete. Six observers each completed 7 sessions, but the first session was discarded as practice.

4.4.2 Results

For each observer, the proportion of perceived Mach bands was calculated for each condition and plotted against the generalised Gaussian exponent. The observers' responses to the generalised Gaussian images were very similar for both stimulus polarities, so means are plotted in figure 4.5.

The curves in figure 4.5 show the general trend of a relatively smooth and generally monotonic increase of the proportion of 'Mach band seen' responses, as generalised Gaussian exponent increases.

Mach bands were perceived more often for scale 3 images than scale 12 images for observers NRH and TAY. This trend is also visible in the low exponent data for SAW and RJS, but is reversed at higher exponent values. The data for each scale are very similar, for observers MAG and DHB. Overall, there appears to be only a small effect of scale. This is more clearly seen when the data are averaged across observer (see figure 4.9).

The sine edge data for all 6 observers have been inserted at exponent 3.42 in figure 4.5, by a method described below. Figure 4.5 shows that the placement of the sine edge data at this exponent value provides a good fit with the rest of the data, and captures the variation across observers particularly well.

4.4.3 Sine edge data

Although the sine edge was not a member of the generalised Gaussian family, its data were inserted within the generalised Gaussian dataset in order to allow subsequent examination of its properties in relation to those of the generalised Gaussians. For example, it may be the case that the data suggest insertion at a generalised Gaussian exponent value where the sine edge's 2nd derivative has the same amplitude as the generalised Gaussian.

The sine edge data were inserted at exponent 3.42. This value was obtained by firstly interpolating between data points as follows. A Naka-Rushton curve was fitted individually to each of the 18 data sets (6 observers, 3 σ values) by using a downhill Simplex algorithm ('fmins' in Matlab) to minimise the RMS error between data and curve. The Naka-Rushton equation was of the form

$$f(x) = k\frac{x^p}{x^p + s}$$

where k, p and s were all free parameters and x was the value of the generalised Gaussian exponent. The fits were good (rms error= 0.027), as shown in figure 4.6. Parameter values are not stated here, as this fit was used only as a method of interpolation. Next the sine edge data were fitted to the Naka-Rushton curves, with the sine edge exponent value (n) constrained to be the same for all data sets. The fit was good (rms error = 0.013), as shown in figure 4.5. Put simply, this means that across the 18 datasets the sine edge consistently behaved like a generalised Gaussian edge with an exponent of 3.4.

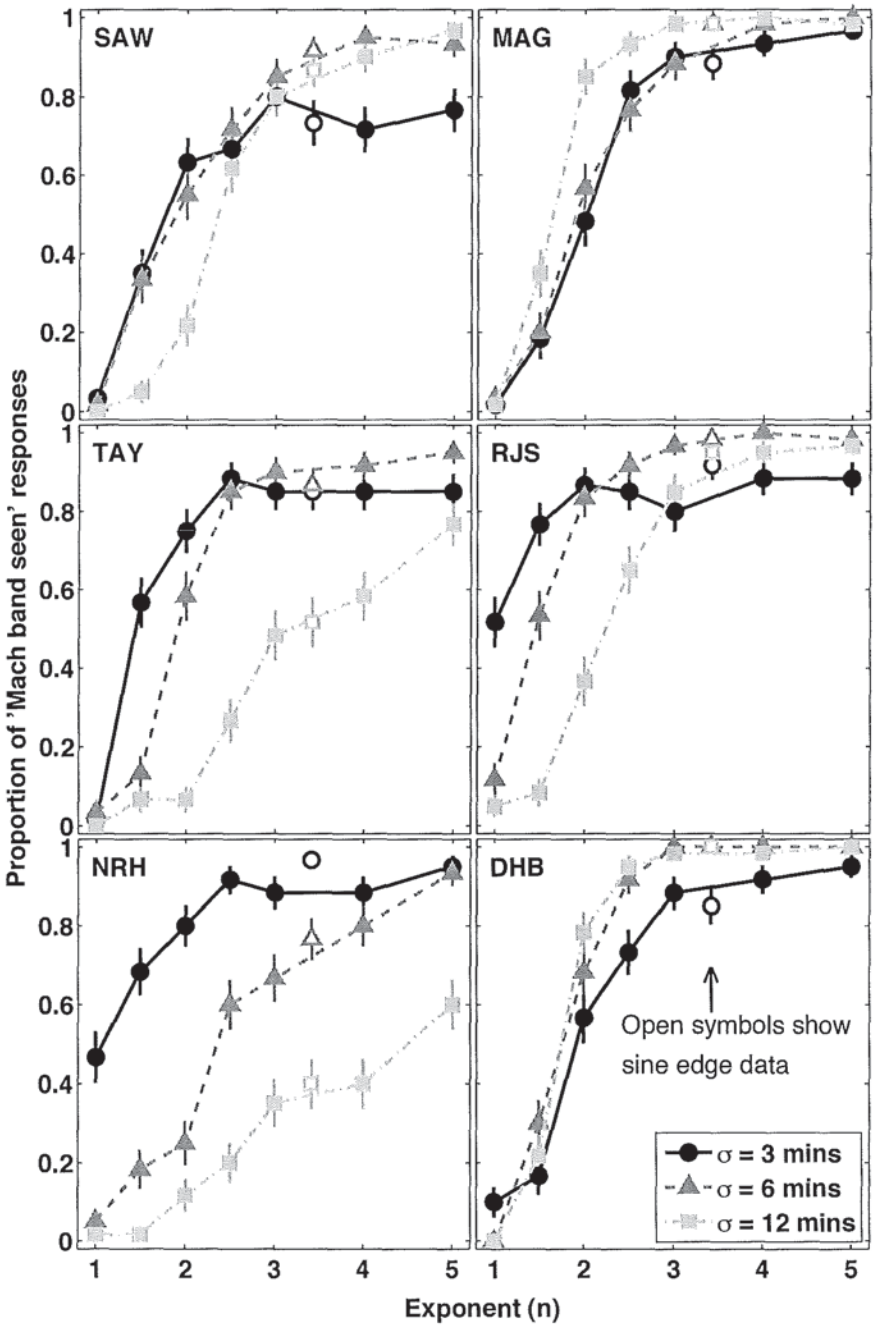

Figure 4.5: Proportion of Mach bands seen in generalised Gaussian images of various exponents and a sine edge. The panels show data from each observer, pooled across polarity. Each closed symbol represents 60 trials, and symbol shape and shade denotes image scale (σ). The sine edge data (open symbols) have been inserted at exponent 3.42 (see text for details). Error bars ($\pm$1sd) are calculated from the expected variance in binomial sampling.

4.5 Discussion

The curves in figure 4.5 are relatively smooth and generally monotonic. This suggests that these images form a coherent set, in which the variation of the generalised Gaussian exponent leads to a systematic

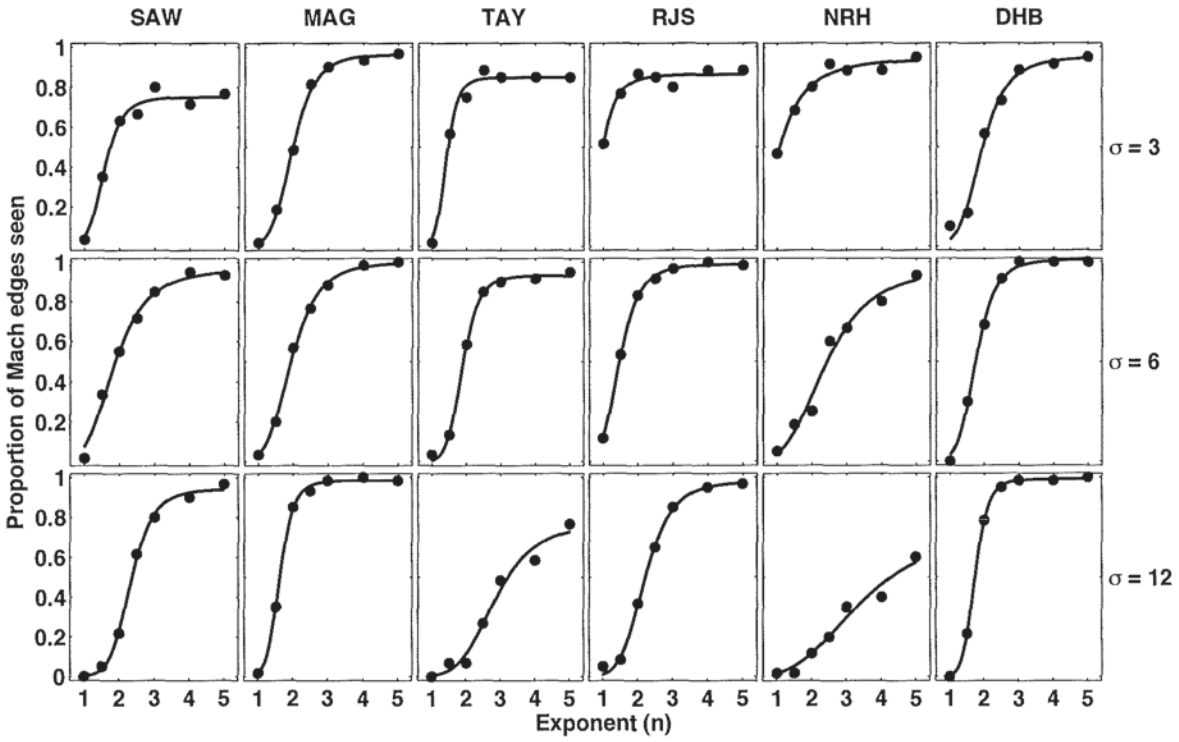

Figure 4.6: Fit of Naka-Rushton equation (lines) to generalised Gaussian data (symbols). Columns show data from each subject and rows show each image scale. This fit is used as a method of interpolation.

change in the Mach band percept. This is a non-trivial result because prior to this experiment it was not clear that the progressive manipulation of the generalised Gaussian exponent would lead to an ordered pattern of responses at all.

The curves that are not quite monontonic are all of scale 3, and they begin to fall at exponent 2.5 or 3. This suggests a performance ceiling, which may correspond to the narrow width of the generalised Gaussian edges. Mach bands are not reported in a classic Mach band image (where the luminance ramp-plateau junction is sharp) when the Mach ramp width is below 4mins (Ross et al., 1989), or about 6mins (Ross et al., 1981). In order to investigate whether this constraint applies to the scale 3 images used in the present experiment, a Mach ramp waveform of width 6mins was blurred with a Gaussian function of spread = 1.25. This amount of blurring was applied in the N3+ model, which provided a good fit to the Mach edge and step-edge feature-marking data of experiments 1.1 to 1.3. It is envisaged to represent the combination of optical blur and the scale of the smallest filter used by the observer (neural blur). The blurred Mach ramp waveform was plotted alongside the scale 3 generalised Gaussian waveform of exponent 3 (see figure 4.7). The close similarity between the two profiles suggests that the present results are constrained by the same lower limit on Mach ramp width that was reported by Ross et al. (1981, 1989).

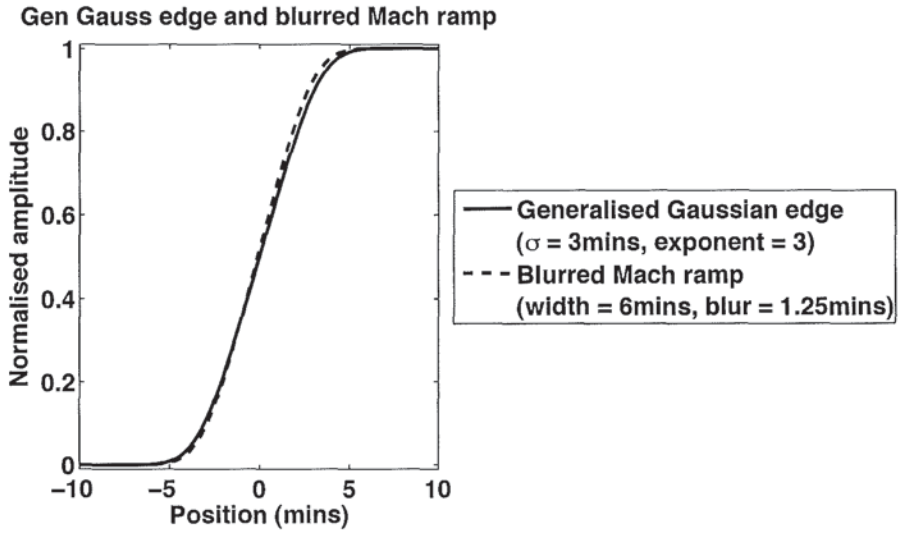

Figure 4.7: The scale 3, exponent 3 generalised Gaussian waveform and a Gaussian blurred ($\sigma = 1.25$) Mach ramp waveform.

4.5.1 Models

the main experimental aim was to test the prediction that Mach band strength is related to 2nd derivative amplitude. In order to examine this relation in detail, several models were fitted to the data. All of them use one or more filters that are, or approximate the shape of, 2nd derivative operators. The models explored in this section are:

1. Single-scale 2nd derivative
2. Scale-space edge-bar
3. Single-scale 3rd derivative, with pre-filter
4. Local energy
5. MIRAGE

1. Single-scale 2nd derivative model

The central aim of this experiment was to test the prediction that Mach band strength is related to 2nd derivative amplitude. So the amplitude of the 2nd derivative of the luminance profile was calculated at a very fine scale, for each experimental waveform. A plot of 2nd derivative amplitude against z-scores of observers' responses is shown in figure 4.8. Z-scores were plotted rather than percent correct, in an attempt to apply Thurstonian scaling (Thurstone, 1927), which can test whether the observers' internal responses in decision space are linearly related to the waveforms' 2nd derivative amplitude. Z-scores are obtained by applying an inverse Gaussian integral function to the data. If the percentage data have the shape of a Gaussian integral then the resulting z-score plot will be linear. The z-score data in figure 4.8 are approximately linear, except for the exponent 4 and 5 data, which are slightly too low.

The 2nd derivative amplitudes were all multiplied by 1.5, in an effort to align with the $\sigma = 6$ data, by eye. It is clear from figure 4.8, however, that a change in waveform scale affects the 2nd derivative

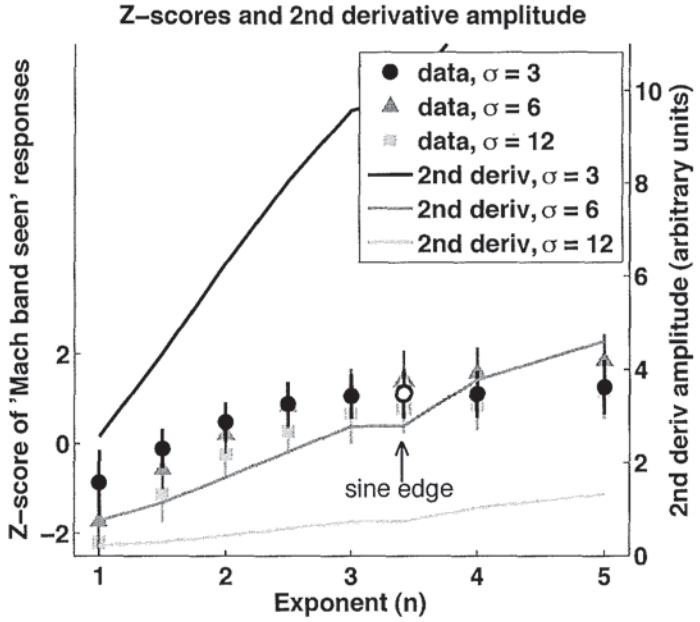

Figure 4.8: Z-scores of 'Mach bands seen' responses against generalised Gaussian exponent (symbols), with the amplitude of the 2nd derivative of each luminance waveform (lines) multiplied by 1.5. The value of σ denotes the image scale. Error bars ($\pm$1sd) are calculated from the expected variance in binomial sampling.

amplitudes much more than the experimental data. This demonstrates the failure of the single-scale approach and suggests that consideration of the 2nd derivative amplitude alone cannot capture the way that Mach band perception is relatively scale-independent. Furthermore, it suggests that the inclusion of multiple-scales is an essential component of any 2nd derivative model of Mach band perception.

However, the trend of increasing 'Mach band seen' responses with exponent *is* captured by the increasing amplitude of the 2nd derivative. Also, the use of the 2nd derivative for the sine data causes only a small kink in the prediction lines. So perhaps a modification of this approach would provide better fits. An alignment of the 2nd derivative amplitudes, for any image scale, would require that either the fine scale amplitudes are reduced, or the broad scale amplitudes are increased. The latter approach is employed in Georgeson's scale-space edge-bar model, discussed next.

2. Scale-space edge-bar model

The experimental data were modelled in two stages, which are described in detail below. Firstly, the images were passed through an existing version of the scale-space model that outputs bar predictions in addition to edge predictions (Georgeson, 2006). Secondly, the ratio of bar to edge responses was used to derive Mach bands predictions.

The first stage is an extension of the previously described N3+ model, without pre-filter (see chapter 3). But in addition to the two edge channels, there are two bar channels; one for dark bars and the other for light bars. In each bar channel, the image is filtered with a bank of Gaussian 2nd derivative filters at a range of scales. The filtered waveforms are then multiplied by a scale-dependant normalisation factor ($\sqrt{\sigma}$) which amplifies the output of larger scales. The normalised filtered waveforms are then

stacked to form a scale-space map. An image of a single bar will produce a peak on this map, whose co-ordinates provide the position and scale (width) predictions.

The scale-space edge-bar model produces 4 scale-space maps for each image, one map for each feature type and polarity; dark-to light edge (DL), light-to-dark edge (LD), dark bar (D), and light bar (L). In order to align the bar and edge maps along a common scale axis, it is necessary to divide the scale of each bar filter by $\sqrt{5}$ before it is applied to the image. The four maps are superimposed and a 'maximum' operator applied. The relative gain (k) of bar channels to edge channels is an important factor; Georgeson set k to 0.4 before the 'maximum' operator is applied. The non-linear map combination vetoes any peak whose amplitude is lower than the amplitude of any of the other 3 maps at that point in scale-space. This crucial operation accounts for the absence of spurious features that would otherwise be predicted. For example, A DL step edge input leads to the following 4 maps. The DL edge scale-space map contains a single peak at the edge location and smallest scale. The model rectifiers ensure that there are no peaks in the LD edge map. The Light bar map contains a single peak just on the right (light) side of the edge at a small scale, and the dark bar map contains a single peak just on the left (dark) side of the edge at a small scale. These bar peaks have lower amplitudes that the central edge peak (due in part to the value of k), so are vetoed. Therefore the model correctly predicts a central DL edge with no Mach bands. This N3+ edge-bar model correctly predicted the presence of Mach Bands on ramp edges and sine-wave edges and the absence of Mach Bands on Gaussian-blurred edges (Georgeson, 2006).

While, the model's max operator ensures that spurious feature are not predicted, its use also means that candidate features are either vetoed or not. There is no middle ground. So in its current form, this model produces predictions of the probability of seeing Mach bands that are 0 or 1 only. In order to produce intermediate probabilities, and account for the current experimental data, a noisy decision process was added to the model.

The rationale behind the noisy decision process is explained as follows. The noise is assumed to be due to the combined effect of sensory noise (that causes repeated presentations of a stimulus to form a distribution of internal responses) and criterion noise (whereby the position of the observer's criterion varies across trials). On each trial, the observer bases his response on the ratio of the outputs of the bar and edge channels. These values are obtained from the scale-space maps at the spatial and scale locations of the peaks in the bar maps. A DL experimental image produces a pattern of responses similar to the step edge described above, except that the peaks in the bar maps are not occluded by the peak in the DL edge map, and the latter peak occurs at a slightly larger scale than in the step edge case. Since the present experimental images are odd-symmetrical, the D bar map is in odd symmetry to the L bar map, and so they have the same peak amplitudes. Since the DL edge map is even-symmetrical, only one bar-edge ratio need be considered (because the D bar-edge ratio equals the L bar-edge ratio).

If the bar-edge ratio is above 1 the 'Mach band seen' response is assumed to be positive, and if it is less than 1, the response is assumed to be negative. But on successive trials, the channels' output to the same stimulus varies somewhat. Thus, over a large number of trials, the bar-edge ratio becomes a response distribution that is assumed to be approximately Gaussian and to have a specific mean and standard deviation. So the z-score of a positive response to the stimulus under consideration is the area under the distribution that is above the criterion. Since the criterion lies at 1, the probability of

a 'Mach band seen' response can be obtained from its z-score with

$$p = \Phi\left(\frac{r-1}{s}\right) \tag{4.2}$$

where r is the bar-edge channel response ratio, s is the standard deviation of the bar-edge ratio's response distribution, and Φ is the standard Gaussian integral. This noisy decision process was amended by the addition of two further factors, which transformed equation 4.2 into

$$p = \Phi\left(\frac{rg-1}{rgs}\right) \tag{4.3}$$

where g is the mean gain of the bar channel relative to the edge channel. Since g multiplies r, it effectively allows the relative gain of bar channels to edge channels (k, described above) to have values other than 0.4. This parameter was added to allow the model sufficient freedom to be able to fit the variations of data across observers. The second addition to equation 4.2 is the inclusion of rg in its denominator. This now transforms the noisy decision process from an 'additive noise' to a 'multiplicative noise' model. This factor was included because the resulting model provided a much improved fit to the data than the additive noise version, as will be described further below.

The model was fitted to the data by using the Solver macro in Excel to minimise the rms error between the data and the probabilities returned by the model, while allowing g and s to vary. Figure 4.9 (left) shows a fit of this model to all of the experimental data, which have been averaged across observers for each image scale. Each datum represents 360 trials. Error-bars ($\pm$1se) represent inter-observer variation. The best-fitting values of gain (g) and noise (s) were 0.95 and 0.55 respectively. The model clearly provides a good fit, thus supporting the use of this 2nd derivative scale-space approach for Mach band detection.

	scale	gain (g)	noise (s)
Fit to all data	**All**	0.95	0.55
Separate fits to each scale	**3**	1.13	0.55
	6	0.99	0.40
	12	0.75	0.69

Table 4.1: Fitted parameter values for the plots shown in figure 4.9.

While the Mach Bands do arise from peaks in the 2nd derivative, in this model the probability of observing Mach Bands depends on the extent to which the 2nd derivative (bar channel) response exceeds the 3rd derivative (edge-channel) response at the same scale-space position. This relationship is illustrated in figure 4.10, which shows the separate (mean) response of the edge and bar channels, and the bar-edge ratio, as a function of generalised Gaussian exponent. As the generalised Gaussian exponent increases from 1, the bar response increases monotonically, but the edge response is non-monotonic. It is initially higher than the bar response, but rises less rapidly. At an exponent of 2, it crosses the bar response and then falls towards zero at higher exponent values. Interestingly, the channel responses for the sine edge (open symbols) are both much higher than the responses to the neighbouring generalised Gaussians, but their ratio is nevertheless is in keeping with the generalised Gaussian ratios. This highlights the importance of the comparison between channels, rather than placing total reliance on 2nd derivative operators (bar channel). The bar-edge ratio shown here (r) was used to produce the fit in figure 4.9 (left).

The right panel of figure 4.9 shows separate model fits to the data of each scale. The model fit is good (rms error = 0.05). The results are (nearly) scale invariant across the 3 scales (3,6,12 mins) despite the fact that the 2nd derivative amplitude increases hugely at the smaller scales, as shown above. The intrinsic scale invariance of the scale-space model accounts very naturally for the observed scale invariance in Mach Bands over this range. The fitted values of gain (g) and noise (s) are shown in the bottom 3 rows of table 4.1. The top row shows the values from the previous fit to the data of all scales, which was shown in the left panel of figure 4.9.

This model is successful with the data averaged across observers. But its ability to capture the inter-observer variation has not been established. In order to address this issue, the model was fitted individually to data for each scale and observer, with g and s free. Best-fitting parameter values are listed in table 4.2 and experimental data with these model fits are shown in figure 4.11. The model fits generally follows the data for each observer and image scale rather well (rms error = 0.060) with only 2 free parameters (albeit set at different values for each observer and scale). The sine edge data have been inserted at exponent 3.42, as before. The sine edge predictions were also generated by the model. These values determine their vertical position on the plots. Their horizontal position was obtained by forcing them to line up with the generalised Gaussian predictions. This determines their horizontal position on the plots, but does not constrain a common exponent value across observers, or across scale for each observer. Nevertheless, each panel of figure 4.11 shows that their exponent value is relatively constant for all scales and observers (mean 3.16, sd 0.03). This means that the model correctly captures the perceptual similarity between the sine edge Mach bands and those of the generalised Gaussian edges of exponent 3 or 3.5, while still allowing for the inter-observer variation.

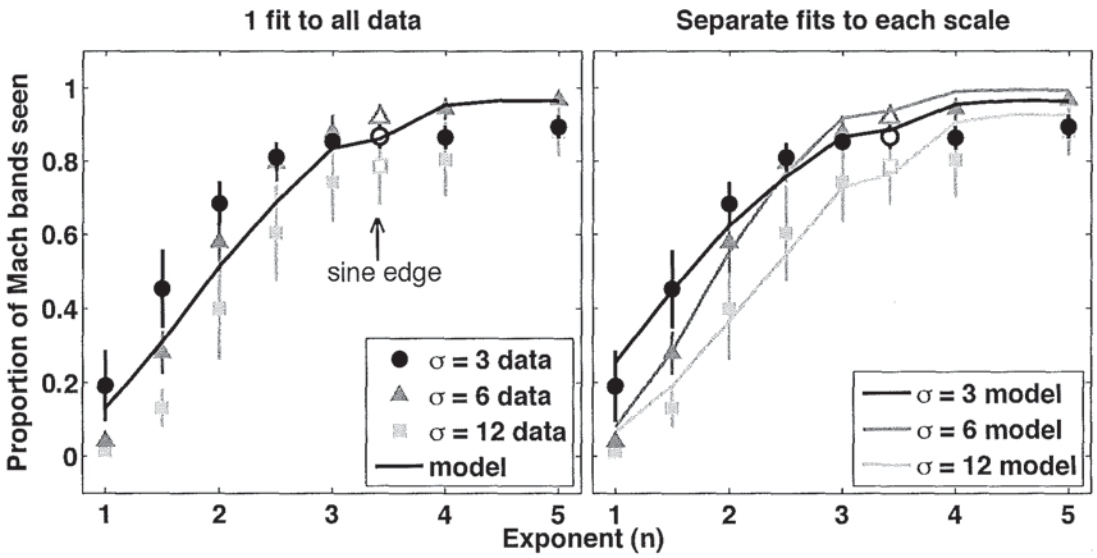

Figure 4.9: Proportion of 'Mach bands seen' responses, averaged across 6 observers (symbols) as a function of generalised Gaussian exponent, with fits of the scale-space edge-bar model (lines). Left panel shows 1 model fit to all data. Right panel shows identical data with separate model fits to the data of each scale. Each datum represents 360 trials. Error-bars (±1se) represent inter-observer variation. Fitted values of gain (g) and noise (s) are shown in table 4.1.

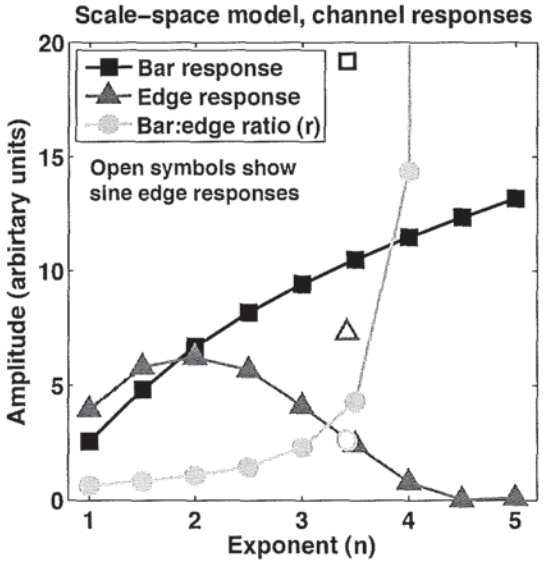

Figure 4.10: Mean responses of the scale-space model's bar and edge channels and the bar-edge ratio, as a function of generalised Gaussian exponent. The bar-edge ratio shown here (r) was used to produce the fit in figure 4.9 (left). At exponents above 4, the value of r is very high, but finite.

observer	SAW		MAG		TAY		RJS		NRH		DHB	
scale	gain	noise	gain	noise	gain	noise	gain	noise	gain	noise	gain	noise
3	0.90	0.88	0.92	0.63	1.16	0.33	1.86	0.66	1.53	0.49	0.93	0.46
6	0.96	0.47	0.95	0.36	0.96	0.30	1.22	0.25	0.70	0.75	1.06	0.25
12	0.73	0.43	1.13	0.18	0.34	1.61	0.81	0.39	0.09	8.03	1.07	0.17

Table 4.2: Fitted parameter values where noise and gain are free. Plots of model predictions that use these values are shown in figure 4.11.

In order to examine whether gain or noise captured more of the data trend, a further 2 fits were undertaken, but with gain or noise fixed. For each observer, the mean value of each parameter was calculated (the average of each column in table 4.2). One model fit employed the bar gain means as fixed parameters. This is shown in figure 4.12. The model fits include most of the data point for observers SAW, MAG and DHB, but do not fully capture the variation across scale for the other three observers. This is due to the common crossing-point that appears to be close to $n = 2$ for each observer.

The next model fit employed the noise means as fixed parameters. This is shown in figure 4.13. The model predictions for each scale converge to a common asymptote for each observer (although the value for the asymptote varies across observers). So some of the high exponent data fall below the model predictions for most observers. Nevertheless, this fit captures more of the inter-observer variation (rms error = 0.102) than the fit where gain was fixed and noise was free (rms error = 0.154), which suggests that variations in gain are more important than variations in noise level in accounting for the effects of scale in individual observers' data.

Figure 4.14 shows data averaged across observers and scale, where gain and noise were free, with 2 model fits. One used the scaled noise described above and the second used fixed (additive) noise. The equation used for the additive noise case was similar to equation 4.3 used previously, except that the

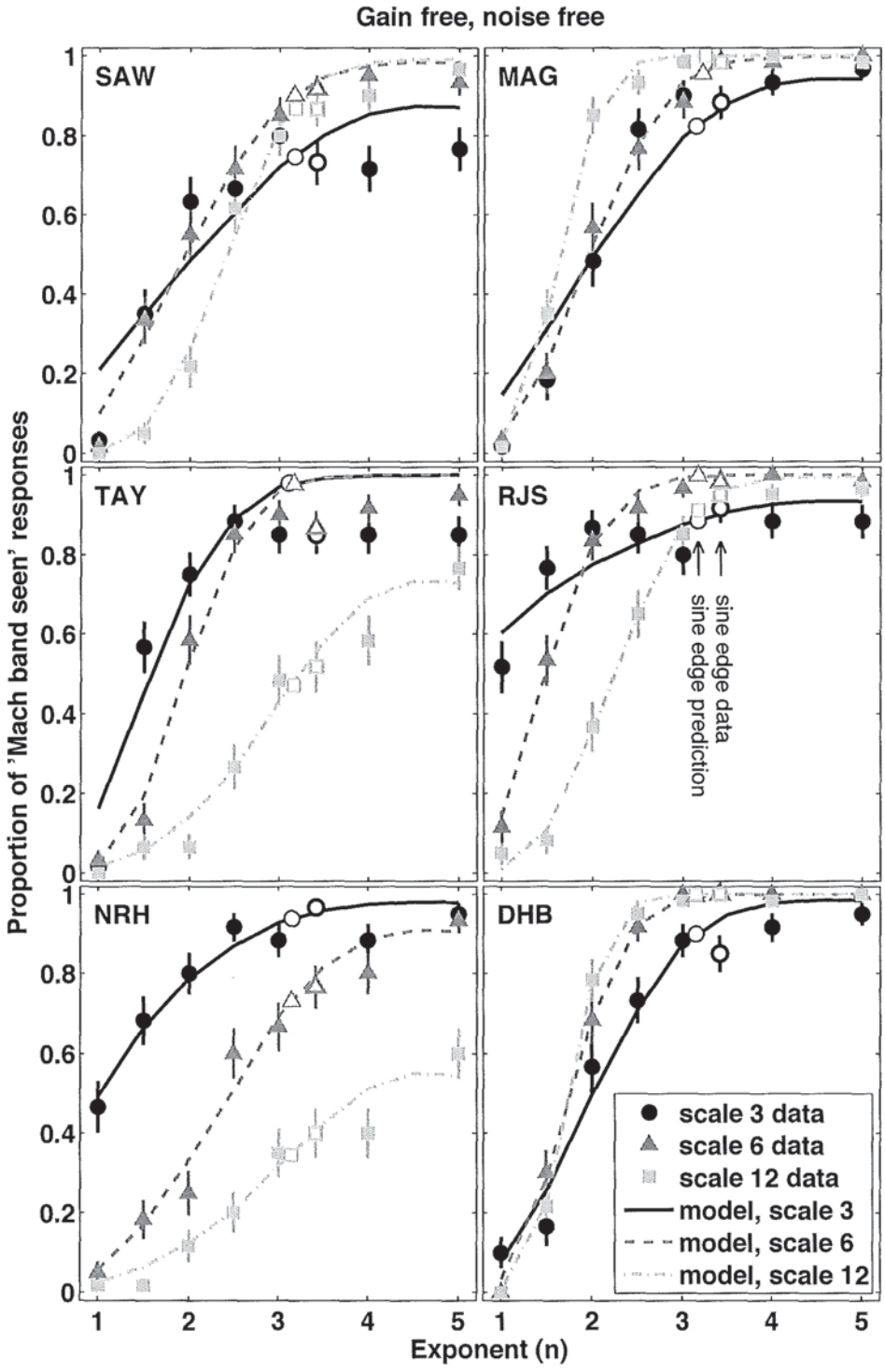

Figure 4.11: Proportion of Mach bands seen in generalised Gaussian images of various exponents, with model fits (lines), where gain and noise parameters are free. Open symbols show sine edge data (with error bars) and sine edge predictions (without error bars). Error bars (±1sd) are calculated from the expected variance in binomial sampling. The fitted parameter values are shown in table 4.2.

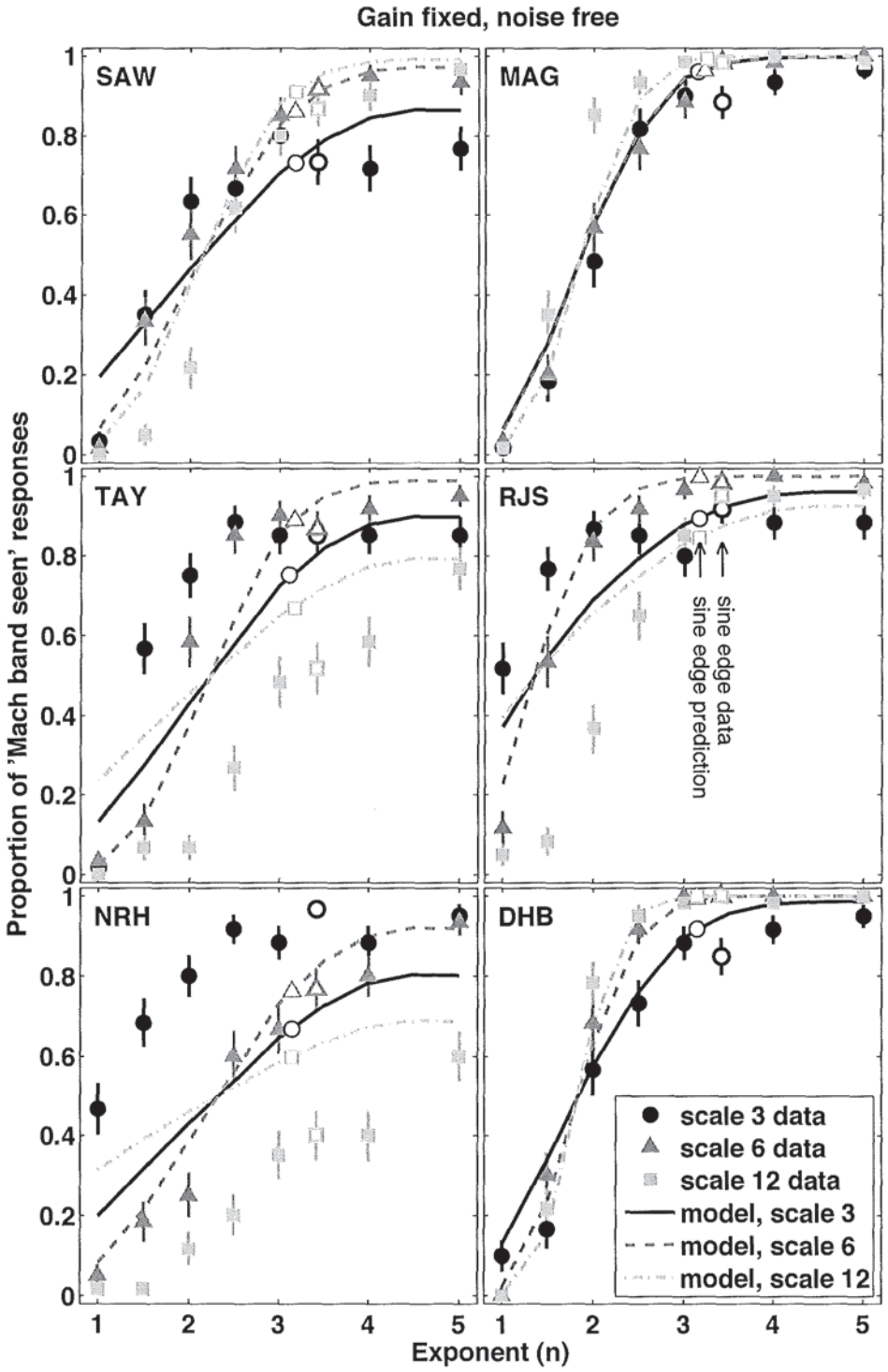

Figure 4.12: Proportion of Mach bands seen in generalised Gaussian images of various exponents, with model fits, where gain is fixed for each observer and noise is free. Open symbols show sine edge data (with error bars) and sine edge predictions (without error bars). Error bars (±1sd) are calculated from the expected variance in binomial sampling.

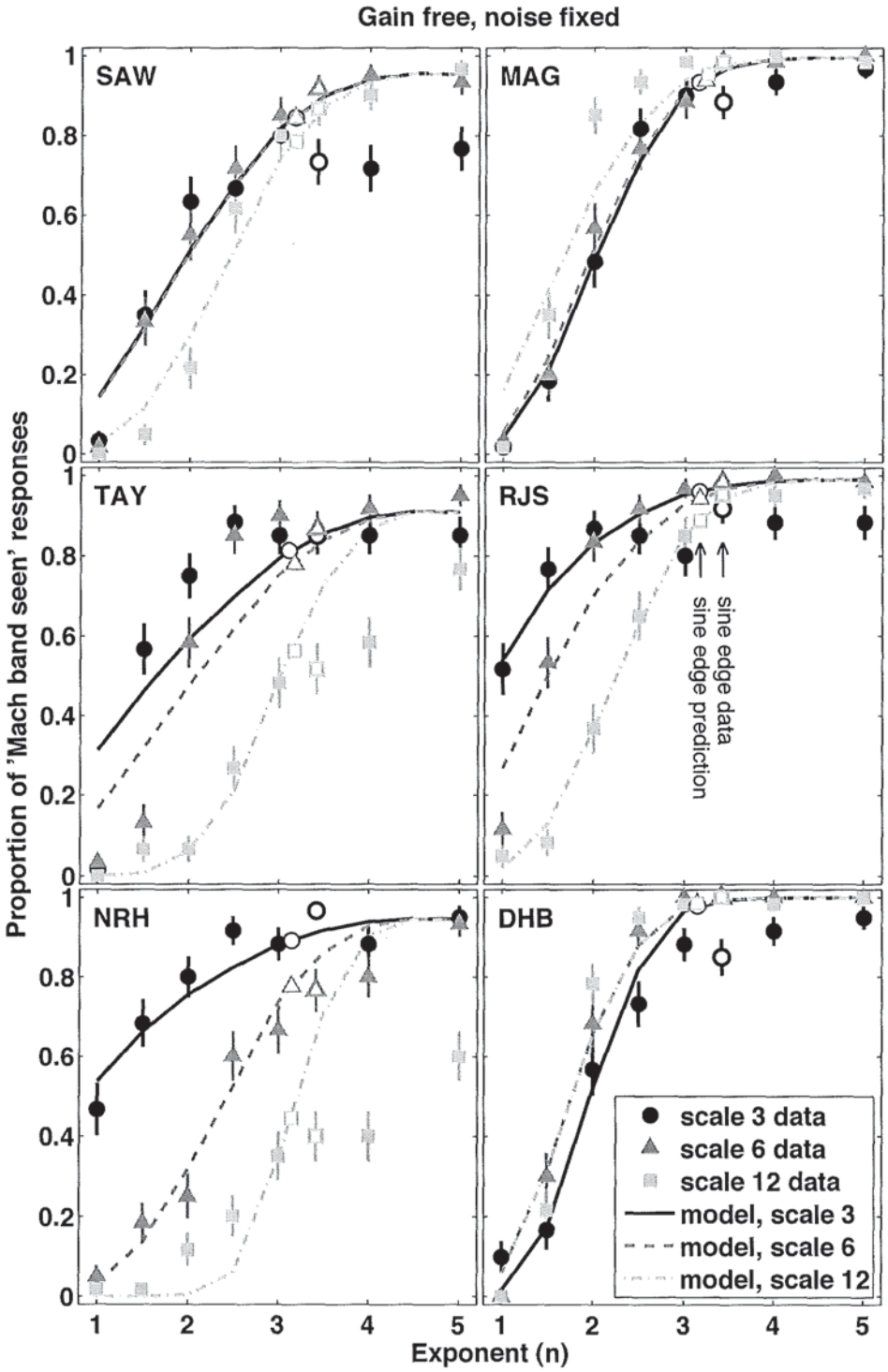

Figure 4.13: Proportion of Mach bands seen in generalised Gaussian images of various exponents, with model fits, where noise is fixed for each observer and gain is free. Open symbols show sine edge data (with error bars) and sine edge predictions (without error bars). Error bars (±1sd) are calculated from the expected variance in binomial sampling.

noise was no longer multiplied by the total bar-edge ratio (rg). Therefore, when additive noise applied, the proportion of Mach edge responses was given by

$$p = \Phi\left(\frac{rg-1}{s}\right) \tag{4.4}$$

where r is the bar-edge channel response ratio, s is the standard deviation of the bar-edge ratio's response distribution, and Φ is the standard Gaussian integral. Although slightly too high at the large exponents, figure 4.14 shows that the fit with scaled noise (rms = 0.073) is better than that with fixed noise (rms = 0.139). This shows why multiplicative noise was used in the model.

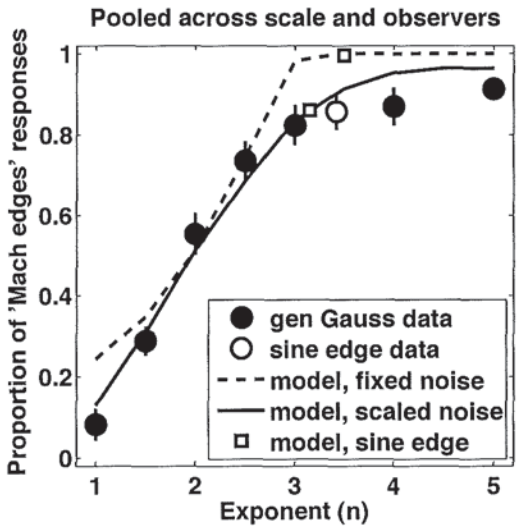

Figure 4.14: Proportion of Mach bands perceived in generalised Gaussian edges (filled circles) and a sine edge (open circle), with model predictions (lines) using scaled or fixed noise. Open squares show sine edge predictions. Data is collapsed across scale and observer. Error bars (±1se) show inter-observer variation for scale-averaged data (n=6).

To summarise, the Yes-No experimental data is well fitted by a two-stage model consisting of 1) a scale-space process that uses peaks in the normalised 2nd derivative to detect bar locations, and 2) a noisy decision process, with multiplicative noise, that uses the bar-edge amplitude ratio at these locations to determine the likelihood of Mach band perception.

3. Single-scale 3rd derivative model, with pre-filter

The scale-space N3+ model (Georgeson et al., 2007) was revised in section 3.8 by the addition of a pre-filter that was based on retinal ganglion cell physiology. This addition was required in order to account for experimental data on Mach edges with a novel image set formed of a blurred triangle-wave with added a luminance ramp. When the gradient of the added ramp was ±1, the image was one side of a blurred Mach ramp (similar to the images used in the present experiment). A single-scale version of the revised model provided a good fit to the data. So that model may also provide a good fit to the present results.

However, the single-scale N3+ model with pre-filter can only produce predictions for edges, not bars. Therefore a Mach band was inferred from the presence of 2 neighbouring edges of opposite polarity.

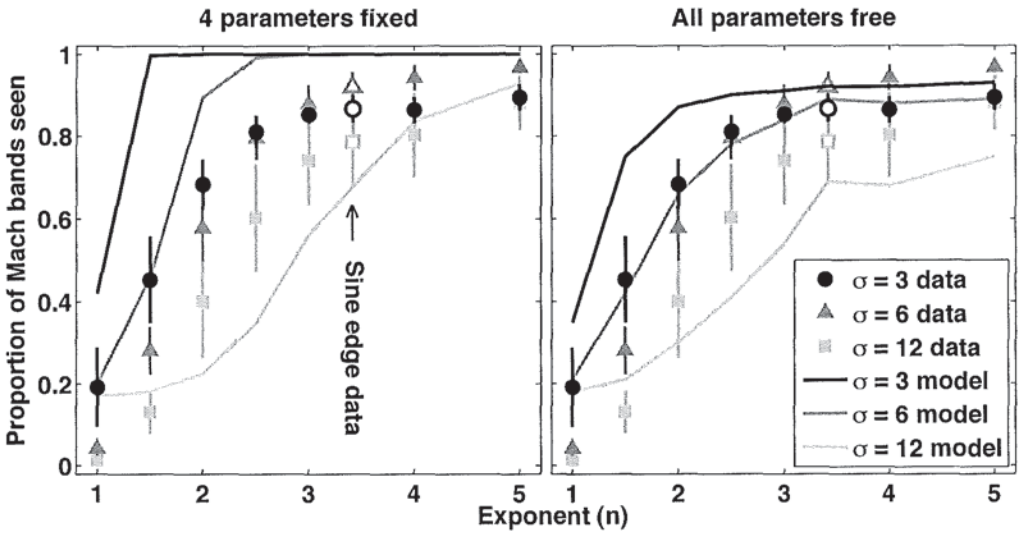

Figure 4.15: Proportion of 'Mach bands seen' responses for generalised Gaussian and sine edge images (symbols) and single-scale N3+ model predictions with pre-filter (lines). Left panel model fits had 4 of the 5 parameters fixed at the optimum values for experiment 2 data (chapter 3). Right panel shows identical data with model fits that allowed all parameters to be free. In each panel, the parameter values were constrained to be the identical for all 3 image scales (σ). Error-bars ($\pm$1se) represent inter-observer variation for scale-averaged data ($n = 6$).

When the model's input was an exponent 5, DL generalised Gaussian image (of any of the 3 scales), it produced edge predictions (from left to right) of LD, DL, DL and LD, where the first 2 edges were left of image centre and were assumed to delimit the dark Mach band. The other 2 edges were right of image centre and were assumed to delimit the light Mach band. This pattern of edge predictions was also obtained with the exponent 2.5 to 4 images of scale 6 and 12. For all other current experimental images, the pattern of predictions was different. The exponent 1 images of scale 6 and 12, and the exponent 1.5 image of scale 12 all produced predictions of one central DL edge. No LD edges were predicted for these waveforms because the pre-filter did not cause a small gradient reversal in their luminance profile. All other current DL experimental images produced predictions of a central DL edge flanked by 2 LD edges. Here, the central DL edge was assumed to delimit the right side of the dark Mach band and the left side of the light Mach band.

Figure 4.15 shows generalised Gaussian data (averaged across image polarity and observers) with fits from the single-scale N3+ model with pre-filter. The left plot shows model fits with 4 of the 5 parameters fixed at the same values that were used to fit the data of experiment 3 (Mach edge stimuli with added luminance ramps). These values are shown in the middle column of table 4.3. Only the 5th parameter was allowed to be free (noise), but it was initially constrained to be the same value for all 3 image scales (σ).

The fit with these parameters values is very poor (rms error = 0.20), so a second fit was performed with all parameters free. The fitted values are shown in the right column of table 4.3, and the model predictions are shown in the right panel of figure 4.15. Although the fit is improved (rms error = 0.13), the false alarm rate of 0.19 prevents model predictions from reaching the 3 data points that fall below this value. Also, predictions still do not fall within error-bars for much of the scale 3 and 12 data.

Parameter	Value	
	Fit 1: 4 parameters fixed	Fit 2: All parameters free
pre-filter centre spread, σ_c	(1)	1
pre-filter surround spread, σ_s	(10)	21.4
pre-filter surround/centre weight, k	(0.55)	0.35
false alarm rate, FA	(0.2)	0.19
noise	8.7×10^{-5}	8.1×10^{-5}

Table 4.3: Parameter values for the fits shown in figure 4.15, which are derived from a single-scale N3+ model with pre-filter. Brackets denote fixed parameters.

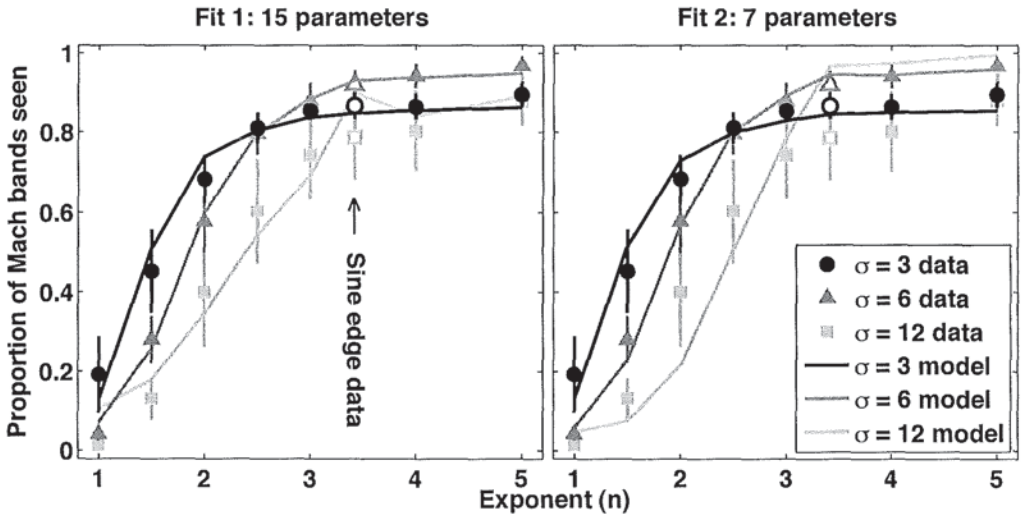

Figure 4.16: Proportion of 'Mach bands seen' responses for generalised Gaussian and sine edge images (symbols) and model predictions that vary across image scale (lines). Left panel model predictions allowed all parameters to vary across image scale. Right panel shows identical data with 4 of the 5 model parameter values common to all image scales.

Thus the model over-emphasises the variation across scale.

So, further fits were performed where all 5 parameters were free to vary across image scale. These are shown in the left panel of figure 4.16, and clearly provide a very good fit with the data. However, since 15 parameters are used to fit 24 data-points, this appears to be an example of 'over-fitting'. But the resulting parameter values can be used to optimise future fits that use fewer free parameters. The parameter values used to generate this plot are shown in table 4.4 (left 3 columns).

The fitted value for the size of the pre-filter's surround (σ_s) was 50.5 for the scale 12 data set, which is implausibly high. So a further fit was performed, but with σ_s fixed at a value of 16, which is within the range spanned by the scale 3 and 6 data sets. Three of the other four parameters were also fixed at levels that were either identical or similar to the values used to model the Mach edge data in experiment 3.

The predictions generated by these values are shown in the right panel of figure 4.16. The fit is still good (rms error = 0.08), with 7 free parameters. However, the predictions are too high at the highest 3 exponent values for the scale 12 data. Nevertheless, the generally good fit suggests that observers

Parameter	Fit 1: 15 parameters			Fit 2 : 7 parameters		
	scale 3	scale 6	scale 12	scale 3	scale 6	scale 12
pre-filter centre spread, σ_c	1	1.2	1	(1)	(1)	(1)
pre-filter surround spread, σ_s	15.2	17.5	50.5	(16)	(16)	(16)
pre-filter surround/centre weight, k	0.58	0.55	0.47	(0.55)	(0.55)	(0.55)
false alarm rate, FA	0.049	0.057	0.093	(0.05)	(0.05)	(0.05)
noise $\times 10^{-4}$	1.61	1.55	0.47	1.46	1.01	0.50

Table 4.4: Parameter values for the fits shown in figure 4.16, which are derived from a single-scale N3+ model with pre-filter, where noise is free to vary across scale. Brackets denote fixed parameters.

may be using Mach edges to detect Mach bands. Yet the better fit (rms error = 0.05, see figure 4.9) of the scale-space bar-edge model, discussed above, suggests that observers use 2nd derivative filters (bar channels) to detect Mach bands directly. That model was fitted to the data with 2 parameters (gain and noise) free to vary across the 3 image scales. Therefore it had 6 parameters, compared to 7 in the present single-scale model. This lower number of parameters, coupled with the lower rms error suggest that the scale-space model should be favoured over the single-scale model. However, the difference between them is relatively small, so the single-scale model cannot be completely ruled out at this stage.

The values for noise in table 4.4, for both fits, show a progressive decrease as image scale increases, which is a linear relationship for fit 2. Observers reported that narrow Mach bands appeared bright and wide Mach bands appeared dim. Assuming that narrow bands are associated with small scale images, the pattern of fitted noise suggests that high noise is related to narrow, rather than dim Mach bands. However this experiment did not collect any data regarding Mach band width, so such a conclusion cannot be drawn at present. Mach band width data will be collected in the next experiment.

4. Local energy

An implementation of the local energy model (Morrone & Burr, 1988) was applied to the data. It was the same version as used in the previous chapter (section 3.5), but without the extra two small scale pairs of filters. It assumes that features are located at peaks in local energy, whose profile is obtained by squaring and summing the output of a quadrature pair of filters. An energy peak is identified as a bar or an edge based on whether the output of the even- or odd-symmetric filter (respectively) has a greater amplitude at the energy peak. Features of different scale are detected by using a number of pairs of filters, where each pair is of a different scale. The model predicts the identity, polarity and location of each feature.

In the implementation used here, each experimental waveform was convolved with 4 pairs of even- and odd-symmetric filters, thus generating 4 local energy profiles. Each fIter-pair was maximally responsive to a different peak spatial frequency, which ranged from approximately 0.25 to 16cpd across the 4 pairs. Filter bandwidth (at 0.6 height) ranged from 2.4 to 1.1 octaves. For all DL experimental waveforms, the two largest pairs of model filters predicted a central DL edge. However, the smallest filter-pair predicted a spurious central LD edge for some waveforms. So this was suppressed by the application of a low threshold to the energy profiles. The smallest and/or 2nd smallest filter-pairs now predicted 2 Mach bands for the images of exponent 2.5 to 5, and no Mach bands at lower exponents, for all 3 image scales. For scale 3 and scale 12 images, the Mach band predictions were obtained from the

smallest filter-pair, and 2nd smallest filter-pair respectively. For scale 6 images, both of these filter-pairs produced Mach band predictions. These predictions are in agreement with the experimental data, except that the Mach bands for exponent 2, scales 3 and 6 were seen on more than 50% of the trials.

The model incorrectly predicted no Mach bands for the sine edge image. This is surprising in view of the sine edge's equivalence to a generalised Gaussian of exponent of 3.4 and the model's prediction of Mach bands for the exponent 3 and 4 images. The model did, however, predict a central DL edge, flanked by 2 LD edges. In the single-scale edge model above, this edge configuration was interpreted as two adjacent Mach bands, but that approach was not adopted here because the local energy model does return bar predictions.

In the previous chapter, the good fit of the energy model to Mach edge data was attributed to the extent to which the shape of the odd-symmetric model filters resembled a 3rd derivative operator. The energy model also uses even-symmetric filters, which approximate 2nd derivative filters, as they have 3 lobes (figure 3.13). So the degree to which the energy model fits the present experimental data may also be attributed to their similarity in shape to 2nd derivative operators.

5. MIRAGE

The MIRAGE model was described in detail in chapter 1, and was summarised in section 4.1.2. It was implemented here in Matlab, using the version specified by Watt & Morgan (1985). There were four filters, which were normalised 2nd derivatives of a unit height Gaussian, given by

$$f(x) = \left(\frac{x^2}{\sigma^2} - 1\right) exp\left(\frac{-x^2}{2\sigma^2}\right) \tag{4.5}$$

where σ is the spread of the Gaussian and x is spatial position. The normalisation makes all the filters of a constant amplitude in the spatial domain. The value of σ determines the filters' width, and following Watt & Morgan (1985), was 0.35, 0.7, 1.4 or 2.8mins. The model convolves each filter with the input image and the resulting 4 outputs are each separated into positive and negative regions. All of the positive regions are summed, and all of negative regions are separately summed (thus forming two channels). The next model stage applies a low threshold. The value of this threshold is not specified by Watt & Morgan, so was obtained empirically by employing Mach ramp stimuli at 100% contrast. The threshold was set at a value that allowed MIRAGE to predict Mach bands when the Mach ramp width was 5mins, but no Mach bands at lower widths. The value of 5mins was selected as it is in good agreement with the experimental findings of 4mins (Ross et al., 1989), and about 6mins (Ross et al., 1981). The next stage of MIRAGE categorises each channel's output into regions of activity and regions of inactivity, and applies parsing rules to obtain feature predictions. The strength of each feature is supplied by the 'mass' of each region of activity, which is the area that it encompasses.

For the present scale 3 experimental images, MIRAGE predicted 2 Mach bands for all images. This agrees with the data of exponent 2.5 and above, but not below 2.5. Mach band strength was predicted to increase with exponent, and the sinusoidal edge was predicted to fall between exponent 3 and 4. Both of these predictions agree with the data. If an additional threshold was applied to the model's output, it would correctly predict the absence of Mach bands at exponents below 2.5.

When MIRAGE was applied to the images of scale 6 and 12, the prediction of Mach bands was reduced. For the scale 6 images, no Mach bands were predicted at exponent 1, and for the scale 12 images, no

Mach bands were predicted at exponents below 3. Where Mach bands were predicted, their strength was lower than those of the scale 3 images, by a factor of approximately 2 and 80, for scale 6 and 12 respectively. The model predictions for scale 3 were improved by the additional threshold on its output. If the same threshold was applied to the scale 6 and 12 output, it would block all Mach bands. It appears that MIRAGE is over-sensitive to the scale of the images used here. To conclude, in this implementation, MIRAGE predictions provided a poor fit to the current experimental data.

To summarise the model-fitting, the data were poorly fitted by the single-scale 2nd derivative model and MIRAGE. In both cases, the failure was due to predictions that were too high for small scale images and/or too low for large scale images. This failure points towards the need for a scale-dependant factor, which amplifies the output of broader scales. The scale-space edge-bar model (Georgeson, 2006) applies this factor, and its predictions provide a good fit to the data. An alternative approach is to use a model that detects the edges of the Mach bands, and generates Mach band perception from a 'filling-in' process, whereby apparent lightness spreads from edges to form regions of coherent apparent lightness (Gerrits & Vendrik, 1970). The single-scale 3rd derivative model with pre-filter was able to supply such edge predictions, and they also provided a good fit to the data. The pre-filter's shape approximates that of a 2nd derivative operator, but it is applied at a single fixed scale for all images. This model's success raises the intriguing possibility that Mach bands are detected by their edges. Most of the data were also reasonably well fitted by the local energy model, except for the sine edge data. This model also uses filters whose shape approximates 2nd derivative operators. Thus the 2nd derivative seems to be a common factor for all 3 successful models. But they differ in how it is applied and how the resulting output is processed. It would be helpful to obtain data that distinguish between these models, as this would help to determine the processing strategy employed in the human visual system. The next experiment attempts to gather such data.

4.6 Experiment 5.2 - Feature-marking applied to Mach band images

The previous experiment applied a Yes/No method to obtain Mach band detection data. The data suggested that 2 models should be rejected. However, 3 other competing models remain. This experiment aims to distinguish between them, by gathering feature-marking data.

4.6.1 Method

The test images were the same as in expt 5.1. The feature-marking method of experiment 1.1 was applied here, with the modification that the location and polarity of bars were indicated, in addition to that of edges. Also, the marker starting position was 32mins from image centre and marker movement was constrained to within the central 192mins. This restriction prevented responses to the relatively peripheral features that (due to the periodic nature of the luminance profiles) are central features in the opposite polarity images. Images were all of contrast 0.4. One session consisted of a randomised presentation of the image sets for all 3 image scales. Three observers each completed 10 sessions, but the first session from each observer was discarded as practice. All other methodological details were as experiment 1.1.

4.6.2 Results

Scatterplots were generated for each condition and observer, and clear button-pressing errors were corrected by eye, e.g. where a single DL edge was marked within a spatial region of several LD responses. This was necessary for only 40 of the 23,988 responses. An example scatterplot is shown in figure 4.17. In each trial, for polarity LD images, observers marked all or a number of the following 7 features, looking from left to right: DL edge, L bar, LD edge, LD edge, LD edge, D bar, DL edge. This represents the light Mach band (flanked by its 2 edges), a central edge, and a dark Mach band (flanked by its 2 edges). For polarity DL images, the polarity of all responses was reversed.

A Matlab program was written to automatically parse the data into 7 feature bins according to logical rules. This was needed as responses in a single trial could contain up to 3 neighbouring edges of the same polarity. The parsing rules were applied to the data from each condition in turn, and consisted of two passes over the data as follows (for polarity LD images):

1. Pass 1:
   (a) Obtain the data for one condition.
   (b) Obtain the data for one trial.
   (c) If there are 7 features, place each one in the 7 bins, in position order.
   (d) If there are fewer than 7 features, then

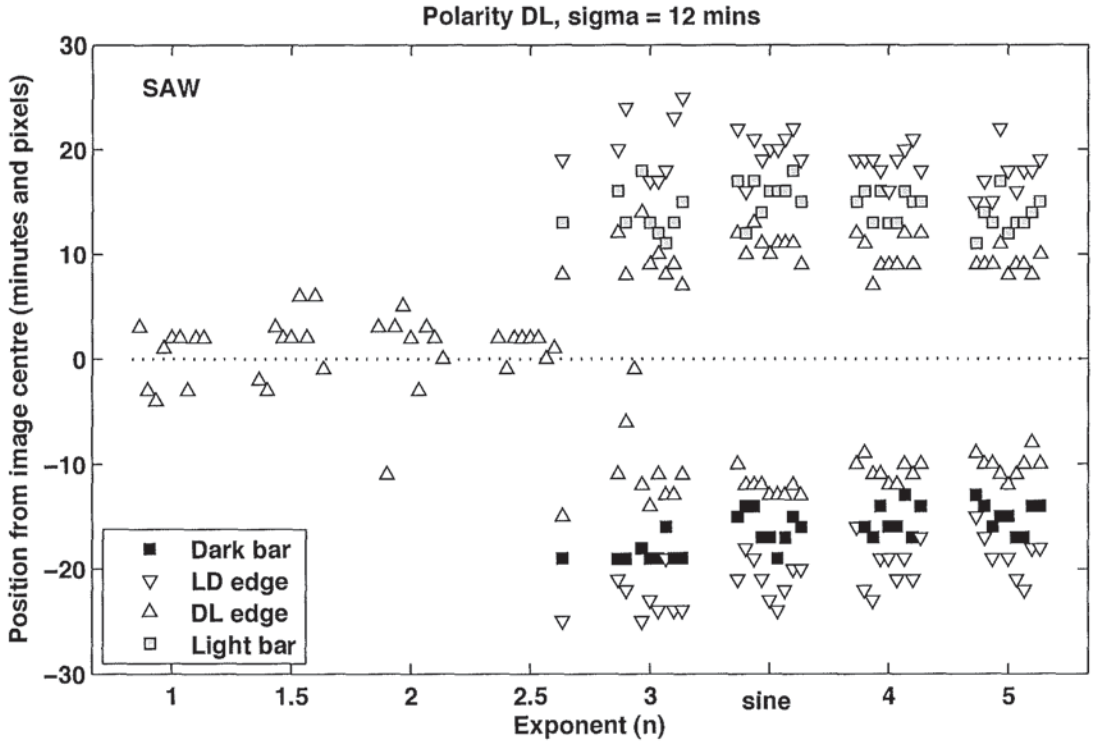

Figure 4.17: Example scatterplot of feature type (symbol shape and shade) and position for each waveform in a single image set. Polarity LD, scale 12mins, observer SAW. Each symbol represents a perceived feature, and each narrow column a single trial.

i. Place any bars into the bar bin of its polarity (bins 2 and 6).

ii. Place any DL edges that are to the left of the image centre in bin 1.

iii. Place any DL edges that are to the right of the image centre in bin 7.

iv. If there are 3 LD edges, place them in bins 3 to 5, in position order, but if there are less than 3 then delay assignment until second pass.

(e) Repeat steps 1b to 1(d)iv for every trial in the same condition.

2. Pass 2:

(a) Calculate the position of each bin from the mean of its contents.

(b) Obtain all of the LD data remaining from pass 1 for the whole condition, and place each datum in its nearest LD bin.

(c) Recalculate the LD bin positions from the mean of their contents.

The mean position of each feature (the mean of each bin's contents) in each condition was then obtained, and these data are shown in figure 4.18, for DL polarity stimuli, and in figure 4.19 for DL stimuli. Each panel of figure 4.18 is generally an inversion of its corresponding panel in figure 4.19, which demonstrates the minimal effect of stimulus polarity.

Each row shows data from one observer and column represents image scale. Each symbol is the mean of up to 9 trials, and its shape and shade denote the type of feature seen. Symbol size indicates the frequency of seeing (9 bins), so a relatively faint feature has a smaller symbol. Position from the image centre is positive where the feature was perceived to the right of image centre, and is negative for left of image centre. Errors bars (plotted behind symbols) show ±1se of each feature position. Based on the Yes-No results with this image set, the sine edge data have been inserted between exponent 3 and 4.

Figures 4.18 and 4.19 show two effects of increasing the image scale. Firstly, Mach band position is shifted away from the image centre. Secondly, a higher exponent value is required before Mach bands are seen. This agrees with the trend seen when the Yes-No method was applied with the same image set. An approximate prediction for the null point between seeing and not-seeing Mach bands for each scale and observer was extracted from the gain free, noise free model fits shown in figure 4.11 by obtaining exponent value that corresponds to a 50% response on each fitted curve. These values are listed in table 4.6.2 and are shown in figures 4.18 and 4.19 as vertical dotted lines. One set of values was derived from the Yes-No data that had been averaged across polarity, rather than a separate set for each polarity, because both Yes-No and feature-marking data show little or no effect of stimulus polarity.

observer / scale	SAW	DHB	NRH
3	2.1	2.0	1.0
6	2.0	1.8	2.5
9	2.4	1.8	4.1

Table 4.5: Thresholds (50% response) for Mach band perception obtained by inspection of the gain free, noise free fits to the Yes-No data, shown in figure 4.11.

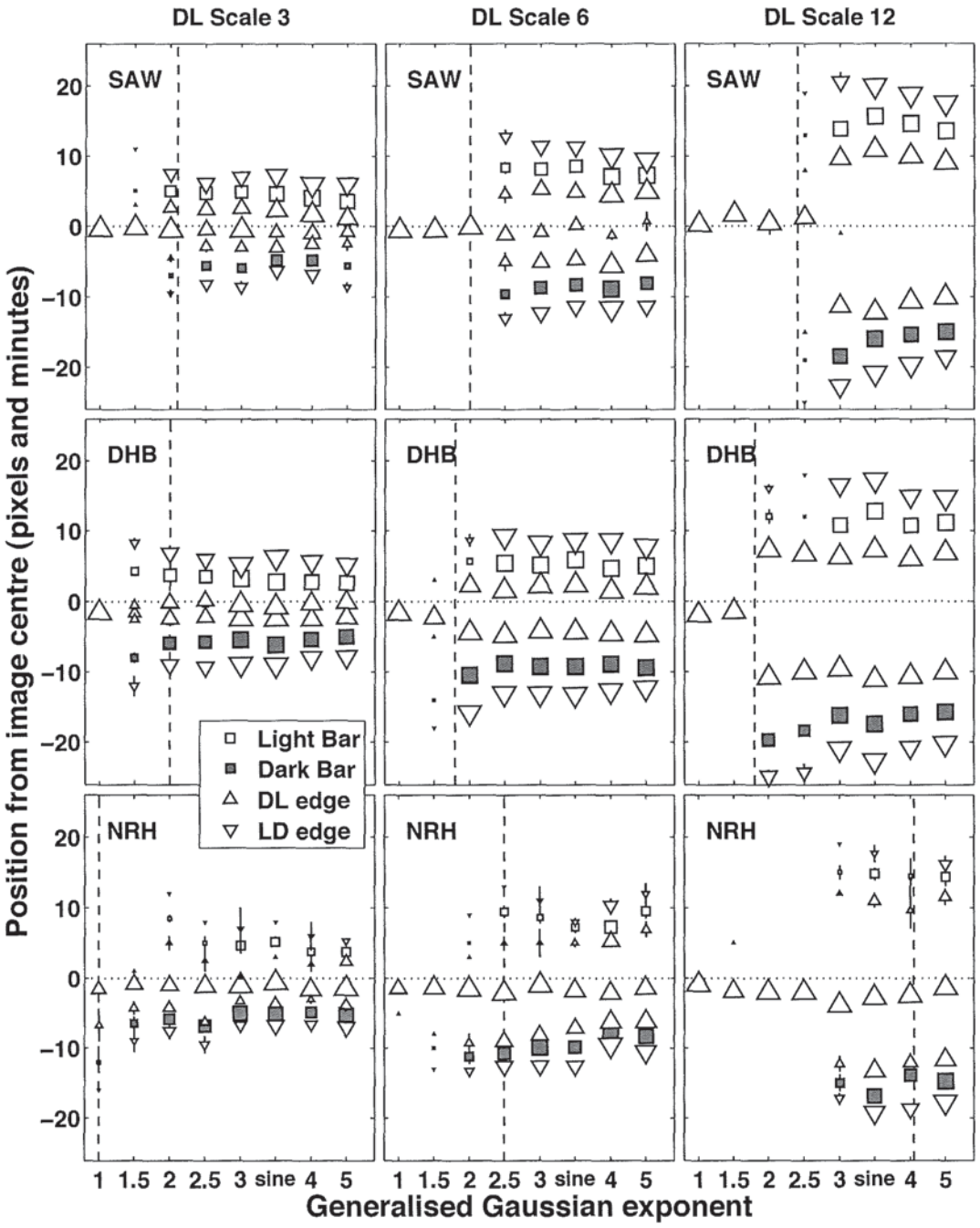

Figure 4.18: Features seen in DL sine edge and generalised Gaussian images for 3 observers (rows) and 3 image scales (columns). Each symbol is the mean of up to 9 trials, and its shape and shade denote the type of feature seen. Symbol size indicates the frequency of seeing (9 bins). Errors bars (plotted behind symbols) show ±1se of each feature position. Based on the Yes-No results with this image set, the sine edge data have been inserted at exponent 3.5. Vertical dashed lines show thresholds for Mach band perception derived from the Yes-No data of experiment 5.1.

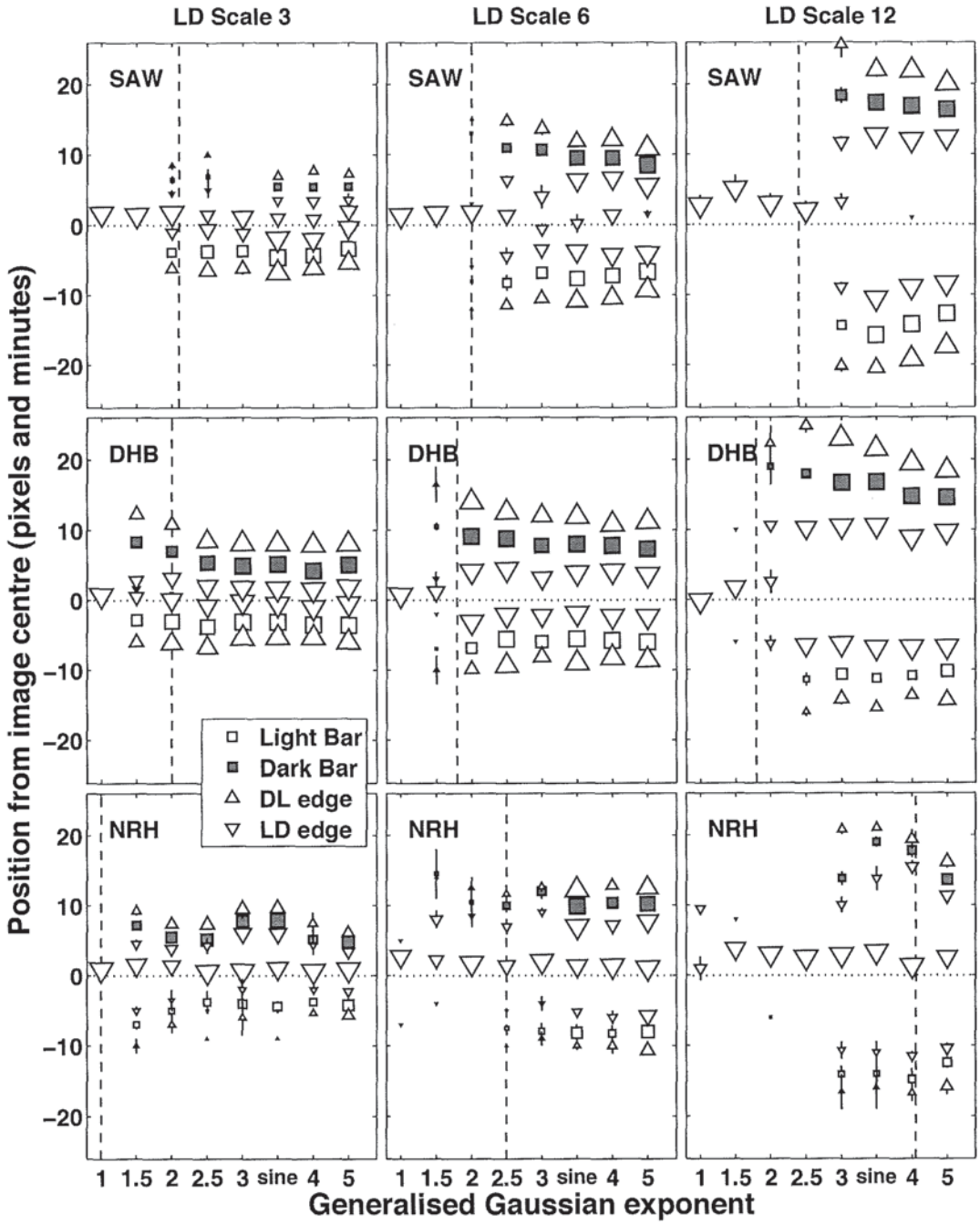

Figure 4.19: Features seen in LD sine edge and generalised Gaussian images for 3 observers (rows) and 3 image scales (columns). Each symbol is the mean of up to 9 trials, and its shape and shade denote the type of feature seen. Symbol size indicates the frequency of seeing (9 bins). Errors bars (plotted behind symbols) show ±1se of each feature position. Based on the Yes-No results with this image set, the sine edge data have been inserted at exponent 3.5. Vertical dashed lines show thresholds for Mach band perception derived from the Yes-No data of experiment 5.1.

In each panel, as the exponent value reduces, the transition between seeing and not seeing Mach bands is gradual. Symbol size generally decreases at this transition point, which represents a lower frequency of seeing, and suggests that the Mach band is weak.

The data are asymmetric about the image centre. The dark Mach band positions are generally further from the centre than the light Mach bands, and where a central edge was marked, its position is often shifted to the dark side, particularly for observers DHB and NRH. These appear to be examples of the Helmholtz irradiation effect (Mather & Morgan, 1986), which can be neutralised by averaging the absolute position data across Mach band polarity.

Mach band positions were averaged across Mach band polarity, observer, and image polarity. These means are shown as symbols in figure 4.20, with model predictions that will be described in the discussion. The scale 12 Mach bands lie at 14-16mins away from image centre, the scale 6 Mach bands lie at about half this distance (8mins), and the scale 3 Mach bands lie at 4-6mins. Mach band position therefore appears to form a near miss to scale proportionality.

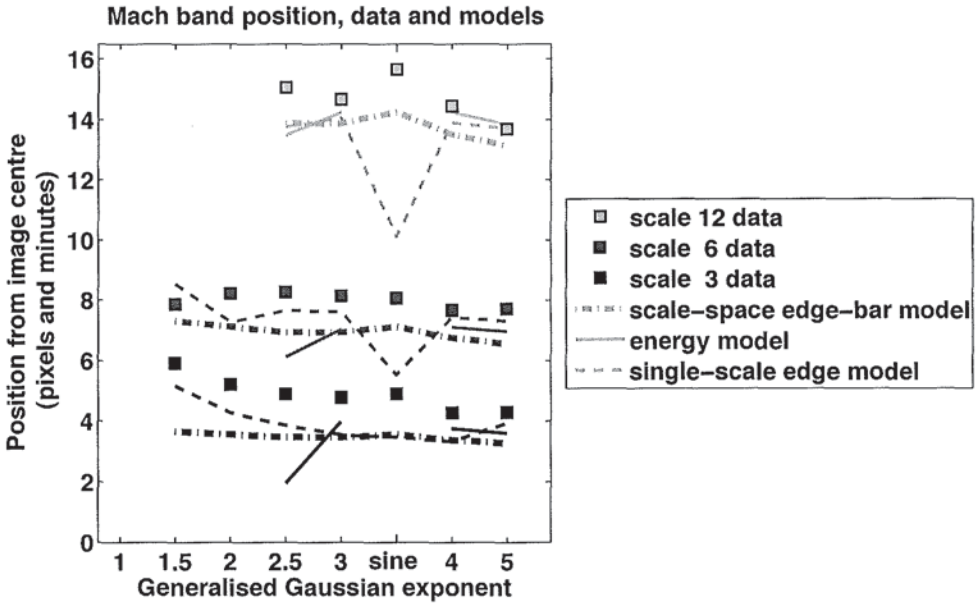

Figure 4.20: Mach band position data averaged across Mach band polarity and observer (symbols) for each of the 3 image scales' and model predictions (lines). RMS errors between models and data are shown in table 4.6, below.

	scale-space edge-bar	energy	single-scale edge
scale 3	1.47	1.57	1.01
scale 6	1.06	1.29	1.12
scale 12	1.02	0.83	2.57
mean	1.19	1.23	1.57

Table 4.6: RMS errors between the data and the three model predictions shown in figure 4.20.

Figures 4.18 and 4.19 carry information about Mach band width as the distance between the marked edges, that is the triangles that are above and below each square. In each panel, the width remains relatively constant across exponent. Also, the width of the light Mach bands for any one image scale appear to be approximately constant across observers, while the width of the dark Mach bands has greater inter-observer variability. This is more easily seen in figure 4.21, which shows Mach bands width as a function of exponent, for data averaged across observer and image polarity. The upper panel shows the dark Mach band data and the lower panel shows the light Mach band data. Error-bars ($\pm$1se) show inter-observer variation ($n \leq 3$), and are tilted to avoid overlap.

The small scale images produce narrower Mach bands than the large scale images. This is unsurprising, because the variation across image scale was achieved by a simple rescaling along the spatial dimension. Thus the width of the 2nd derivative profile would also be rescaled by the same factor.

The longer error-bars in the upper panel clearly show greater inter-observer variation for the dark Mach band data. Both panels show a clear trend of decreasing width as exponent increases. This appears to be more marked for the broader scale images.

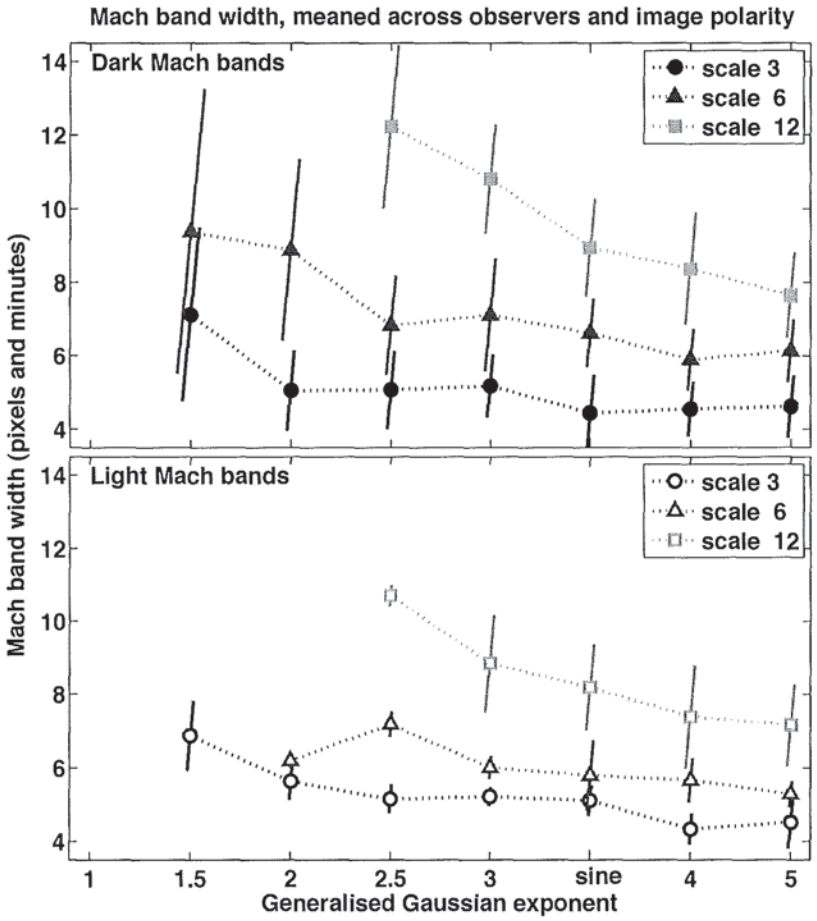

Figure 4.21: Mach band width data averaged across stimulus polarity and observer for each of the 3 image scales. Top and bottom plots show dark and light Mach band data respectively. Error-bars ($\pm$1se) show inter-observer variation (n=2 or 3), and are tilted to avoid overlap.

4.6.3 Discussion

Several aspects of the results are noteworthy. Firstly, the small errorbars in figures 4.18 and 4.19 indicate that the responses were repeatable, and suggest that the method was reliable. Secondly, the position of the vertical dotted line on each panel allows a comparison to be made between the Yes-No and feature-marking methods. The line was obtained from the Yes-No data of the previous experiment, and represents the transition (50 percent point) between seeing and not seeing the Mach bands. If the Yes-No and feature-marking methods are comparable, then the vertical lines should mark the transition between marking and not marking Mach band position (i.e. the presence and absence of square symbols). This can be seen to be true in almost all panels, which suggests that both methods are reliably investigating the same perceptual process.

Observer NRH marked dark Mach bands more frequently than light Mach bands, for all exponent values. This asymmetry is also seen in DHB's scale 12 data for LD stimuli, and is consistent with the previous research that has indicated that dark Mach bands are stronger than light Mach bands (Whiteside & Davidson, 1971; Gur & Syrkin, 1993). However SAW shows an asymmetry in the opposite direction at scale 3. Therefore, when all observers are considered, the present data only slightly support this previous research.

The main aim of this experiment was to attempt to distinguish between the 3 competing models of 1) the scale-space edge-bar model, 2) the local energy model, and 3) the single-scale edge model with DoG pre-filter. Predictions were generated from each model and plotted as lines in figure 4.20. The scale-space edge-bar model was the same version as used in the previous Yes-No experiment with this image set. Its predictions follow the horizontally flat trend of the data, although they are slightly too close to image centre, by an amount that is greater for the scale 3 images than the scale 12 images.

The energy model implementation applied a low threshold to the energy peaks, which suppressed faint features. This was necessary as its absence resulted in a large number of spurious features, which were not marked by observers. The resulting predictions exist only for exponent values of 2, 3, 4 and 5, and are shown as solid lines in figure 4.20. The absence of prediction for exponents of 1.5, 2 and the sine edge indicate the poor performance of this model when compared to the other two.

The single-scale edge model was the same version as used in the previous Yes-No experiment with this image set. The DoG pre-filter parameters were fixed at the best-fitting values used in the Yes-No experiment (see table 4.4). This model cannot directly produce bar predictions. So the existence of a Mach band was inferred from a pair of opposite polarity edges. For example, when the image was DL polarity, the model often predicted a central DL edge flanked by a pair of LD edges. This was assumed to indicate a dark Mach band, positioned left of image centre (at the midpoint between the left LD and DL edges), and a light Mach band, positioned right of image centre (at the midpoint between the DL and right LD edges). For some images, the predicted central DL edge was replaced by a pair of neighbouring DL edges, close to the image centre. In those cases, each Mach band had its own pair of unique edges. This model's predictions are shown as thin dashed lines in figure 4.20. They provide a fit to the data that is generally as good or better than the scale-space edge-bar model, except for the sine edge data of scale 6 and 12. Here the predictions are too far towards the image centre.

The rms error between each model and the data of each scale is shown in table 4.6, with the mean error for all 3 scales.

The single-scale edge model has a higher rms error (1.57) than the other two models (1.23 and 1.19), which quantifies its poorer fit. Although the local energy model's RMS error is similar to that of the scale-space edge-bar model, its absence of predictions for 9 of the 19 data points shown in figure 4.20 illustrates its poorer predictive ability. The single-scale edge model provided a poorer fit than the scale-space edge-bar model for the sine data of scale 6 and 12, and the exponent 2 datum of scale 12. However the rest of its predictions are as close or closer to the data. So it is difficult to reject this model outright on the basis of 3 out of the 19 data points. In summary, the present experiment's position data support the scale-space edge-bar model and suggest rejection of the energy model. They hint at a tentative rejection of the single-scale edge model.

The Mach band position predictions of the edge-bar scale-space model and single-scale edge model were too similar to each other and the data to provide a clear-cut rejection of either model. So perhaps consideration of Mach band width can distinguish between them. These data are shown in figure 4.21. To a first approximation, the width of the dark bands is equal to light bands for each scale and exponent value. So the data were pooled across both polarities for each observer and then averaged across observer. Figure 4.22 shows pooled data with model predictions from the scale-space edge-bar model and the single-scale edge model with DoG pre-filter. The energy model was unable to provide quantitative Mach band width predictions as only the position of each bar is provided. Inspection of the outputs of the energy filters of each scale would indicate an approximate scale. For example, if only the smallest scale pair of filters produced bar prediction and the larger filters did not, then the predicted bar would be small. However, as filters of only 4 sizes are used in the model, such an approximation would be excessively gross.

Figure 4.22 shows that the single-scale edge model with DoG pre-filter provides Mach band predictions

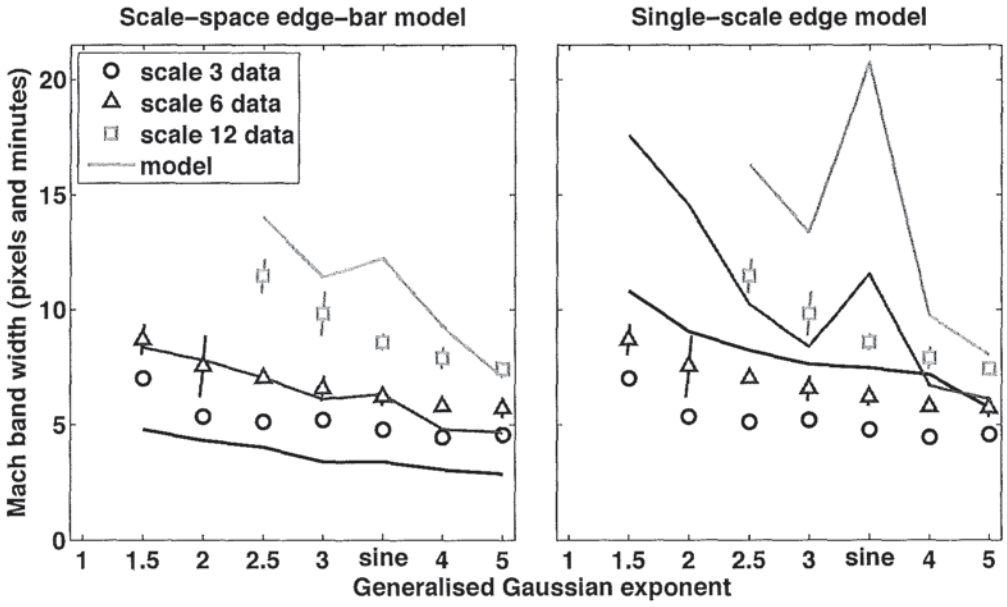

Figure 4.22: Mach band width data pooled across band polarity and observer for each of the 3 image scales (symbols) and model predictions (lines). Left panel shows the scale-space edge-bar model predictions and the right panel shows the single-scale edge model predictions. Both panels show identical data. Error-bars show inter-observer variation (n=2 or 3).

that are always too wide when compared to the data. This is particularly noticeable for the sine edge, as the scale 6 and 12 predictions suddenly increase here, breaking the trend formed by the other generalised Gaussian exponents. The comparatively poor fit of the sine edge predictions may be due to the assumptions used to generate the position and width predictions. Where the model predicted, for a DL stimulus, the following edges (from left to right) LD, DL, DL, LD, then the first pair delimited a dark Mach band and the 2nd pair delimited a light Mach band. So the position and width of each Mach band is explicit. However, where a central DL edge was flanked by 2 LD edges, two abutting Mach bands were assumed, which shared the central DL edge. This assumption was applied in the sine edge case. The resulting atypical predictions suggest that the assumption may not be correct. Nevertheless, the exponent 2.5, 3 and 4 predictions for scales 6 and 12 (which were derived without that assumption) still fall above the data, suggesting that this model is unlikely to characterise Mach band perception.

In contrast to the single-scale edge model, the scale-space edge-bar model predictions, shown in figure 4.22 (left), provide a much closer fit. The scale 6 data is fitted particularly well, but the predictions are a little too high for the scale 12 data and too low for the scale 3 data. To conclude, the Mach band width data suggest that the single-scale edge model can be rejected in favour of the scale-space edge-bar model.

Although the scale-space edge-bar model provides the best fit to all of the Mach band data, it nevertheless departs slightly from the width data for images of scale 3 and 12. So the addition of blur to the model may improve its fit at scale 3. Figure 4.23 shows the effect of adding a small amount of Gaussian blur to the image waveforms before they were passed through the model. The amount of

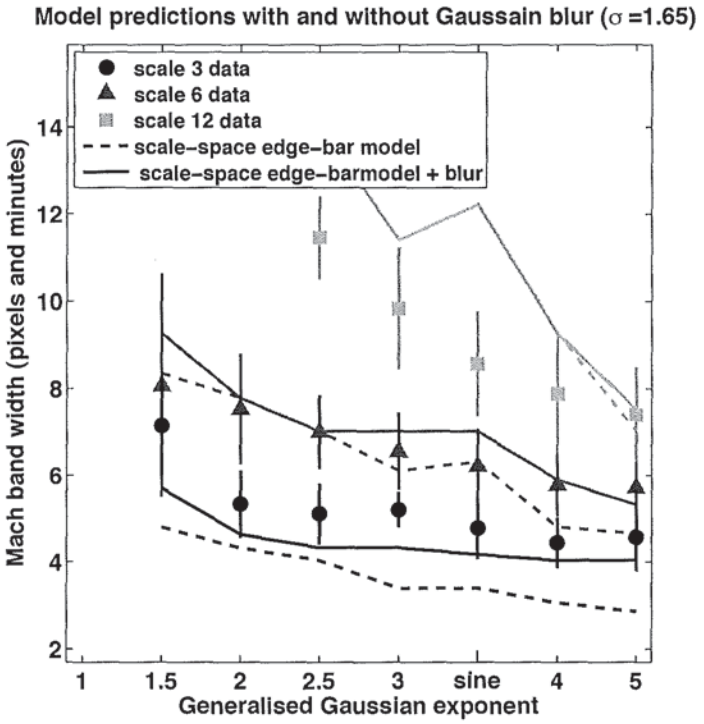

Figure 4.23: Mach band width data, averaged across 3 observers and Mach band polarity (symbols), and model predictions with and without Gaussian blur of $\sigma = 1.65$ (lines).

blur was optimised by using a Matlab program to find the least rms error between data and model predictions for a range of blur σ values, which were initially 0 to 3 in 0.5 increments. This identified a minimum rms error between 1 and 2.5, so a second run was performed with σ values of 1 to 2.5 at 0.05 increments. This run produced the final value of $\sigma = 1.65$. The addition of blur to the model improves its fit to the data (rms error with blur = 1.68, no blur = 1.91). Although the fit is good for the scale 6 data, it remains slightly to low for three scale 3 data points and too high for four of the scale 12 data points. This suggests that there may be one or more additional factors that this modelling requires. Their identities are presently unclear.

4.7 Chapter summary

Previous researchers have generally considered the existence of Mach bands on Mach ramps as evidence for a lateral inhibition process, which can be approximated by a 2nd derivative filter. In this chapter a novel set of images was investigated, in which the sharp ramp-plateau junction of the Mach ramp was replaced with a more curved transition.

In experiment 5.1, a single-interval Yes-No method was applied with a Mach band detection task. The resulting data were matched against the predictions of 5 competing models that all used 2nd derivative filters, or approximations thereof. The data did not fit the predictions of 2 models, so the models were rejected. But 3 models remained. In an effort to distinguish between them, experiment 5.2 applied a feature-marking method to the same set of images. The Mach band position data suggested a rejection of the energy model. The Mach band width data were poorly fitted by the single-scale edge model with pre-filter, despite its reasonable success with the Mach edge image sets used in chapter 3. The one remaining successful model was the scale-space edge-bar model, which therefore appears to characterise Mach band perception.

Chapter 5

Determinants of bar detection

5.1 Experiment 6 - Detection of a 3f cosine on a f cosine background

5.1.1 Introduction

Since the seminal work of Campbell & Robson (1968), a popular notion of early visual processing is that the visual signal is passed though parallel, independent, bandpass channels that are each sensitive to a limited range of spatial frequencies (Blakemore & Campbell, 1969; Blakemore et al., 1970; Graham & Nachmias, 1971; Sachs et al., 1971; Stromeyer & Julesz, 1972; Sullivan et al., 1972). Campbell & Robson (1968) found that a square-wave grating could not be distinguished from a sinusoidal grating until the amplitude of its 3rd harmonic had reached its own detection threshold. This suggests that the bandwidth of the spatial frequency channels is sufficiently narrow that independent processing occurs for 2 sinusoids that have spatial frequencies of f and 3f. Graham & Nachmias (1971) examined contrast detection thresholds for gratings composed of f and 3f components at various phase and contrast relationships. Their data supported independent processing of the f and 3f components, due to a square-shaped summation square (which is plot of fundamental contrast against 3rd harmonic contrast, when the compound grating is at threshold).

However, evidence for interaction between spatial frequency channels has been provided by some researchers. Tolhurst (1972a) used a square-wave stimulus and measured the amount of adaptation at its fundamental and 3rd harmonic. The amount of adaptation was much less than that predicted from the effects of f and 3f when viewed in isolation. The conclusion reached was that the f and 3f components mutually inhibit one another (when above threshold). Similar results were obtained when the square-wave grating was replaced with a f + 3f grating, which suggested that higher square-wave harmonics (5th, 7th, etc.) had little or no impact.

Nachmias & Weber (1975) used a 2AFC procedure to explore the detection of 3f in the presence of f, and found that a suprathreshold f component *facilitated* the detection of 3f. This finding was replicated by Tolhurst & Barfield (1978) and Lawden (1983). However, in the latter study, when the suprathreshold component was 3f rather than f, it was found to mask (not facilitate) f detection. This suggested asymmetric frequency-dependent interactions.

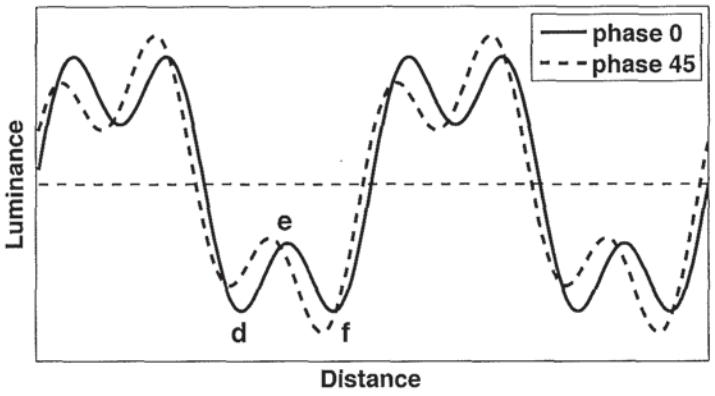

Figure 5.1: f + 3f 'peaks subtract' luminance waveforms of the type used by Badcock (1984b). The legend refers to the phase relation between the f and 3f sinusoids. Extrema d, e and f are used to obtain measures of 'local contrast' (see text for details).

Meese (1993) employed compound grating stimuli that were composed of 2 sinusoidal components in peaks subtract phase, at a range of frequency ratios that ranged from 1:3 to 1:12. The contrast of the lower frequency component (c_f) was fixed, and the contrast of the higher frequency component (c_{nf}) was altered to find the threshold for detection of the central dark bar that was caused by the higher frequency component. The resulting threshold value of c_{nf} was close to double c_f at the lowest frequency ratio (1:3), but the threshold was higher at higher frequency ratios. When the c_f/c_{nf} contrast ratio was plotted as a function of the f/nf spatial frequency ratio on log-log axes, the data formed a straight line, whose slope suggested a filter exponent of about 2.4. This exponent was higher than the value obtained in previous experiments with plaids (Meese, 1993). Although no clear reason was apparent for this discrepancy, Meese does suggest that different mechanisms, with different gain functions, may be involved in the two cases: bar detectors here, and edge detectors in the earlier work. The idea of separate mechanisms for edges and bars is central to the scale-space model of feature detection developed by Georgeson (2006).

A 'local contrast' alternative to the spatial frequency channel approach was proposed by Badcock (1984b). He used a fundamental grating of fixed suprathreshold contrast and added a 3f component in 'peaks add' (triangle-wave) or 'peaks subtract' (square-wave) phase at a range of contrast levels. A 2AFC procedure was applied to determine the sensitivity to phase shifts of 3f relative to f. A similar procedure was then used with a single contrast of 3f and various contrast levels of f. The pattern of the data was consistent with a model that used a measure of 'local contrast' that is computed over pairs of (usually) adjacent dark and light bars. For the 'peaks subtract' waveforms, shown in figure 5.1, the local contrast between extrema d and e is given by

$$C_d = \frac{d-e}{d+e} \tag{5.1}$$

Likewise, the local contrast between f and e is given by

$$C_f = \frac{f-e}{f+e} \tag{5.2}$$

Badcock (1984b) derived a 'relative contrast difference' percentage (RCD) by

$$RCD = \frac{C_f - C_d}{C_d} \times 100 \tag{5.3}$$

The RCD is zero when the phase difference between f and 3f is zero, because C_d has the same value as C_f. But when the phase difference increases, the waveform becomes asymmetrical, such that C_d has a different value from C_f (see figure 5.1). This means that the RCD now has a non-zero value that depends on the phase difference. Badcock (1984b) found that the RCD predicted the observed superiority of 'peaks subtract' performance over 'peaks add' performance, and the differences in slopes of the corresponding psychometric functions. A subsequent study added an additional parameter to the RCD model, which represented a compressive contrast transducer (Badcock, 1984a). The revised model produced quantitative predictions that fitted the experimental data for f + 3f gratings and f + 4f gratings. Thus apparent sensitivity to phase could actually represent sensitivity to local contrast.

Thus far, only two accounts of f + 3f grating perception have been discussed; spatial frequency analysis, and local contrast. A 3rd account uses the edge blur between adjacent bars. Hess & Pointer (1987) interpreted edge blur to mean the inverse of luminance gradient. They used a 3AFC procedure to perform the tasks of:

1. detection of a f or 3f sinusoidal grating
2. detection of 3f in squarewave phase to a fixed high contrast f.
3. phase discrimination of the 3f component relative to a fixed high contrast f (i.e trianglewave phase *vs* squarewave phase

They applied two contrast models to the resulting compound detection and phase discrimination data. One model involved the computation of the peak-to-peak difference in contrast between the pair of compound waveforms that were at discrimination threshold. The other model computed a measure of local contrast, following Badcock (1984a,b). However, both models provided a poor fit to the contrast discrimination sensitivity of the visual system, which was measured at each of ten pedestal contrasts. This suggested that the compound detection and phase discrimination performance was not based on a contrast metric.

However, a model that employed a measure of edge blur performed better. The amount of blur was defined as the reciprocal of the maximum luminance gradient. The difference in blur was calculated between the pair of compound waveforms that were just discriminable from each other. This blur difference was the same magnitude as the blur discrimination thresholds that were obtained for sinusoidal edges of various slopes and at various pedestal blurs. This equvalence supported the notion that edge blur is the cue that observers were using in the f and 3f tasks. This conclusion was also supported by the finding that performance worsened when the edge blur was randomised within each trial, even though the local contrast was largely unaltered.

While the luminance gradient appeared to be the salient cue, Hess & Pointer (1987) conceded that it may have been some other cue that covaried with it, such as the apparent width of the edge. This ambiguity highlights a problem with this type of 'objective' experimental method: it does not tell us directly which cue the observer uses (or what the observer actually sees). The observer will tend to use whatever combination of cues affords the greatest number of correct responses.

If we are interested in the coding and perception of local features then a more 'subjective' method, in contrast, may provide more information about what the observer sees (but at the expense of being more subject to criterion effects). The present experiment will use a variable contrast 3f on a f background, but with a more subjective single interval procedure without feedback, accompanied by instructions that specify which stimulus feature to attend to. A number of models will be used to fit the resulting data. The model that provides the best fit to the data will be assumed to best characterise human visual processing in this task. Of particular interest is the pre-filter + 3rd derivative model, because it has provided a good fit to almost all of the experimental data in this thesis, thus far. It has previously been applied to the 'illusory' Mach edges and Mach bands. It will now be tested with 'real' bars; that is, bars that are peaks of luminance.

A second aim of this experiment is to examine the observers' criterion. In a typical psychophysical experiment, the effect of the criterion is generally factored out, because it is thought to represent bias, which might otherwise contaminate the data (see section 1.11 for a more detailed account of signal detection theory). However, the criterion can be considered not as a nuisance, but as a source of useful information regarding the location of the observer's perceptual category boundary. In the present experiment it will be used to examine the stimulus that lies at the border between seeing and not-seeing thin 3f bars on a background consisting of wide f bars.

5.1.2 Null point prediction

Consider the f + 3f waveforms, shown in figure 5.2. As the contrast of 3f is increased from zero (dashed line), the peak of the fundamental gradually becomes less pointed, until eventually it becomes a plateau. Further increases produce a central dip. The observer's task is to report whether a central

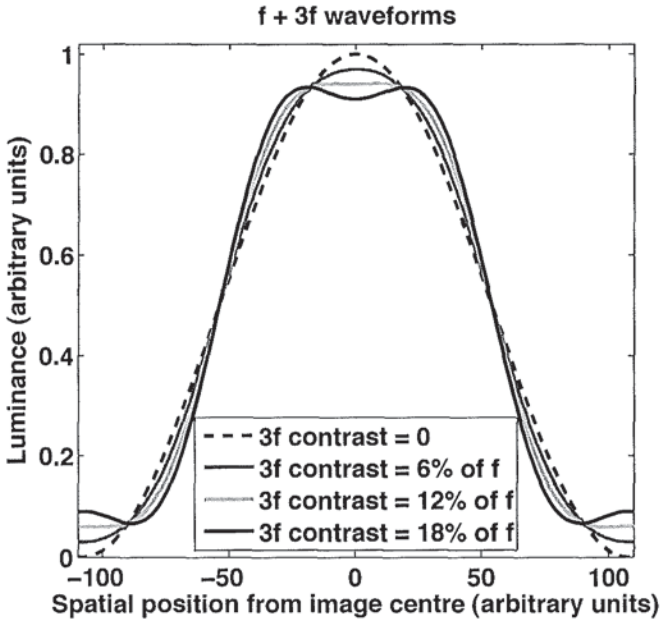

Figure 5.2: One period of f + 3f waveforms, where fundamental (f) contrast is fixed and the 3rd harmonic (3f) contrast varies. Polarity shown is 'f peak central'.

dark bar is seen. The 'null point' (50%) between reporting and not reporting the central thin bar identifies the location of the observer's criterion, and this could correspond to the perceptual category boundary for identification of this dark bar feature. It can then be hypothesised that this will be where the 3f component produces a plateau in the waveform. This hypothesis is based on the idea that if the background was of uniform luminance, and the task was polarity identification of a bar, an unbiased observer would provide a 'light' response when the bar was a luminance peak and a 'dark' response when the bar was a luminance trough, and the boundary between these 2 responses would be where the bar had zero contrast.

The hypothesised 'null point' will occur when the 2nd derivative at the image centre is zero. This occurs when the 3f contrast is 11% of the f contrast, as shown in the following proof:

The f + 3f waveform is defined as

$$y = c_1.cos(2\pi fx) + c_3.cos(6\pi fx) \tag{5.4}$$

where c_1 and c_3 are the contrasts of f and 3f respectively.

Differentiating twice produces

$$y'' = -c_1(2\pi f)^2.cos(2\pi fx) - c_3(6\pi f)^2.cos(6\pi fx) \tag{5.5}$$

At the waveform centre, x=0, so at that point equation 5.5 reduces to

$$y'' = -c_1(2\pi f)^2 - c_3(6\pi f)^2$$

At $y'' = 0$ this can be rearranged

$$\frac{c_3}{c_1} = \frac{-(2\pi f)^2}{(6\pi f)^2} = -\frac{1}{9} \tag{5.6}$$

This ends the proof.
The minus sign before the 1/9 in equation 5.6 indicates that the polarity of the 3f waveform is opposite of the f waveform at its centre; in other words, that the phase relationship between the f and 3f waveforms is square-wave.

If the detection threshold for the thin bar is defined as the 3f contrast that provides a certain above-chance level of performance (such a 84% 'thin bar seen' responses), then it can be posited that this threshold will be somewhat higher than 11% of the f contrast. Previous research has suggested that a f pedestal facilitates detection of a 3f component, when a forced choice (Georgeson & Georgeson, 1987; Hess & Pointer, 1987; Lawden, 1983; Nachmias & Weber, 1975) or rating scale (Stromeyer & Klein, 1974) procedure was used. On this basis, the threshold 3f contrast on a f background would be expected to be less than the sum of its simple detection threshold (on a blank background) and 11% of the f contrast. However, the forced choice procedure does not constrain the observer to any particular cue(s). The detection of 3f on a f background could be achieved by using the presence of the thin 3f bars, the reduction of f edge blur, and/or the increase in width of the f bars. Thus the subsequent analysis is relatively complicated. The present experiment will simplify matters, by instructing the observer to use as a cue only the visible presence of the thin 3f bars. One experimental aim is to determine whether a f pedestal will facilitate thin bar detection (of a 3f) under these conditions.

Furthermore, the forced choice procedure factors out the observer's internal criterion, because he is

assumed to respond to the difference between the interval 1 internal response and the interval 2 internal response. Therefore the 2AFC method does not provide criterion position information. The present experiment will use a single interval procedure so that the criterion's position (i.e. the null point) can be measured.

Although Stromeyer & Klein (1974) used single interval, they also used feedback, which would be expected to train the observers to stabilise their criterion at a position that maximised correct feedback. Yet this position may not represent their inherent perceptual category boundary. The present experiment will not use feedback, in order to avoid its influence on the criterion's position.

5.1.3 Method

The stimuli were vertical gratings, consisting of two components: a fundamental cosine-wave (f) and its 3rd harmonic (3f), added in square-wave phase. For a single set of images, the contrast of the fundamental was constant, and the contrast of the 3rd harmonic varied from a low to a high value above its detection threshold (on this f background). The exact range was determined in pilot sessions. The 3f produces a thin dark bar centred on each f peak and a thin light bar on each f trough. A method of constant stimuli was used with a single interval procedure to determine the threshold contrast required to detect the narrow bars (generated by the 3f component) that were centred in each wide bar (generated by the f component). Threshold performance was defined as 84% 'thin bar seen' responses. Waveforms were also inverted in opposite polarity conditions (termed 'f trough central').

The images were 256 × 256 pixels and subtended 3.56 degrees (9.19cm) at the viewing distance of 148.1cm. Their border abruptly changed to a uniform mid-grey background of luminance 39.8 cd/m^2 that filled the rest of the screen. The Michelson contrast of the fundamental was 0, 0.1, 0.2 or 0.4. Its spatial frequency was 1/3 or 1cpd. Thus stimuli contained 1.19 or 3.56 periods of the f component. Stimulus duration was 50 or 300ms. The shorter duration was used to test the hypothesis that the pre-filter + 3rd derivative model would provide a better fit to the 50ms condition data when the weight of the pre-filter's inhibitory surround was substantially reduced. This hypothesis was based on previous physiological (Derrington & Lennie, 1984) and psychophysical work (Georgeson & Georgeson, 1987; Legge, 1978b), which has suggested (from the shape of the CSF) that the weight of the inhibitory surround would be reduced at short duration. Moreover, the results of experiment 4 (chapter 3), which employed short duration stimuli, were consistent with this hypothesis.

There were 28 conditions in total: 24 conditions for the conditions where the fundamental contrast was not zero (3 contrasts, 2 spatial frequencies, 2 durations, 2 polarities) plus 4 conditions where the fundamental contrast was zero (2 spatial frequencies, 2 durations). For each condition, the 3f contrast levels were obtained from pilot data, and were selected to ensure that the percentage of positive responses ranged from at least 10% to 90%. The number of 3f contrast levels in each condition was between 8 and 13.

Each trial began with a 500ms presentation of a central black fixation dot (2 × 2 pixels) on a blank mid-grey screen. This was followed by 300ms or 50 ms presentation of the stimulus, without the fixation point. The stimulus was then replaced by the blank screen with fixation point. For conditions where both f and 3f components were present, the task was to indicate the presence or absence of a central thin bar on the central broad bar background (bar detection). For conditions where only the 3f component was present, the task was to indicate the polarity of its central bar (polarity identification).

The trial ended with a 500ms pause, to reduce the effect of after-images upon the next stimulus. The observer had unlimited time to respond, and was given no feedback of the correctness of response.

Each session consisted of all of the trials for a single condition interleaved randomly with its opposite polarity pair. Each session contained ten presentations of each 3f contrast level for both polarities, in a randomised order, and took about 8mins to complete. The data from the first session of each pair of conditions was discarded as practice. The observers completed a further 6 or 8 sessions for each condition-pair in a block. The blocks were presented in a randomised order. There were two observers.

5.1.4 Results

For the 24 conditions where f and 3f were both present (task was bar detection) psychometric functions were plotted separately for each condition. Plots were of percentage of 'thin bar present' responses against 3f contrast, and an example for a pair of conditions is shown in figure 5.3 (top right). For the 4 conditions where only 3f was present (task was polarity identification), data from both stimulus polarities were used in a single plot. These plots were of percentage of 'light bar' responses against contrast, where positive contrast indicated a central light bar, and negative contrast indicated a central dark bar. An example plot is shown in figure 5.3 (top left). Data from both polarities were shown on a single plot because 1) they are two halves of one condition-pair, and 2) all of the polarity identification conditions included zero contrast stimuli (blank screens). Since stimuli of both polarities were interleaved between trials, the allocation of half of the blank screen responses to each polarity would be entirely arbitrary. So all of the blank screen data for a given session were pooled.

The psychometric functions were fitted using the psychometric fitting program described in section 2.5.1. Curves commonly used in this circumstance include Gaussian integral, Logistic, Power or Weibull. The Gaussian integral has a long history in vision science, as responses along many scales have often followed a normal distribution. Furthermore, the noise in a response to a visual stimulus may arise from many independent sources, and the central limit theorem states that the sum of a large number of independent random variables, each with finite mean and variance, will tend to have a Gaussian distribution. This is the case even when the individual random variables do not have a Gaussian distribution.

Separate psychometric functions were also plotted in z-units, because linear performance would be immediately apparent as a linear psychometric function. In contrast, linear performance cannot be surmised from simple inspection of a plot of percentage values because that psychometric function still has a sigmoidal shape whether performance is linear or not. The linearity of performance will be seen to emerge as an interesting issue in this chapter, and forms the basis of chapter 6.

In order to determine whether it was appropriate to fit with a Gaussian function here, the linear axes z-score plots were individually examined by eye for linearity. Also, the correlation coefficient was calculated between each set of z-scores and the contrast of the thin bar. For observer RJS, $r \geq 0.94$ in all conditions except 4, where the data appeared to be linear except for a floor and/or ceiling effect. An example of this pattern is shown in figure 5.3 (lower right), where RJS's z-scores do not fall below about -2. For observer SAW, $r \geq 0.95$, in all conditions except one, which showed a ceiling effect. The floor/ceiling effects suggest that any curve-fitting algorithm applied to these data should incorporate a parameter that allows for lapses. The linearity of the z-scores not subject to lapses suggests that a Gaussian can be fitted to the data. The values of r decreased by an average of 0.06 for SAW and 0.10

for RJS when a logarithmic scale was used for the contrast of 3f. Therefore all further analysis used a linear contrast axis.

Most of the polarity discrimination data for both observers showed a central kink, as illustrated for SAW in figure 5.3 (left). This is characteristic of an accelerating non-linear transducer (Legge & Foley, 1980) that maps contrast onto response at the low contrast levels used here. These curves were not fitted with a simple Gaussian integral, as it would not capture the central kink. Rather, the fitting program passed the contrast levels through an accelerating nonlinear transducer, before adding the Gaussian noise, as described in chapter 2 (section 2.5.1).

Initial fits allowed all parameters to be free. However, for observer SAW, the threshold confidence limits for several fits were very wide. But they were much improved, with little alteration to the resulting thresholds, by constraining the fits. The lapse rate was fixed at 0.5% for the bar detection data, and constrained to an upper limit of 30% for the polarity identification data. For 27 of the 28

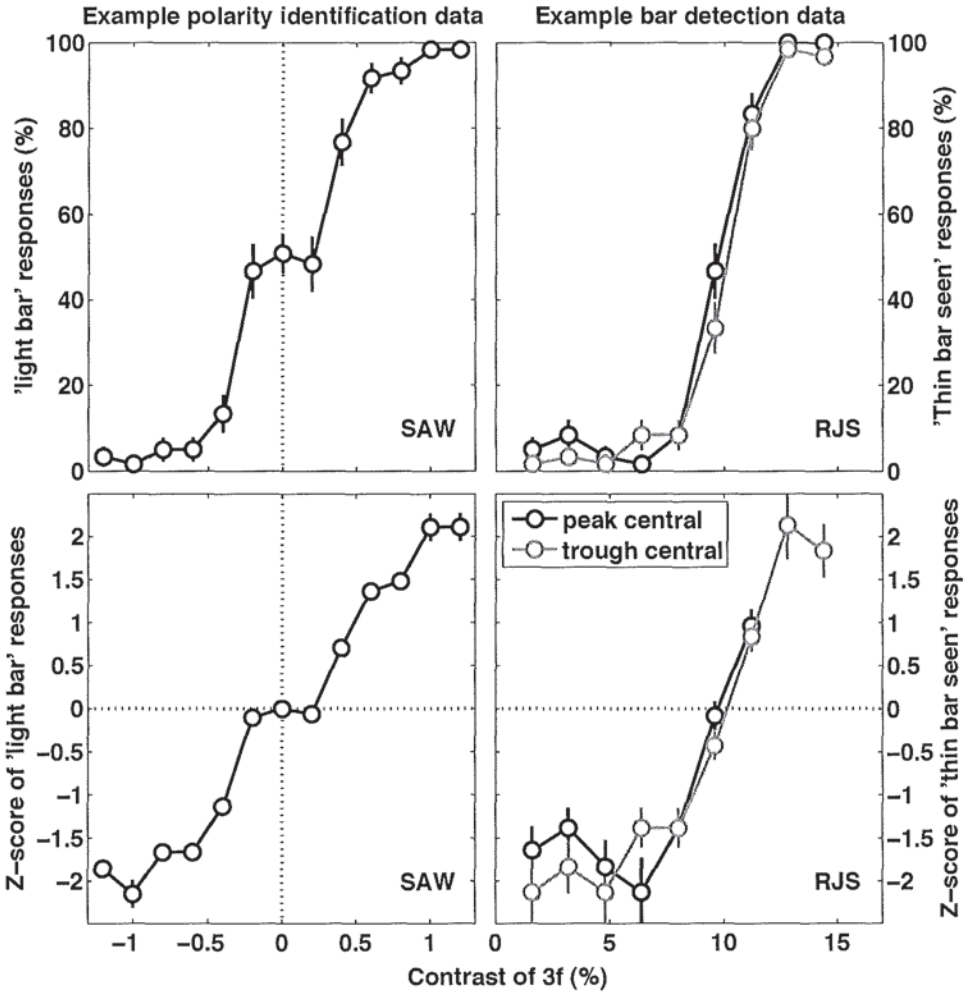

Figure 5.3: Example experimental data from 2 specific conditions. Top left panel shows percentage of 'light bar' responses for observer SAW in the 1cpd, 300ms condition. Peak central polarity data are positive and dark central polarity data are negative. Top right panel shows the percentage of 'thin bar seen' responses for observer RJS in the f3 = 3cpd, f contrast = 0.4, 300ms condition-pair. The lower panels show the z-scores of the upper panels.

fitted curves, the deviance of the fitted curve from the data fell within the 95% confidence limits. For observer RJS, all parameters were free, and the threshold confidence limits were relatively narrow. For 27 of the 28 fitted curves, the deviance fell within the 95% confidence limits. Attempts to constrain the fits further worsened the fit to some of the data for this observer.

Figure 5.4 shows all of the data and fitted curves for observer SAW. Each column shows a separate f contrast, and each row shows a different combination of f spatial frequency and stimulus duration. The left three columns show the bar detection data. The grey symbols refer to the data where the f component had a peak at the stimulus centre, and the black symbols refer to the f trough central data. There is little difference between these two polarities, in most of these 12 plots. Pearson's X^2 statistic (without Bonferroni correction) showed no significant difference between the 2 data sets in each of these panels, except for the two denoted by a $p(X^2)$ value. A Bonferroni correction was applied by dividing the criterion probability level by 5. The value of 5 was chosen because inspection of figure 5.4 suggested that only 5 of the bar detection panels were likely to show a polarity effect; the three on the 2nd row, and the lower two panels in the 40% column. After this correction was applied, only the panel endorsed with $p(X^2) < 0.01$ remained significant at the 1% level.

For all bar detection data, the null point (c_{null}) between seeing and not seeing the thin bar was assumed to be the contrast level that resulted in a 50% response. The threshold contrast level (c_{thd}) was defined as 1 z-unit above the null point, which corresponds to a 84.1% response. The null point and threshold are subject to the effected response bias (e.g. for 'dark bar' responses), but if the observer's bias is consistent across conditions, then the bias can be factored out by averaging across polarity (f peak central and f trough central).

The right column shows the data for polarity identification of the 3f sinusoidal grating. The data and fitted curve in 3 panels show a central kink that characterises nonlinear performance. A response bias was shown to shift the central kink vertically, in chapter 1 (section 1.11). In figure 5.4, the kink falls below the 50% response level, which indicates a bias for dark bar responses. If the null point and threshold contrast levels here were defined as above (at 50% and 84.1% responses), then the response bias would cause both contrast levels to be artifactually high. While it would be possible to factor out the bias by averaging across polarity, the odd-symmetrical nature of the accelerating z-score curve means that an equivalent correction is achieved by assigning the null point to be at zero contrast and the threshold to be 1 z-unit above that. Therefore, for these polarity identification data, c_{null} and $cthd$ were defined in this way.

Figure 5.5 shows all of the data and fitted curves for observer RJS. Compared to observer SAW, higher contrast levels were required, and there are now 4 panels where the Pearson's X^2 statistic was significant, without Bonferroni correction. A Bonferroni correction was applied by dividing the criterion probability level by 5; the five plausible panels for a polarity effect were the 4 endorsed with the Pearsons statistic and the panel in the 2nd row, 3rd column. After correction, the only significant polarity effects, at the 1% level, were in the row for the 3f=1cpd, 50ms conditions. This is the same row where observer SAW showed a significant Pearson's X^2 statistic, and therefore suggests a polarity effect that is specific to these conditions. The plots in the right column of figure 5.5 show a more pronounced central kink for RJS than for SAW, which indicates a greater degree of nonlinearity. Also, the kink is lower, which shows a greater bias for 'dark bar' responses.

The psychometric fit program that produced the curves in figures 5.4 and 5.5 had 4 free parameters (c_{null}, σ, $lapse$ and m). c_{null} was the 3f contrast at the null point, σ was the standard deviation of the

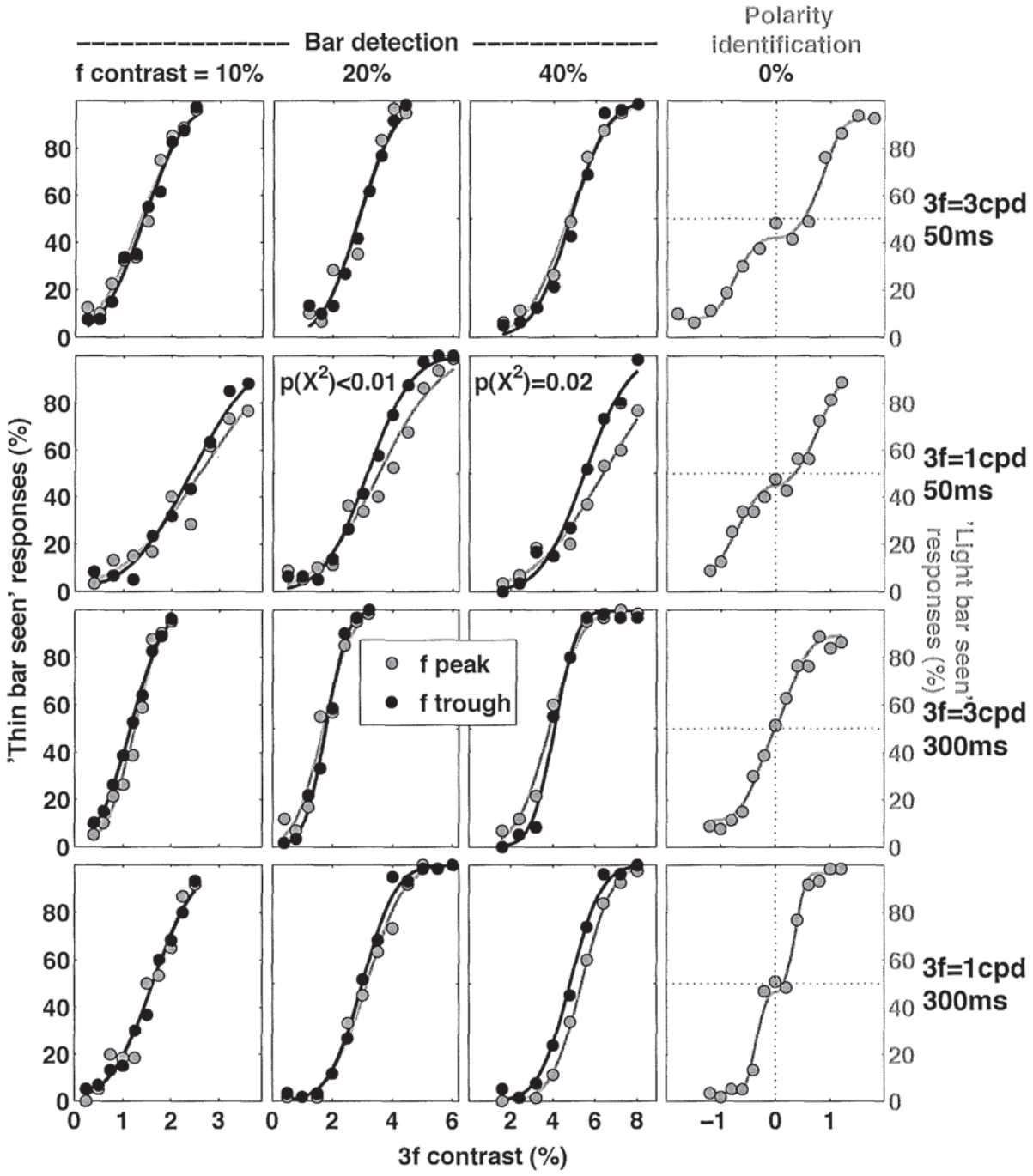

Figure 5.4: Percentage of 'thin bar' or 'light bar' responses (symbols) as a function of 3f contrast, with fitted curves (lines) for observer SAW. Each column shows a separate f contrast, and each row shows a different combination of f spatial frequency and stimulus duration. The legend refers to the polarity of f. A Bonferroni correction has not been applied to the the Pearsons X^2 statistic shown in the panels.

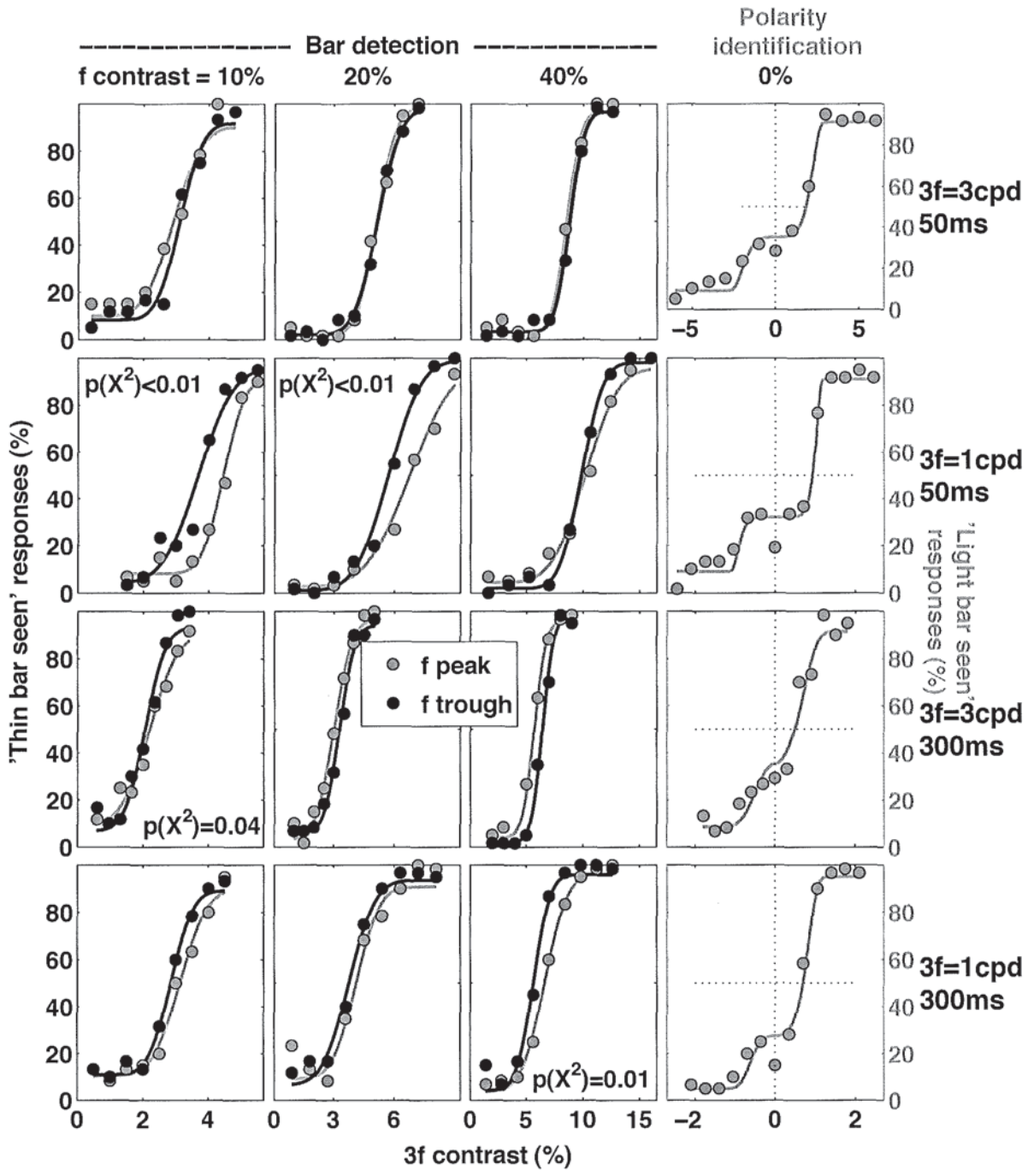

Figure 5.5: Percentage of 'thin bar' or 'light bar' responses (symbols) as a function of 3f contrast, with fitted curves (lines) for observer RJS. Each column shows a separate f contrast, and each row shows a different combination of f spatial frequency and stimulus duration. The legend refers to the polarity of f. A Bonferroni correction has not been applied to the the Pearsons X^2 statistic shown in the panels.

fitted Gaussian, *lapse* was the lapse rate, and m was the transducer exponent. For the bar detection data (left 3 columns), m was fixed at 1. The program also returned the threshold contrast level (c_{thd}).

Plots of c_{null} (circles) and c_{thd} (squares) are shown for observer SAW in figure 5.6. Dashed lines show

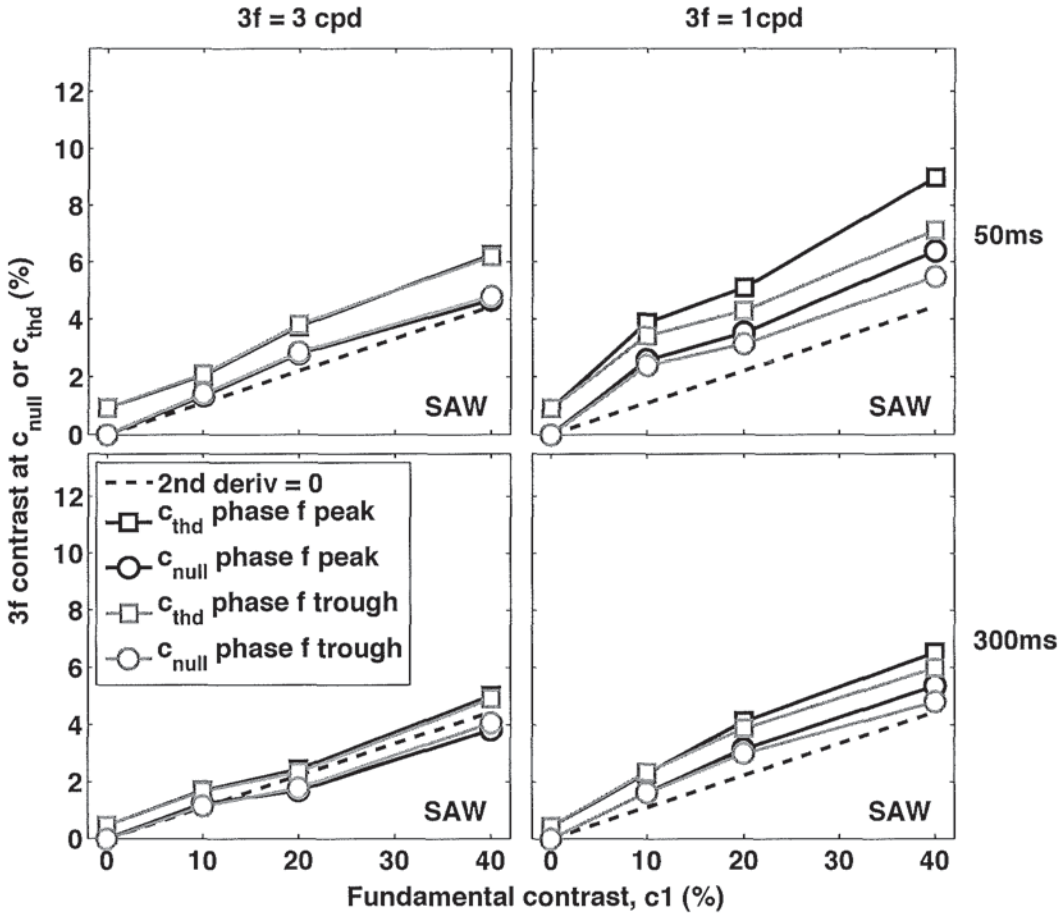

Figure 5.6: Threshold (squares) and c_{null} (circles) contrast levels for bar detection, or polarity identification of the 3f component, for observer SAW. Dashed lines show c_{null} predictions based on zero 2nd derivative values. Column headings refer to the 3f spatial frequency. Each row shows data from a given stimulus duration.

c_{null} predictions, representing 3f contrasts at which 2nd derivative values are zero (ie. where 3f contrast is 11% of f contrast). Column headings refer to the 3f spatial frequency. Each row shows data from a given stimulus duration. Plots with the same format were produced for observer RJS, and are shown in figure 5.7. The c_{null} data are generally close to the '2nd derivative = 0' line for observer SAW, but are above it for observer RJS. The c_{thd} data are all higher than their corresponding c_{null} data points, by an increment that generally increases slightly with fundamental contrast. One exception to this trend is caused by RJS's c_{thd} point for 3f = 3cpd, duration 50ms, phase f peak and zero fundamental contrast (leftmost square in the top left panel in figure 5.7), which is much higher than the same point in the other 3 panels.

In order to examine the degree to which the f component facilitated 3f bar detection, the data were normalised and replotted in figure 5.8. The normalisation process was similar to that used to normalise

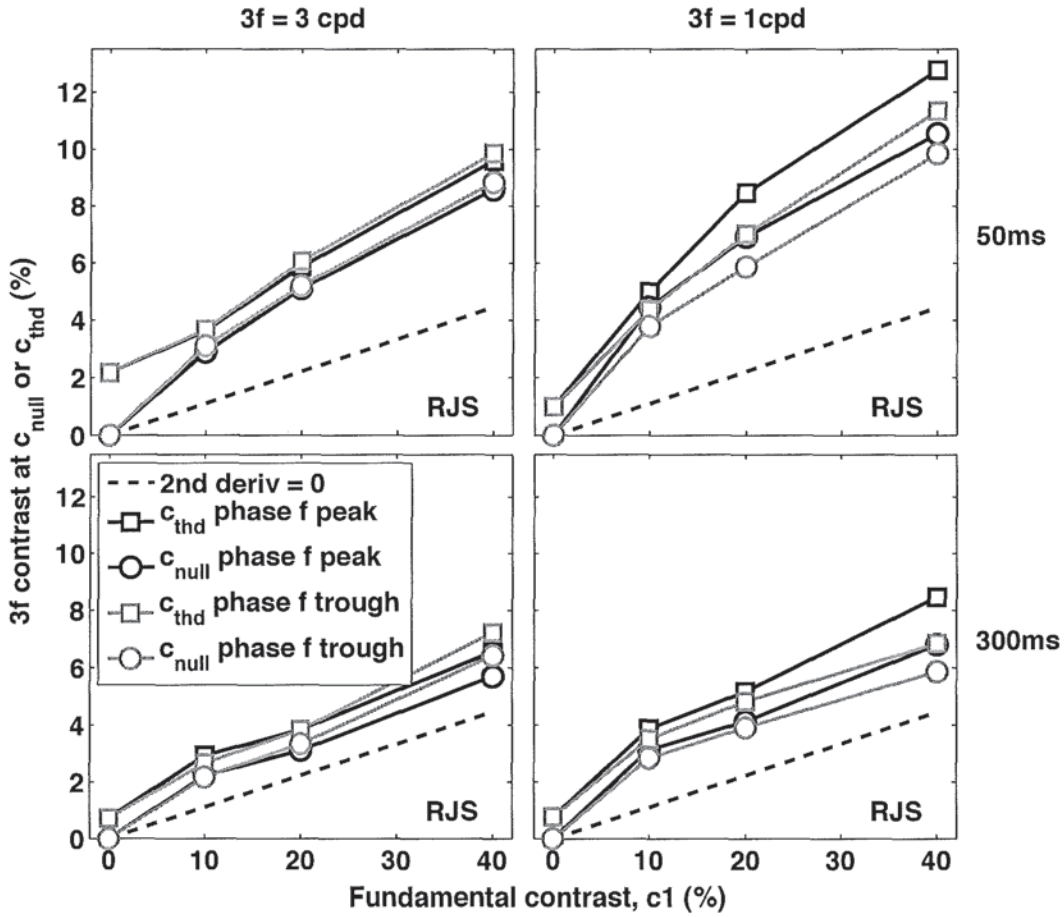

Figure 5.7: Threshold (squares) and c_{null} (circles) contrast levels for bar detection, or polarity identification of the 3f component, for observer RJS. Dashed lines show c_{null} predictions based on zero 2nd derivative values. Column headings refer to the 3f spatial frequency. Each row shows data from a given stimulus duration.

increment threshold contrast -v- mask contrast 'dipper' functions (Legge, 1979), where data are divided by the contrast detection threshold. In figure 5.8, the ordinate shows the c_{thd}-c_{null} increment divided by the c_{thd} value for the polarity identification task, termed c0 in the figure. The abcissa shows c_{null} divided by the same c0 value. The separate panels shows data from each observer. Symbol shade represents stimulus polarity. Each dashed line is the best (least squares) linear fit to the data. The data all fall on, or close to, the dashed line (on these logarithmic axes) of slope 0.63 for SAW and 0.92 for RJS.

5.1.5 Discussion

One result of this experiment was that the f trough central psychometric functions were significantly shifted to lower contrast levels than the f peak central psychometric functions, in 4 of the 6 panels that show the 3f=1cpd, 50ms data. This polarity effect appears to be mainly specific to this lower spatial frequency and short duration. In figures 5.4 and 5.5 this leftward shift can be seen to a smaller

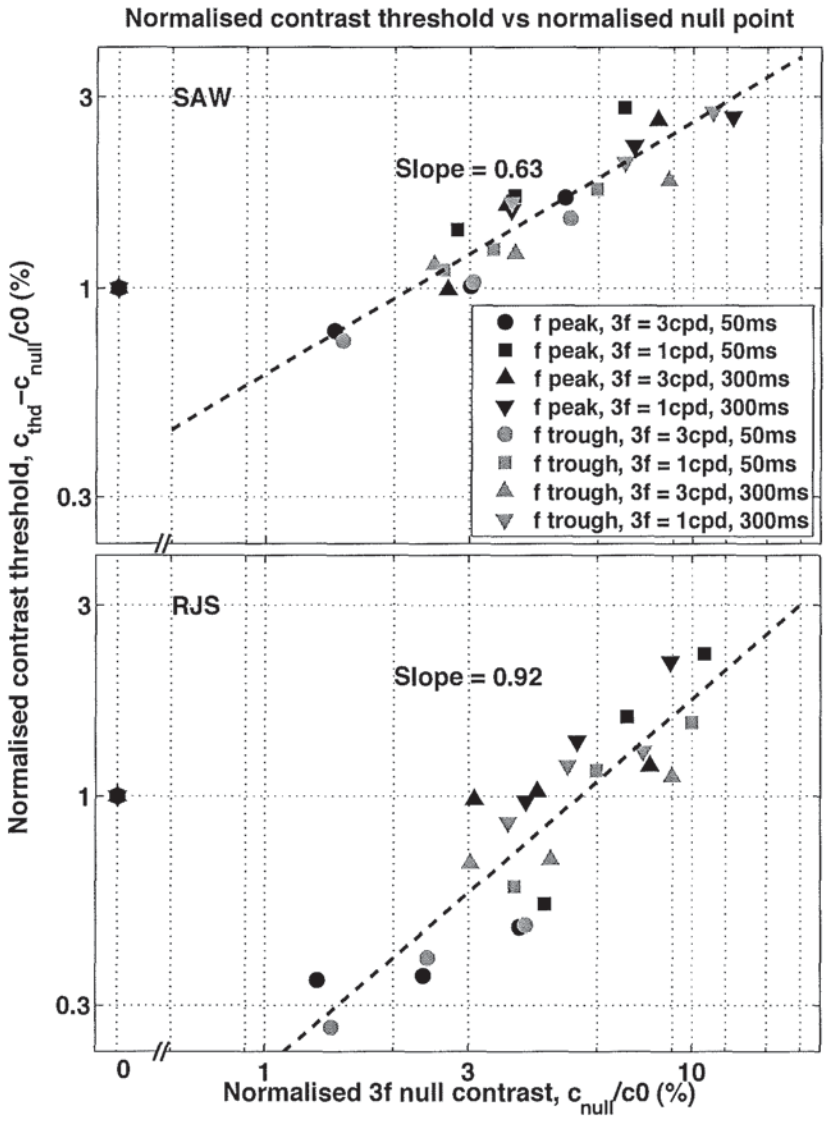

Figure 5.8: Normalised contrast threshold as a function of normalised c_{null} contrast. Upper panel shows data from observer SAW and lower panel shows data from RJS. Symbol shade represents stimulus polarity. Each dashed line is the best (least squares) linear fit, and resembles the 'handle' of a traditional threshold -v- contrast 'dipper' function.

extent in the 3f=1cpd, 300ms conditions (lowest row) and it is absent in the 3f=3cpd, 50ms conditions (top row). This suggests that the spatial frequency factor is slightly more important than the duration factor in accounting for the polarity effect. The existence of this effect is consistent with a compressive transduction of local luminance (Mather & Morgan, 1986), as that would reduce the effective contrast of higher luminance features (i.e. f peak central) more than lower luminance features (i.e. f trough central). However, this account on it's own cannot explain why the polarity effect is reduced or absent at a higher spatial frequency and long duration. But a process of local adaptation may provide a suitable account, as follows.

A well-established property of the early visual system is that of gain control. It has been shown to

depend on the amount of luminance flux collected by the central region of retinal ganglion cells in cat (Enroth-Cugell & Shapley, 1973b). The time course of this process is known to be rapid (Enroth-Cugell & Shapley, 1973a; Saito & Fukada, 1986; Shapley, 1986), so can be assumed to be complete by 300ms, but may still be ongoing at 50ms. Therefore the 300ms stimuli in the present experiment would be subject to the full effect of gain control, which would be expected to cancel out the effect of background luminance level which is dominated by the f component, and thus lead to the observed absence of a polarity effect. The 50ms stimuli, however, would not be so strongly affected by gain control, and would therefore be expected to be subject to background luminance effects.

In order to account for the observed spatial frequency dependance of the polarity effect, it can be proposed that the 'gain pool' that inputs to the gain control process has a fixed width of about 1/3 degree. This would encompass 1 cycle of the 3f=3cpd stimulus, which means that the mean luminance here is at the background level. Therefore the responses to f peak and f trough stimuli would be expected to be similar. But for the 3f=1cpd stimulus, the gain pool encompasses only the f central bar. This would lead to different responses to the f peak and f trough stimuli, but only in the 50ms conditions, where gain control is incomplete. A fixed-width gain pool was suggested by Kingdom & Whittle (1996) to account for a dip-down of contrast discrimination thresholds at high contrasts that was present for 4cpd and 8cpd gratings, but not for 0.0625cpd and 0.5cpd. Although they do not specify the gain pool's width, it seems likely that it should encompass at least 1 cycle of a 4cpd grating and no more than 1/2 cycle of a 0.5cpd grating. Thus it would fall within the range 0.25 - 1 degree.

The suggested gain pool width of about 1/3 degree is also consistent with the work of Kelly & Martinez-Uriegas (1993). They used a method of contrast cancellation, whereby a sinusoidal grating afterimage was nulled by an opposte phase grating (of identical spatial frequency and orientation), whose contrast was varied in order to produce no apparent contrast change at afterimage onset. The experimental factors were spatial frequency and contrast of the adapting grating. The nulling contrast was plotted as a function of adapting contast separately for each spatial frequency and observer. The slopes of these functions followed a low-pass pattern that fell rapidly above about 1cpd. The highest spatial frequency that produced a measurable afterimage was 3cpd for one observer and 4cpd for the other observer. But the adapting contrast here was at least 60%, which is above the maximum f contrast level used in the present experiment (40%). When the adating contrast was 20% (the lowest level used in Kelly & Martinez-Uriegas' experiment), the highest spatial frequency grating that produced a measurable afterimage was 1.5cpd for one observer and 2cpd for the other. When the spatial frequencies were increased to 2cpd and 3cpd no measurable afterimages were produced at this lower contrast. If this lack of an afterimage was due to spatial summation across 1 grating cycle, then this result is consistent with a spatial integration width of 1/3 to 1/2 degree.

One aim of this experiment was to test the hypothesis that the null point between perceiving and not perceiving the thin bar caused by the 3f component would be at the 3f contrast that caused the waveform's 2nd derivative to be zero at the centre. If this prediction was correct then the c_{null} data (circles) in figures 5.6 and 5.7 should lie on or close to the dashed line, which shows zero 2nd derivative. While figure 5.6 shows that this appears to be approximately true for much of SAW's data, figure 5.7 shows that RJS required a greater 3f contrast at the null point. Data from other observers would be required to determine whether SAW's or RJS's data are typical. But at present, the null point prediction is not consistently supported.

It might be expected that the thresholds for perceiving the 3f component (c_{thd}) would fall above the

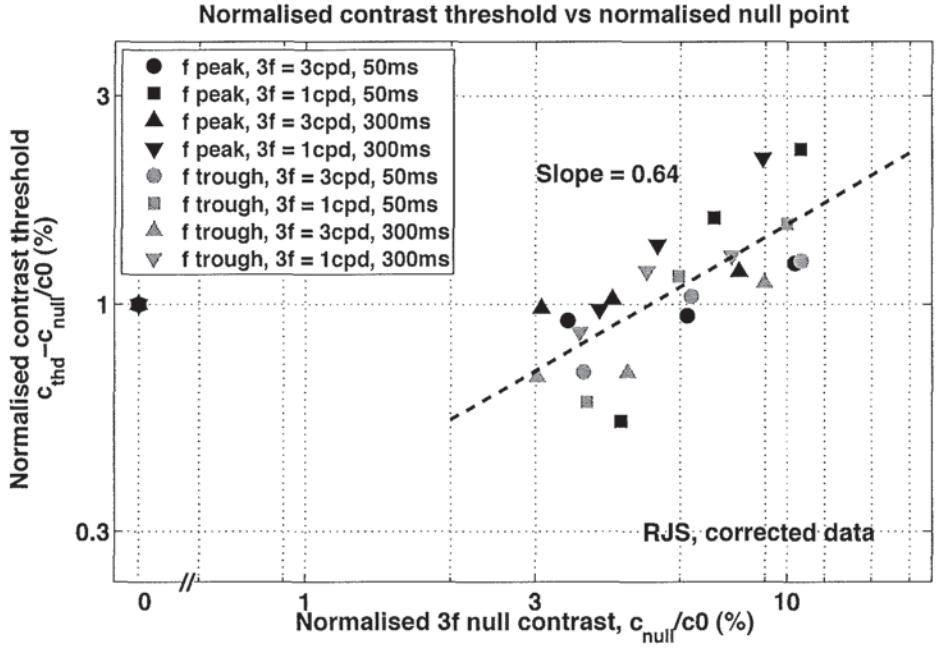

Figure 5.9: Normalised contrast threshold as a function of normalised c_{null} contrast, for RJS's data that have been corrected for one outlier.

c_{null} thresholds by a constant amount that is related to the 3f polarity discrimination threshold. The 3f polarity identification threshold is plotted as leftmost square in each panel of figures 5.6 and 5.7. The degree to which the c_{thd} data are parallel to the corresponding c_{null} data shows that this expectation is approximately true. However, a closer examination shows that for most of the data, the c_{thd} data rise slightly more rapidly than c_{null}. This trend was examined more closely in figure 5.8, which shows the contrast increment between the c_{null} and c_{thd} data plotted against c_{null}, with all data normalised by the 3f polarity discrimination threshold. The fact that the best-fitting linear functions are not horizontal demonstrates that the contrast increment is not constant across conditions.

The linear functions resemble the 'handle' of a traditional threshold -v- contrast 'dipper' function, and have a slope of 0.63 for SAW and 0.92 for RJS. The relatively high slope for RJS is largely due to the high 3f polarity discrimination threshold (2.2%) in the 50ms, 3f = 3cpd condition (figure 5.7, top left panel). RJS obtained a threshold of 0.72% to 0.99% in the other three 3f polarity discrimination conditions, which suggests that the threshold of 2.2% may represent an outlier. This assumption is further supported by the commonality between the same four polarity discrimination thresholds for SAW, which all fall in the range of 0.43% to 0.93%. In order to correct for RJS's outlier, it was replaced by the mean of his other three 3f polarity discrimination thresholds, and the normalised data re-plotted in figure 5.9. The slope of the best-fitting linear function (0.64) is now very similar to SAW's slope (0.63). On the log-log axis shown here, this slope suggests that the normalised contrast threshold is approximately equal to the normalised 3f null contrast raised to an exponent of 0.64. If we treat the 3f c_{null} contrast as a pedestal on which an increment is to be detected, then this value is consistent with the contrast discrimination literature, e.g. Legge & Foley (1980); Meese (2004); Meese et al. (2007).

In chapter 3, we saw that a DoG pre-filter was required to model the appearance of Mach edge stimuli with a luminance ramp. Its effect was to accentuate the curvature of the luminance profile at the

junction of 2 luminance ramps (of different gradients). Moreover, in chapter 4, the same pre-filter in combination with the single-scale 3rd derivative model provided a good fit to the reported proportions of Mach bands seen. Given the previous success of the pre-filter, a second aim of the present experiment was to determine whether the luminance waveforms would require this pre-filter before subsequent analysis revealed a coherent pattern of results.

The pre-filter's parameters were fixed at the values that provided the best fit to the previous data ($\sigma_c = 1, \sigma_s = 10, k = 0.55$). It was applied to the f + 3f experimental waveforms where the 3f contrast level was set at the value that provided a 50% response (c_{null}) for each of the 24 conditions that contained the f and 3f components, separately for each observer. In every case, the pre-filter's effect was similar to an increase in 3f contrast; for 'f peak central' stimuli, it deepened the central dip in the waveform caused by the 3f component. In order to quantify this effect, the amplitude of the 2nd derivative of each waveform was calculated at the waveform's centre. The 'f trough central' 2nd derivative amplitudes were generally similar to the negative of their corresponding 'f peak central' values, so the former values had their sign inverted and were then averaged with the 'f peak central' values. These averaged data are shown in figure 5.10. The columns show data without (left) and with (centre and right columns) the pre-filter. The centre column shows the pre-filter where its inhibitory surround had a weight (k) of 0.55 in all conditions. In the right column, k was 0.55 in the 300ms conditions, but was reduced to 0.25 in the 50ms conditions, for reasons given below. Error-bars show ±1se and were derived from data of each polarity (n=2). The top left panel shows data from RJS without the pre-filter. The 2nd derivative amplitudes are all positive, which indicates that the waveforms always had a small central dip (f peak central) or peak (f trough central). The top centre panel shows that when the waveforms were passed through the pre-filter, the 2nd derivative amplitudes increased by a factor of approximately 3.5, but their pattern remained largely unaltered.

The aim of this analysis was to attempt to bring the whole set of data onto a single function, which would then define the underlying 'rule' behind bar detection under these experimental conditions. However, the 2nd derivative amplitudes with (or without) the pre-filter do not converge onto a common function for either observer. In an attempt to improve on this, the weight of the inhibitory surround (k) was reduced below 0.55 in only the 50ms conditions. This manipulation was based on the finding that the inhibitory surround of both magnocellular and parvocellular retinal ganglion cells is relatively less sensitive to higher temporal frequencies, when compared to the cells' central region (Derrington & Lennie, 1984). Furthermore, psychophysical work has found that the low frequency cut in the CSF is rather mild at short durations (25-100 ms) compared with long durations (500-1000 ms) (Georgeson & Georgeson, 1987; Legge, 1978b). The right column of figure 5.10 shows the effect of $k = 0.25$. This value of k provided the best overlap, by eye, of the two sets of f3=3cpd data in each panel (black symbols).

The pre-filter has brought the f3=1cpd data of each panel onto a common horizontal function. Furthermore, the selective pre-filter has approximately brought the f3=3cpd data of each panel onto a common linear function, and has brought the data of each observer into an almost identical pattern, differing only in scale. These features suggest that a selective pre- and 2nd derivative filter does partially characterise perception of the experimental stimuli. However, the f3=3cpd data in the right column of figure 5.10 still fall above the f3=1cpd data, and progressively increase with f contrast. This suggests that there are other factors that need to be taken into account, one or more of which depend on f contrast and/or stimulus spatial frequency.

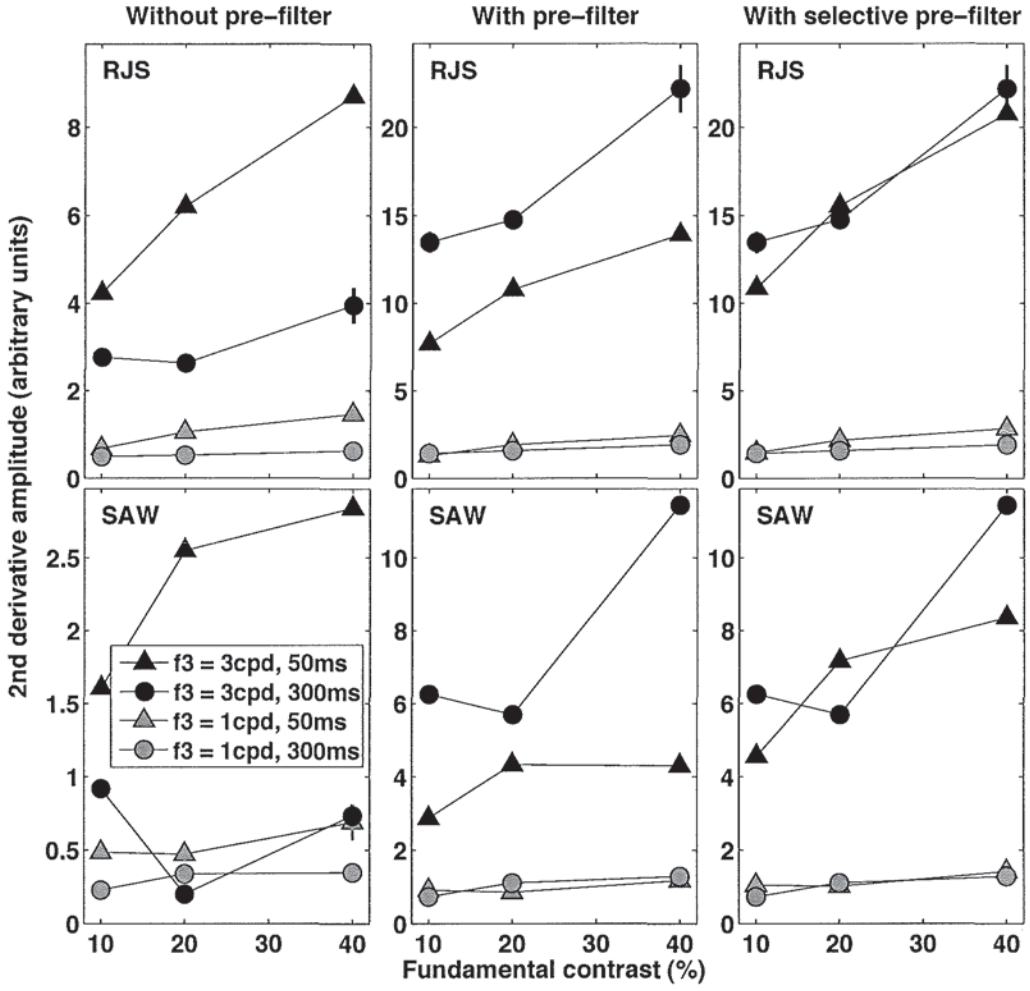

Figure 5.10: Amplitude of the 2nd derivative at the centre of the 84% response waveforms, averaged across image polarity, without pre-filter (left column) and with the pre-filter with full (centre column) and selective (right column) inhibitory surround. Error-bars show ±1se and were derived from data of each polarity (n=2). The rows show data from each observer.

If a selective pre- and 2nd derivative filter does characterise perception of the thin bar at the 84% threshold, to some extent, then it might be expected that when the same filter is applied to the 50% response waveforms, a common pattern of data would also be obtained. Figure 5.11 shows the result of the filter on the 50% response waveforms. The pattern in the right column is the same as for the 84% response waveforms, but shifted to lower 2nd derivative amplitudes.

The bottom left panel shows data from SAW, without the pre-filter. Here, there are 2 data points that fall below zero, which indicates that these waveforms retained their central curvature caused by the f component; they did not form a plateau. The bottom centre panel shows that when the waveforms were passed through the pre-filter, the 2nd derivative amplitudes all become more positive, including the two that were previously negative. Thus, the pre-filter has caused a small central dip (f peak central) or peak (f trough central) in the two waveforms that previously had none. The pre- and 2nd derivative filter has produced a consistent pattern of data even where some waveforms did not form a

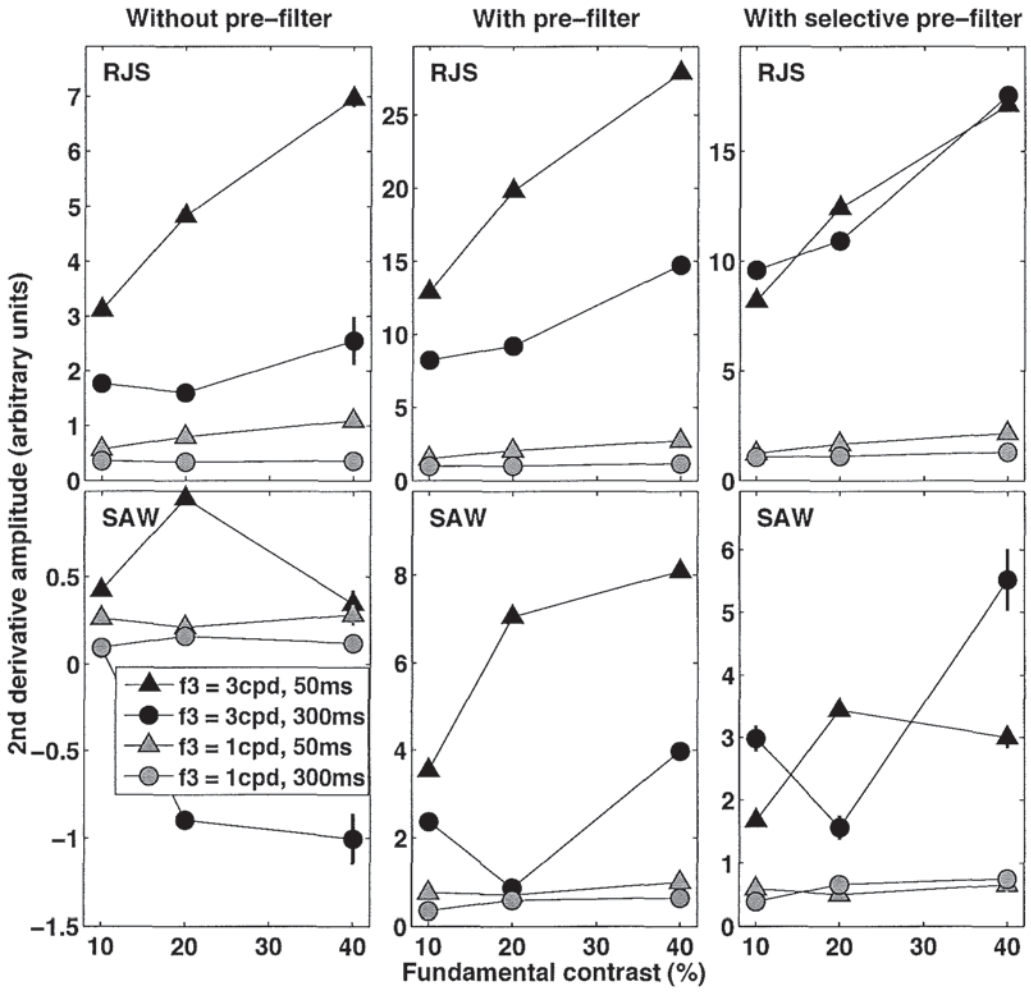

Figure 5.11: Amplitude of the 2nd derivative at the centre of the 50% response waveforms, averaged across image polarity, without pre-filter (left column) and with the pre-filter with full (centre column) and selective (right column) inhibitory surround. Error-bars show ±1se and were derived from data of each polarity (n=2). The rows show data from each observer.

plateau. This demonstrates the robustness of this filter.

The scale-space edge-bar model was the best fitting model to the Mach band data in chapter 4. The question arose whether it might fit the data here, particularly as it uses 2nd derivative filters to derive bar locations. However, when the 84% response waveforms were passed through it, a central thin bar was not predicted for several of the waveforms. This problem was rectified by reducing the model's small threshold that is applied to the feature strength. But this reduction then caused spurious feature to be predicted, such as 3 central bars of the same polarity, but of 3 different scales. Further fits of the scale-space edge-bar model were attempted with the selective pre-filter. However, the same problems arose. The failure of this approach reinforces the notion that performance is not characterised solely by the amplitude of the 2nd derivative; other factors need to be added to the modelling.

The right columns of figures 5.10 and 5.11 suggest that the missing factors relate to the fundamental

contrast and stimulus spatial frequency. The generally linear pattern of the data in the normalised contrast increment functions shown in figures 5.8 and 5.9 suggest a lawful relation between the 3f contrast at the 50% response level and the incremental 3f contrast that drove performance up to 84%. In these figures, each symbol is plotted 3 times; once for each fundamental contrast level. In every condition, the 3 identical symbols are plotted in order of increasing f contrast from low to high contrast levels. So an encapsulation of this lawful behaviour will need to incorporate f and 3f contrast levels. However there are a number of possible interpretations.

The plots in figures 5.8 and 5.9 resemble contrast discrimination 'dipper' functions. Legge & Foley (1980) collected contrast discrimination data for sinusoidal gratings of various spatial frequencies proposed that the visual signal was passed through a non-linear transducer of the form

$$R = \frac{gc^{2.4}}{Z + c^2} \tag{5.7}$$

where c is the stimulus contrast, and g and Z are parameters that relate to the transducer's gain and its degree of acceleration at low contrasts. However, when applying this equation to the present experiment, it is not immediately obvious whether one or both of the c terms should be replaced by the contrast of f, or 3f, or some combination of both. However, Foley (1994) subsequently showed that the denominator in equation 5.7 required the addition of a term for mechanisms of non-target orientation. This could be interpreted in the present experiment as a non-target spatial frequency, thus:

$$R = \frac{gc_3^p}{Z + c_1^q + c_3^q} \tag{5.8}$$

where c_1 and c_3 are the contrasts of the f and 3f components respectively, and p and q are model parameters. Equation 5.8 has 4 parameters. But now that two spatial frequencies are included, we also need a term that reflects the model's relative sensitivity to each one. Furthermore, there are 3 parameters included in the pre-filter, and the scale of the 2nd derivative operator is a possible further parameter. Thus we now have a total of 9 free parameters. Whilst it is possible to build such a complicated model, its consideration is beyond the scope of this thesis. Furthermore, with so many free parameters, there is a real danger of over-fitting. The situation is complicated further by the larger number of stimulus cycles in the f=3cpd than the f=1cpd conditions, which may preferentially reduce thresholds because of probability summation (Graham et al., 1978; Legge, 1978b,a; Meese & Williams, 2000; Meese, 2004; Robson & Graham, 1981; Williams & Wilson, 1983). However the next experiment removes this unwanted complication.

In an effort to explore other, more simple alternatives, the 50% and 84% response waveforms were passed through a program that computed a measure of local contrast (C) that was inspired by Badcock (1984b), defined as

$$C = \frac{L_c - L_b}{L_c + L_b} \tag{5.9}$$

where, L_c was the luminance at the image centre and L_b was the luminance at the adjacent pair of peaks or troughs. Two of SAW's waveforms did not form a central little dip (or little peak), so they were omitted from this analysis. For each observer, the difference between the contrast levels derived from the 84% and 50% response waveforms was then plotted as a function of f contrast, separately for the 4 combinations of stimulus spatial frequency and duration. The aim was to determine whether these 4 functions would coincide. However they did not (not shown). This process was repeated for

Michelson and Weber definitions of contrast, with a similar result. An other likely stimulus cue is the luminance gradient between adjacent extrema (Hess & Pointer, 1987). However, when this measure replaced local contrast, the resulting plots showed no striking pattern.

5.2 Experiment 7 - Bar detection on a Gaussian background

5.2.1 Introduction

The previous experiment used f +3f waveforms to examine thresholds for polarity discrimination of the 3f component. These cosine-waves contained more than one bar. So although the task involved only the central bar, the neighbouring bars may also have been influencing performance due to probability summation. Futhermore lateral suppression may also have had an influence (Chubb et al., 1989; Cannon & Fullemkamp, 1991; Adini et al., 1997; Yu et al., 2001). In order to eliminate this complication, the present experiment uses stimuli that consist of one small Gaussian dark bar resting on a larger Gaussian light bar, or a small light bar resting on a large dark bar. The aim is to characterise bar perception and to determine whether a suitably designed model can account for the results.

5.2.2 Method

The method was very similar to that used in the previous experiment; the main change being the stimulus. A method of constant stimuli was used with a single interval procedure, without feedback. The stimuli were vertical 1D images, where the luminance (L) at each position (x) was given by

$$L = c_1 exp\left(\frac{-x^2}{2\sigma_1^2}\right) - c_2 exp\left(\frac{-x^2}{2\sigma_2^2}\right) \tag{5.10}$$

This represents a small inverted Gaussian (σ_2 = 2, 4 or 6mins) of various contrast levels (c_2) sitting on a large upright Gaussian (σ_1 = 12mins) of fixed contrast (c_1). σ_1 was fixed at 12mins because that Gaussian's width at half height approximates half a period of the 1cpd fundamental sine-wave used in the previous f + 3f experiment (calculation shown below). So the Gaussian bar appeared to be of the same width as a wide bar in one set of the f + 3f images. Similarly, a σ_2 of 4mins was used for the thin inverted Gaussian bar because it appeared similar to the narrow bars in the f + 3f images.

The wide bar contrast (c_1) was fixed at 0, 0.1, 0.2 or 0.4, and the narrow bar contrast (c_2) ranged from zero to above its detection threshold in 11 to 13 levels (values determined in a pilot). Waveforms were also inverted to form opposite polarity conditions (termed 'trough central'). An example of a set of peak central waveforms is shown in figure 5.12.

A Weber definition of contrast was applied, given by

$$\text{contrast} = (L_c - L_b)/L_b \tag{5.11}$$

where L_c is the luminance at the image centre, and L_b is the background luminance. For peak central images, $L_c = L_{max}$ and for trough central images, $L_c = L_{min}$.

The images were 256 pixels square and subtended 4.267 degrees (9.19cm) at the viewing distance of

123.4cm (1 pixel = 1 minute). The image border abruptly changed to a uniform mid-grey background of luminance 36.0cd/m^2 that filled the rest of the screen. Since the base of the wide bar was set to mid-grey, this abrupt change was not visible, except at the top and bottom of the bar. The observers used a chin and head rest, and viewed the stimuli binocularly in a dark room.

Each trial began with a 500ms presentation of a central black fixation dot (2 x 2 pixels) on a blank mid-grey screen. This was followed by 300ms presentation of the stimulus, without the fixation point. The stimulus was then replaced by the blank screen with fixation dot. When c_1 was above zero, the task was to indicate the presence or absence of the thin dark (or light) bar in the centre of the broad light (or dark) bar. When c_1 was zero, the task was polarity identification of the small Gaussian. Stimuli from successive trials were separated by at least 1sec.

One set of images consisted of all of the narrow bar contrast levels (c_2) for a single narrow bar width (σ_2) and a single contrast of the wide bar (c_1), but included both peak central and trough central polarities. Each session consisted of 10 repetitions of a whole image set, in a randomised order, and took about 7mins to complete. The data from the first session was discarded as practice. Two observers each completed a further 6 sessions before moving on to the next image set. Image sets were shown in the same randomised order for both observers. There were 12 image sets (4 c_1 levels $\times$ 3 σ_2 levels) in total.

Calculation of image scale

The aim is to obtain the scale of a Gaussian bar whose width appears to be the same as a single bar of a 1 cpd sine-wave. Thus the Gaussian's width at half-height should equal 30mins. A Gaussian of unit height and zero mean is given by

$$\exp\left(\frac{-x^2}{2\sigma^2}\right)$$

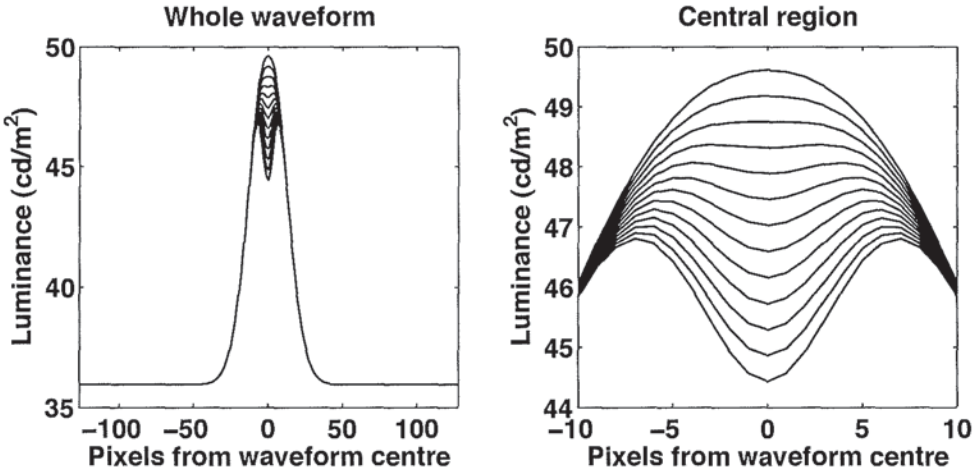

Figure 5.12: An example of a set of peak central stimulus waveforms used in experiment 7. The left plot shows the whole waveforms. The right plot shows their central region. The top waveform is the large Gaussian alone. The waveforms below it are the result of successive increases of contrast in the small inverted Gaussian component.

At half-height this expression becomes

$$\exp\left(\frac{-x^2}{2\sigma^2}\right) = \frac{1}{2}$$

Solving for σ

$$\sigma = \sqrt{\frac{-x^2}{2ln(0.5)}} \tag{5.12}$$

Substituting $x = 15$ (the Gaussian's half-width at half-height) into equation 5.12 produces $\sigma = \pm 12.74$. This is very close to 12, which is the value of σ_1 used in this experiment.

5.2.3 Null point prediction

Following the null point prediction made in the 3 + 3f experiment, this section derives an equivalent expression that defines the required thin bar contrast level that causes the experimental waveform to form a central plateau. As before, the logic is that this level can be used as a prediction of the thin bar contrast that defines the observers' 'null point' between seeing and not-seeing the thin bar.

If we assume that the null point will be where the 2nd derivative is zero at the centre of the waveform, then this will occur when the contrast of the small Gaussian (c_2) is given by:

$$c_2 = c_1 \frac{\sigma_2^2}{\sigma_1^2} \tag{5.13}$$

where c_1 is the contrast of the large Gaussian, σ_1 is its standard deviation, and σ_2 is the standard deviation of the smaller Gaussian.

Proof

The luminance profile of the experimental waveforms was defined by a difference of two unit height Gaussians

$$L(x) = c_1 \exp\left(\frac{-x^2}{2\sigma_1^2}\right) - c_2 \exp\left(\frac{-x^2}{2\sigma_2^2}\right) \tag{5.14}$$

If we consider one Gaussian

$$G = \exp\left(\frac{-x^2}{2\sigma^2}\right)$$

Its 2nd derivative is

$$G'' = \left(\frac{-x^2}{\sigma^4} - \frac{1}{\sigma^2}\right) \exp\left(\frac{-x^2}{2\sigma^2}\right) \tag{5.15}$$

At the centre of the waveform, $x = 0$, so equation 5.15 reduces to

$$G''(0) = -\frac{1}{\sigma^2}$$

So at $x = 0$, the 2nd derivative of equation 5.14 becomes

$$L''(0) = \frac{c_2}{\sigma_2^2} - \frac{c_1}{\sigma_1^2} = 0$$

Rearranging the central and right terms gives equation 5.13. This ends the proof.

5.2.4 Results

The data were analysed in the same way as the f + 3f data in the previous experiment. Psychometric functions were generated for each observer and condition. When the contrast of the wide bar was above zero, the percentage of 'thin bar seen' responses was plotted as a function of thin bar contrast, separately for each stimulus polarity. When the contrast of the wide bar was zero, the ordinate was the percentage of 'light bar' responses; the responses to peak central wide bars were plotted against positive contrast values, and the responses to trough central wide bars were plotted against negative contrast values.

Separate psychometric functions were also plotted in z-units, because linear performance would be immediately apparent as a linear psychometric function. In contrast, linear performance cannot be surmised from simple inspection of a plot of percentage values because that psychometric function still has a sigmoidal shape whether performance is linear or not. Example plots for one pair of conditions for observer NRH are shown in figure 5.13. The percentage psychometric functions were fitted using the psychometric fit program described in section 2.5.1. In order to determine whether it was appropriate to fit with a Gaussian integral, the z-score plots were individually examined by eye for linearity. Also, the correlation coefficient was calculated between each set of z-scores and the contrast of the thin bar. For observer SAW, $r \geq 0.96$ in all conditions. For observer NRH, $r \geq 0.96$, in all conditions except five. Out of these five, four had a wide bar contrast above zero and the data suggest a floor and/or ceiling effect (lapses). One such example is shown in figure 5.13 (right), where responses do not fall below about 10%. Providing lapses are taken into account, the bar detection data do appear to be suitable for a Gaussian fit.

Figure 5.14 shows the z-zcore data for all of the polarity identification conditions for both observers. The lower left plot shows the fifth instance where NRH's data did not have a very high correlation coefficient. The data clearly have a pronounced central kink. Therefore the fit applied to all polarity identification data raised the contrast levels to a (sign-preserving) power before adding the Gaussian.

Each panel in the lower row of figure 5.14 shows that the data for observer NRH fall below the (0,0) point, which indicates a response bias in favour of dark bars. The data for observer SAW show little or no response bias.

The psychometric fit was applied separately to the data of each observer and each condition. An initial run of the program placed no constraints upon the lapse rate for all conditions or the exponent power (m) for the polarity identification conditions (for the bar detection conditions m was always fixed at 1). For observer SAW, this resulted in good fits (judged by eye) and reasonably narrow confidence intervals from the bootstrap (1000 runs). The resulting values of m were 1.11, 1.25 and 1.26 for $\sigma_2 =$ 2, 4 or 6 mins respectively. This parameter (m) represents the exponent of a non-linear transducer, so would be expected to be a constant across the 3 polarity identification conditions. So a further fit was performed with m fixed at the mean of the previous values (1.21) for these 3 conditions. The resulting fits are shown in figure 5.15 with all of SAW's data. For all 21 fits, the lapse rate was between zero and 8%, and the deviance of the data from the fitted curve fell within the 95% confidence limits.

Where the task was bar detection (left 3 columns), the peak central and trough central data generally superimpose in most panels. In those conditions where they do not generally superimpose, the peak central data are shifted slightly to the left of the trough central data, except for the $c_1 = 10\%, \sigma_2 = 2$ condition (top left panel), where the shift is in the opposite direction. Since the shift's direction is not

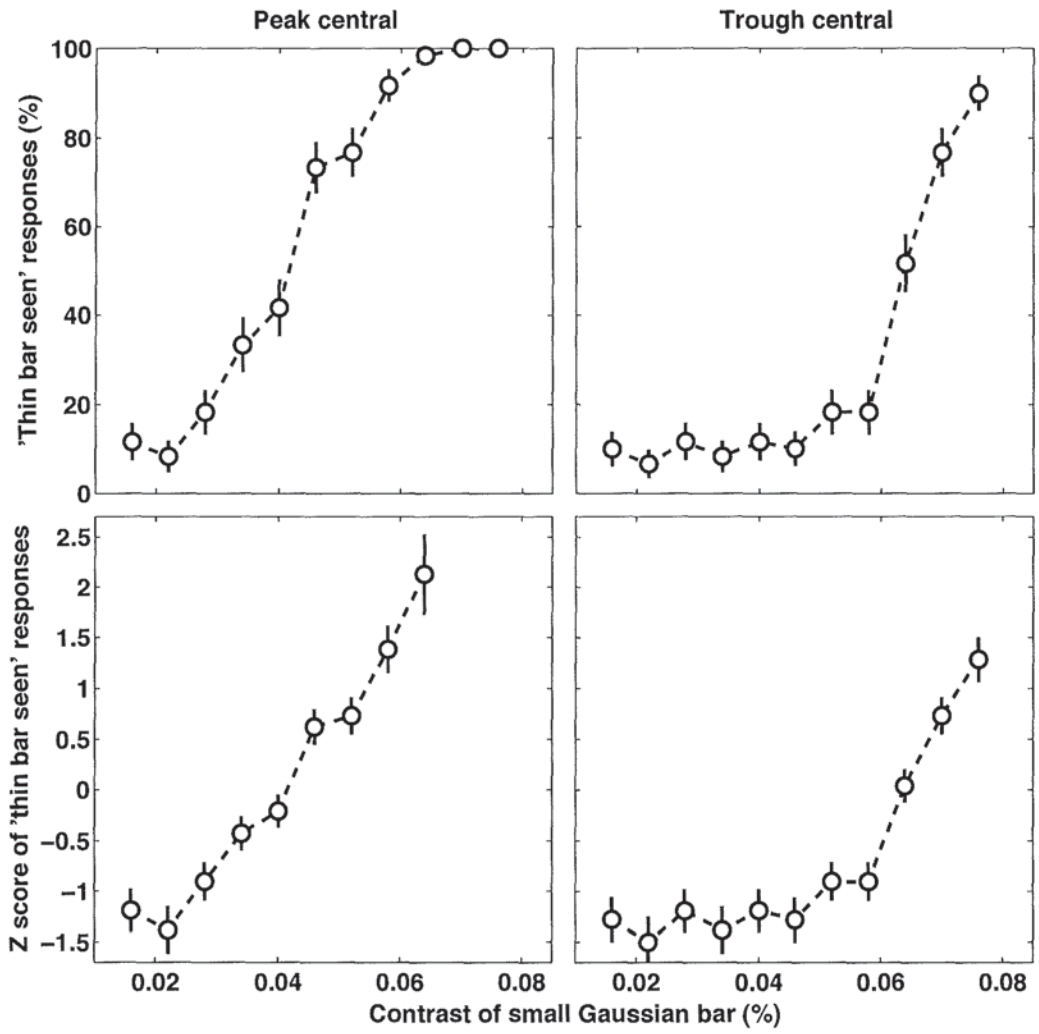

Figure 5.13: Data from one pair of conditions for observer NRH, where $c_1 = 0.2, \sigma_2 = 6$mins. Top row shows the percentage of 'thin bar seen' responses, with error bars are estimated from binomial sampling. Bottom row shows z-scores of the same data, with error-bars estimated from the expected variance. Columns show separate polarity stimuli. Each point represents data from 60 trials.

consistent across all conditions, it does not appear to represent a systematic effect. Where the task was polarity identification (right column) the fitted curves pass through or very close to the (0, 50%) point that is at the centre of each of these panels. This indicates little or no response bias.

The same analysis was performed for NRH's data. The initial psychometric fit was also good, and yielded exponent values (m) of 1.93, 3.29 and 1.69 for $\sigma_2 = 2$, 4 or 6 mins respectively for the polarity identification conditions. A 2nd fit constrained m to be at the mean of the previous values (2.30) for these 3 conditions. The resulting fits are shown in figure 5.16 with his data. There is a clear response bias in favour of thin dark bars, as the fits in the right column all pass below the (0, 50%) point. Moreover, all of the other panels show a leftward shift of the peak central data, which is consistent with a response bias that favours the thin dark bars that were centred on the wide light (peak central) bars. The similarity of slope of the peak central and trough central data in each panel is also consistent

with a simple bias effect.

As before, on each fitted bar detection curve, the contrast level that resulted in a 50% response was assumed to be the null point between seeing and not seeing the thin Gaussian bar (c_{null}). The null point on each polarity discrimination fitted curve was defined as 0% contrast. For both tasks, threshold (c_{thd}) was defined as 1 z-score unit above the null point.

Plots of the c_{null} and c_{thd} contrasts as a function of wide bar contrast are show in figure 5.17 for both observers. In four of the six plots, the c_{null} data are very close to the line of zero 2nd derivative that forms the null-point prediction described above. The close similarity between the data from each observer is particularly striking, which suggests that averaging across stimulus polarity has removed the majority of inter-observer differences. As expected, each c_{thd} datum falls slightly above its corresponding c_{null} datum, but it can now be seen that it is by an amount that is larger when the wide bar contrast is zero than when it is above zero. However, it is difficult to see how this amount changes as the wide bar contrast increases above 10%.

In order to examine this relation more closely, the data for each observer were normalised by the c_{thd} value for polarity identification of the thin bar (c_0), and replotted in figure 5.18. The ordinate shows the c_{null}-c_{thd} increment divided by c_0. The abcissa shows c_{null} divided by the same c_0 value. The

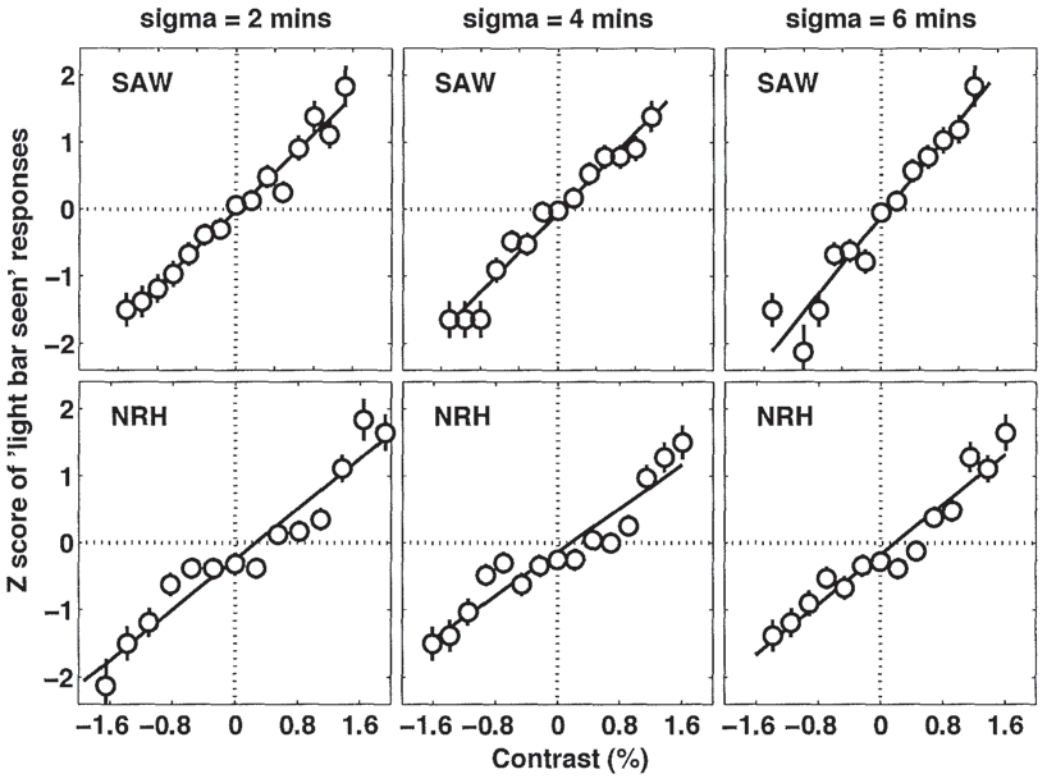

Figure 5.14: Z-scores of 'light bar seen' responses as a function of Gaussian bar contrast. Positive and negative contrast refers to light and dark bars respectively. Rows show data from each observer and columns show separate thin Gaussian bar standard deviations. Each data point represents 60 trials except those for zero contrast, which represent 120 trials. Error bars are derived from the expected variance of each z-score, given the number of trials. Solid lines are the best-fitting (least rms error) linear function.

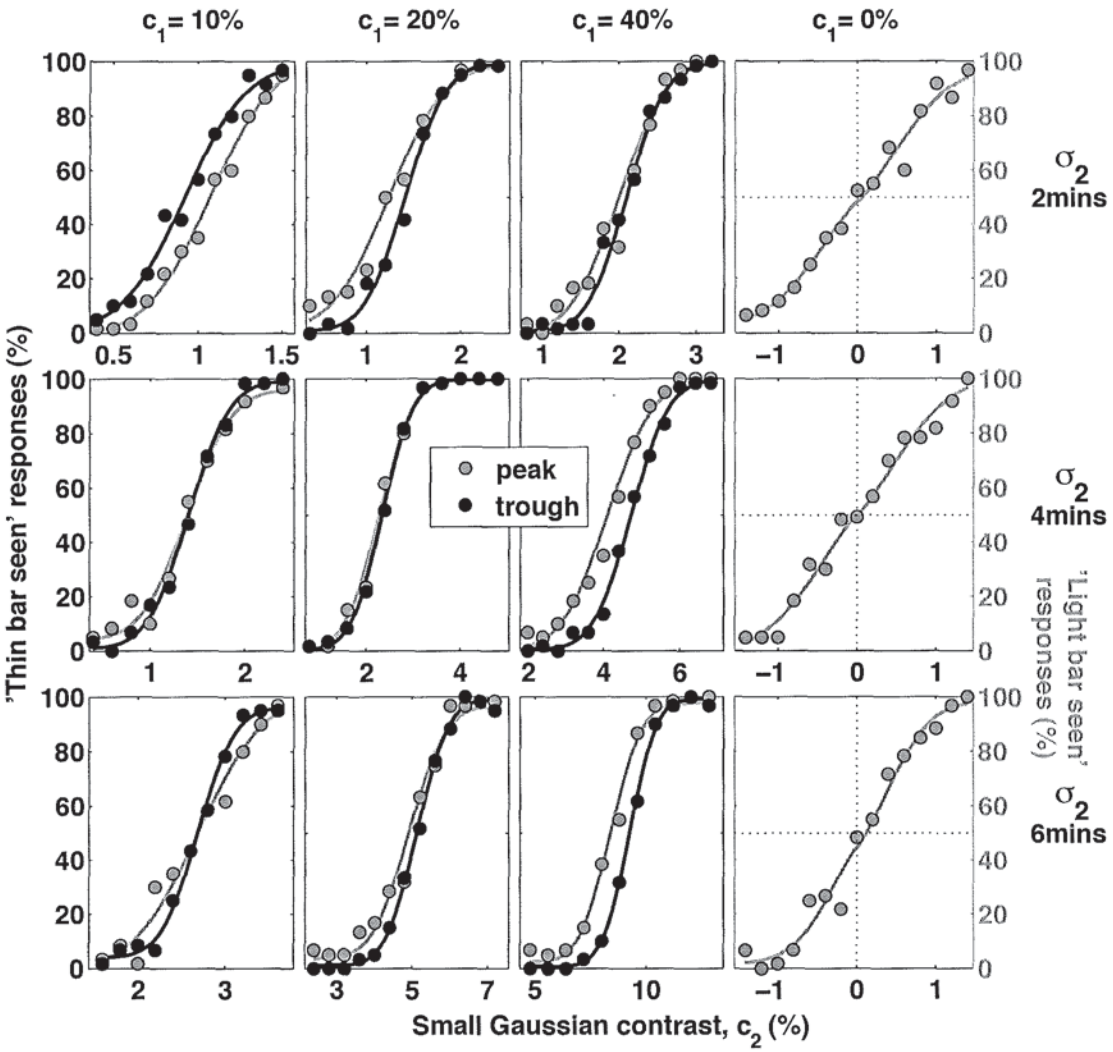

Figure 5.15: Percentage of 'thin bar' or 'light bar' responses (symbols) as a function of narrow Gaussian contrast, with fitted curves (lines) for observer SAW. Each column shows a separate wide Gaussian contrast, and each row shows a different narrow Gaussian spread. The legend refers to the phase of the wide Gaussian bar.

normalised data for both obsrvers fall on or close to a straight line (on these logarithmic axes) of slope 0.6, which demonstrates that a power relationship applies here.

In order to examine whether a local contrast measure might characterise performance, the luminance waveforms were plotted for each bar detection condition, with the contrast of the narrow bar set at c_{null} or c_{thd}. Some of these critical waveforms are shown in figure 5.19 for observer SAW. The waveforms for NRH were broadly similar, so are not shown. The top row of figure 5.19 shows the peak central c_{null} waveforms, which represent the null point between perceiving and not perceiving the narrow bar. It was predicted above that this point would occur when the luminance profile has a central plateau (i.e. zero 2nd derivative). This was based on the idea that if the background was of uniform luminance, and the task was polarity identification of a bar, an unbiased observer would provide a 'light' response

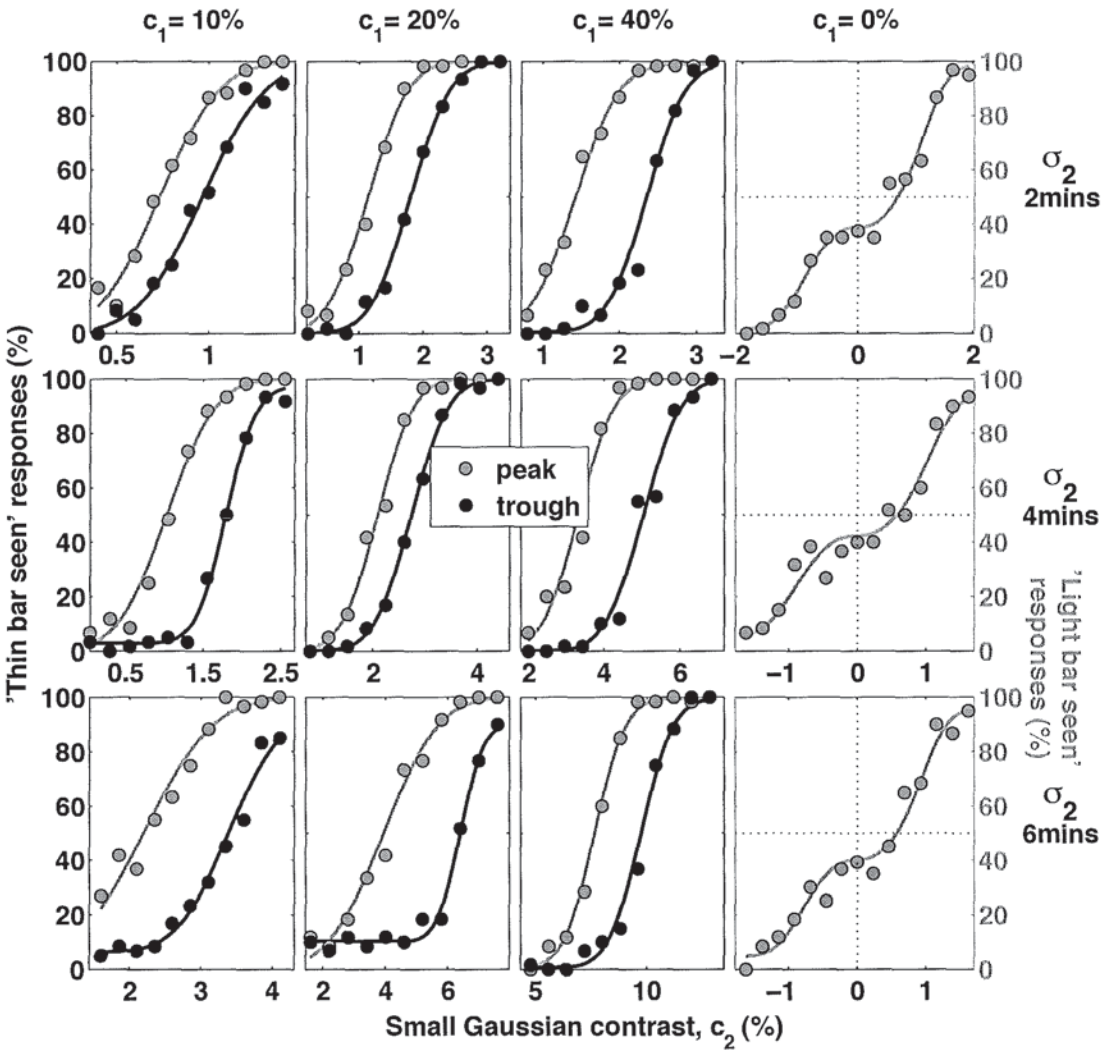

Figure 5.16: Percentage of 'thin bar' or 'light bar' responses (symbols) as a function of narrow Gaussian contrast, with fitted curves (lines) for observer NRH. Each column shows a separate wide Gaussian contrast, and each row shows a different narrow Gaussian spread. The legend refers to the phase of the wide Gaussian bar.

when the bar was a luminance peak and a 'dark' response when the bar was a luminance trough, and the boundary between these 2 responses would be where the bar had zero contrast. Whilst the null point prediction appears to be true for some waveforms, it is clearly not the case when the narrow bar $\sigma = 2$ or the wide bar contrast = 10%, because a central dip is visible.

The middle row of figure 5.19 shows the c_{thd} waveforms, for the same polarity. As they represent the narrow bar perception threshold, we might expect these to all exhibit a central dip. Surprisingly, the waveform for wide bar contrast 40% and narrow bar $\sigma = 6$ appears to have no dip. Moreover, in each panel the amplitude of the dip varies across narrow bar scale. Thus, a simple local contrast mechanism based on luminance peaks and troughs does not appear to be the determining factor in the observers' responses. However, inspection of the luminance profile provides only rough information about the

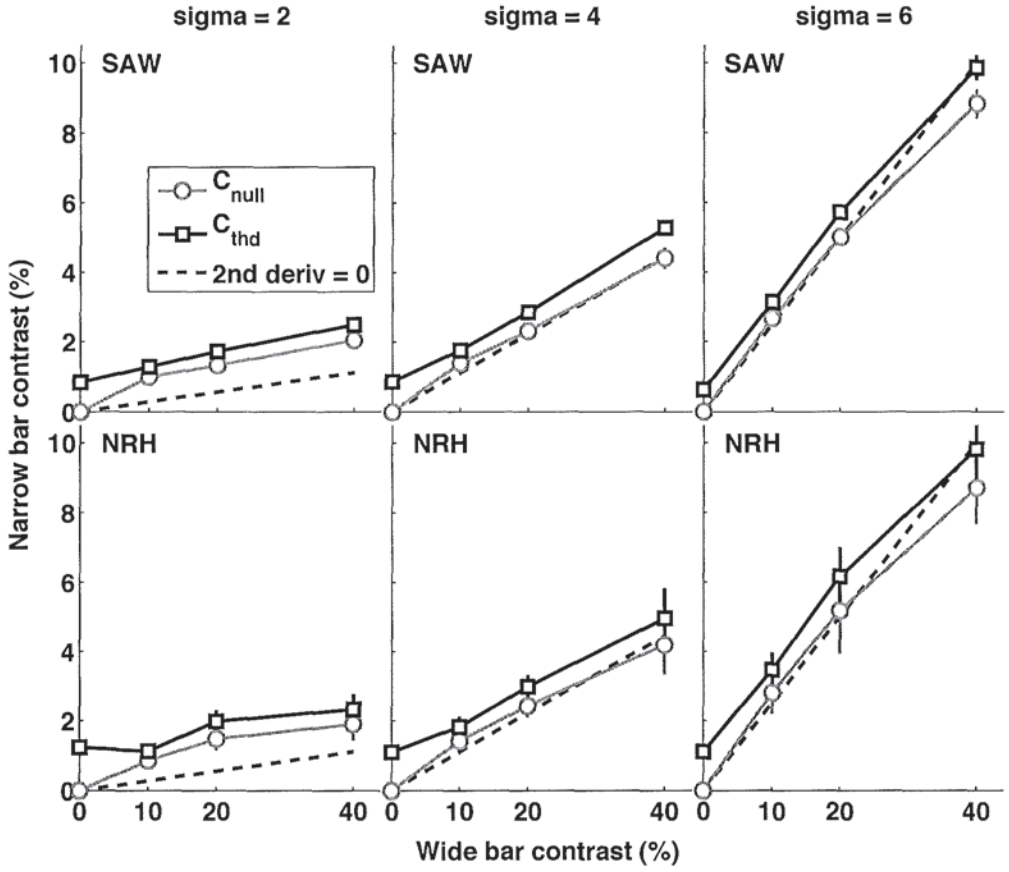

Figure 5.17: Thresholds (squares) and 50% performance contrast levels (circles) for polarity identification of a narrow Gaussian bar on an inverted wide Gaussian background. Dashed lines show null-point (c_{null}) predictions based on zero 2nd derivative values. Column headings refer to the standard deviation of the narrow bar. Each row shows data from each observer. Data are averaged across stimulus polarity, so error bars ($\pm$1se) are calculated from 2 values.

stimulus, because the higher derivatives amplify subtle changes in gradient or curvature. Thus a more rigorous analysis would be to examine higher derivatives directly.

Figure 5.20 shows the amplitude of the 2nd derivative at the centre of the waveforms at 50% response for each observer (rows). In the right column, the waveforms were filtered by the pre-filter (same parameters as previously) before the 2nd derivative was obtained. In the left column, the pre-filter was not applied. If the null point prediction was correct, then the data should all fall at zero 2nd derivative amplitude. While some of the data do fit this prediction, most of the $\sigma = 2$ data do not. A comparison between the left and right columns of figure 5.20 suggests that the addition of the pre-filter does little to improve the fit of the data to the null point prediction. The pre-filter was similarly ineffective at bringing the c_{thd} 2nd derivatives onto a common function (not shown).

The c_{null} waveforms shown in figure 5.19 were blurred by convolution with a Gaussian and then their 2nd derivative was obtained, in order to determine whether the resulting waveforms would all form a central plateau. However a central inflection persisted until the standard deviation of the blurring

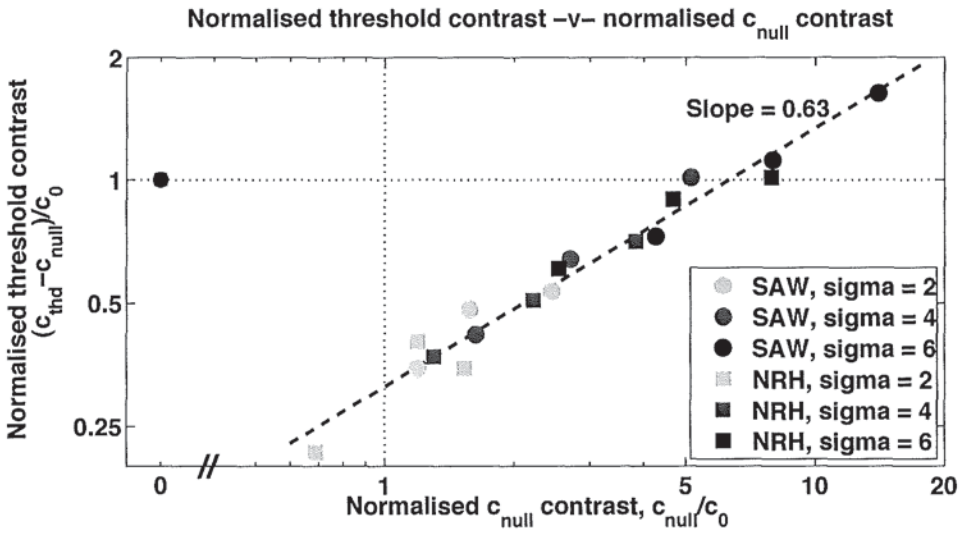

Figure 5.18: Normalised contrast threshold as a function of normalised c_{null} contrast. Circles show data from observer SAW and squares are from NRH. Symbol shade represents the spread of the narrow Gaussian bar. The dashed line is the best linear fit to the data.

Gaussian was 6 mins. This amount of blur might represent the combination of optical blur and the scale of the smallest filter used by the observer ('neural blur'). While this amount of neural blur seems plausible for the narrow bar $\sigma = 6$ images, it seems implausible when the narrow bar $\sigma = 2$ or 4, because the observer would surely employ a finer filter.

In figure 5.17, the c_{thd} data are generally parallel to the 2nd derivative = 0 line. In order to examine this relation more closely, figure 5.21 shows $c_{thd} - c_0$ as a function of the narrow bar contrast levels that provide a zero 2nd derivative (c_{d0}). Most of the data fall on or close to the line of unity slope (not fitted to data), demonstrating a clear linear relationship, which can be expressed as

$$c_{thd} - c_0 = c_{d0} \tag{5.16}$$

Rearranging equation 5.16 produces

$$c_{thd} = c_{d0} + c_0 \tag{5.17}$$

Equation 5.17 shows that the threshold contrast level is the simple sum of the contrast required to bring the waveform to a plateau (c_{d0}) and the contrast required to identify the polarity of the narrow bar on a background of even luminance (c_0).

5.2.5 Discussion

The aim of this experiment was to characterise the perception of a thin Gaussian bar when it was presented on a background of a wider inverted Gaussian bar. The results suggest that the 2nd derivative of the luminance profile plays a major role. In figure 5.17, the c_{null} and c_{thd} data all follow the trend of the 2nd derivative = 0 line; how it increases in slope as sigma increases. This concordance with the 2nd derivative is much clearer here than in the previous experiment (figures 5.6 and 5.7) where the slope of the 2nd derivative = 0 line did not vary. The c_{thd} data are generally parallel to the 2nd

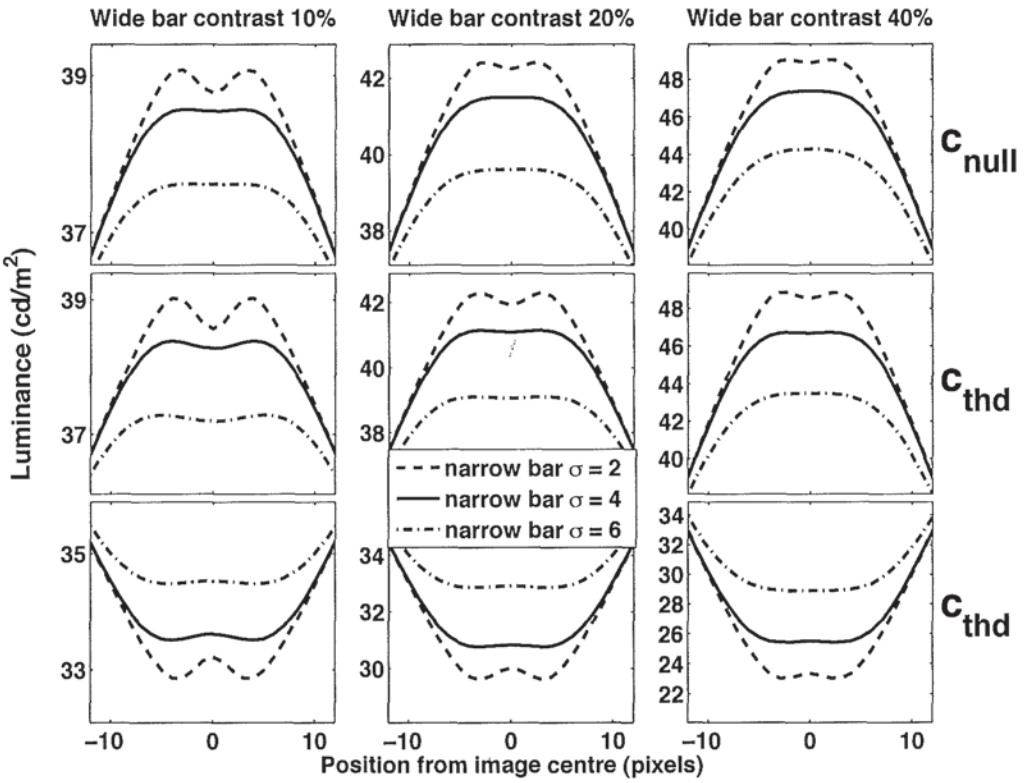

Figure 5.19: Luminance waveform profiles at the 'null point' between perceiving and not perceiving a thin inverted Gaussian bar on a wide Gaussian bar background, for one stimulus polarity (c_{null}, top row), and at the thin bar detection threshold for both polarities (c_{thd}, bottom two rows), for observer SAW.

derivative = 0 line, except for the $c_1 = 40\%, \sigma_2 = 6$ values, which fall on the line. This generally linear relationship is shown clearly in figure 5.21. Thus it can be concluded that the threshold for seeing the thin bar is the sum of the of the contrast required to bring the waveform to a plateau (c_{d0}) and the contrast required to identify the polarity of the narrow bar on a background of even luminance (c_0). This suggests that the wide bar serves only to alter the background luminance, and does not facilitate polarity identification of the narrow bar. An interpretation based on interactions between a wide bar mechanism and a narrow bar mechanism would appear to be unnecessarily complicated here.

However, a different interpretation is suggested by dipper-like function shown in figure 5.18. The majority of the normalised $c_{thd} - c_{null}$ increment values fall below 1, which suggests that the narrow bar polarity identification mechanism is facilitated. But it is not clear how much of the facilitation is due to the wide bar contrast and how much is due to the narrow bar contrast, because the normalised c_{null} term (which is the abscissa of this plot) is influenced by the contrast of both bars. The interpretation of this dipper-like function is not straightforward because it has an atypical abscissa. A dipper function usually plots the contrast discrimination threshold as a function of stimulus contrast, with both factors normalised by the detection threshold. In other words, the abscissa is derived predominantly from the stimulus contrast levels. However, in figure 5.18 the abscissa relates entirely to response levels of contrast. The slope of the best-fitting linear function in figure 5.18 is 0.63, which is the same value as that obtained in the equivalent plot in the previous experiment (figures 5.8 and 5.9). This once again

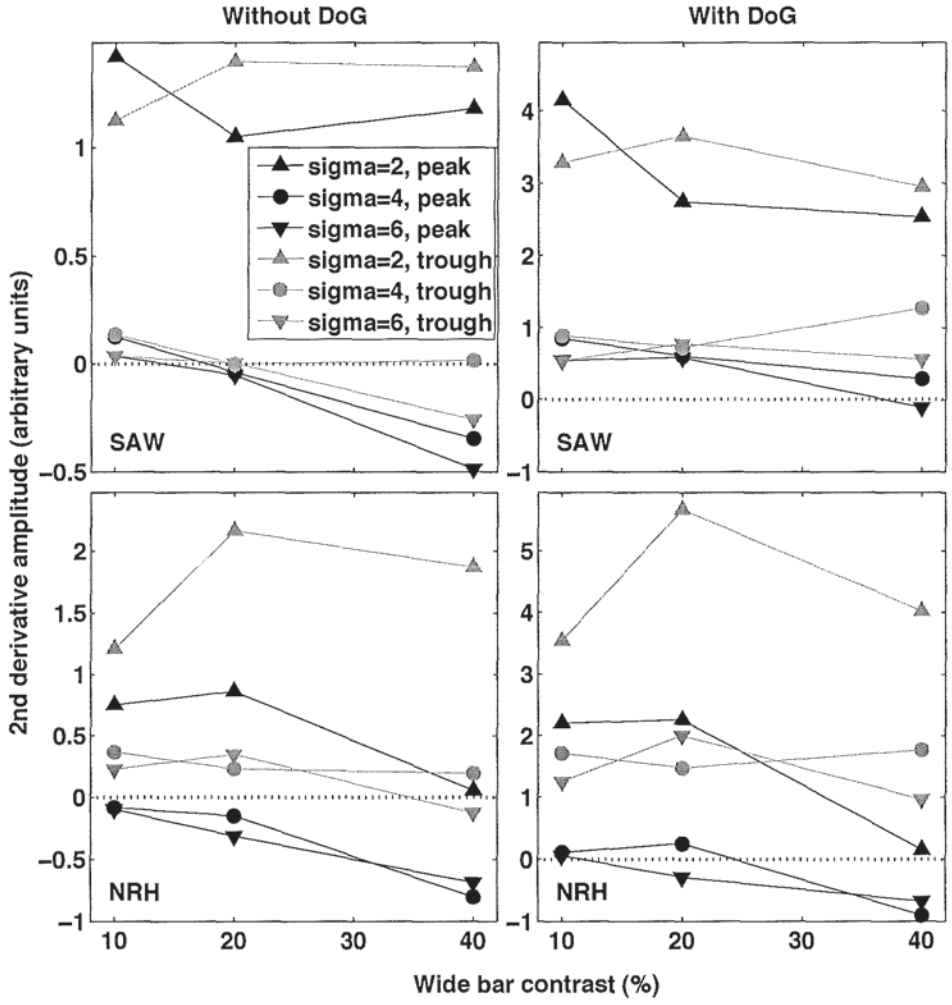

Figure 5.20: Amplitude of the 2nd derivative at the centre of the c_{null} waveforms for each observer (rows), without and with a DoG pre-filter (columns).

shows a lawful relationship, which is now shown to apply across different types of stimuli.

It now appears that the results are inconsistent, because figure 5.21 and equation 5.17 suggest no facilitation, and yet figure 5.18 shows facilitation. This apparent discrepancy can be resolved by consideration of the reference contrast level that c_{thd} is compared against in each case. In equation 5.17, c_{thd} is defined as the contrast required to bring the waveform to a plateau (c_{d0}) added to the contrast required to identify the polarity of the narrow bar on a background of even luminance (c_0). This can be visualised in each panel of figure 5.17 by starting from the dotted line (c_{d0}), and moving upwards by the distance between the leftmost square and circle. The degree to which this manipulation is an accurate description of the threshold data is the same as the degree to which the squares are parallel to the dotted line.

The dipper-like function shown in figure 5.18 applied a different definition of performance, which was $(c_{thd} - c_{null})/c_0$. For a given narrow bar σ and observer, the value of c_0 is constant. Therefore, when considering the data in a given panel of figure 5.17, this measure of performance is governed by the vertical difference between the squares (c_{thd}) and the circles (c_{null}). In most panels, this distance

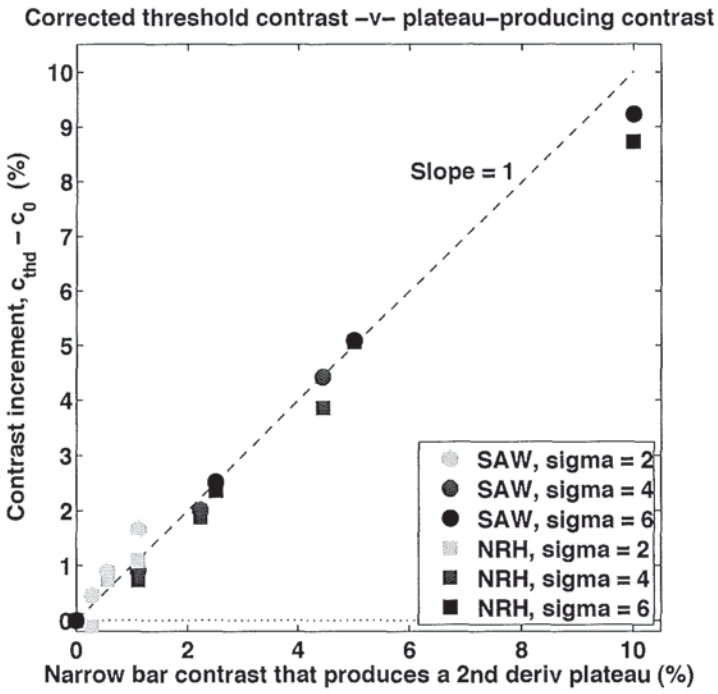

Figure 5.21: Contrast increment for threshold performance as a function of the narrow bar contrast that provides a zero 2nd derivative.

is large at zero wide bar contrast, and is relatively small at higher wide bar contrast levels. This reduction at wide bar contrasts above zero indicates facilitation. This analysis is not in conflict with the lack of facilitation described above, because it is merely using a different reference level to compare threshold performance against. Here the reference level is c_{null}, whereas the above analysis used c_{d0}. Nevertheless, at the time of completing this thesis, further data analysis was underway, which aimed to further reconcile the 'facilitation *vs* no facilitation' issue, and to draw further conclusions about human visual processing.

All of the c_{null} and c_{thd} values were derived from the curves produced by the psychometric fit program. The closeness of the curves to the data, shown in figures 5.15 and 5.16, attests to the program's efficacy, and supports the use of its structure (e.g. non-linear transduction for the polarity identification data, and the inclusion of the lapse parameter). However, the fitted value of *lapse* was not as consistent a might be expected. Each panel of figures 5.15 and 5.16 shows the bar detection data for peak central and trough central stimuli. Since both of these data sets were collected in an interleaved manner within the same session, it would be expected that the value of *lapse* would be the same for both fitted curves in a given panel.

However this was not always the case. In figure 5.16, the $c_1 = 20\%, \sigma_2 = 6$ panel shows data that had a fitted *lapse* of zero for the peak central data, and 20% for the trough central data. For the $c_1 = 10\%, \sigma_2 = 6$ condition, these two values of *lapse* were zero and 12%. For the $c_1 = 10\%, \sigma_2 = 2$ condition, they were zero and 6%. In each of these 3 conditions there are usually no peak central data that fall below the trough data's lower asymptote. Therefore, this pattern of data might be thought to represent a strategy of guessing 'thin bar present' on a proportion of trials when the observer is

unsure. A guess rate parameter is sometimes incorporated when fitting psychometric functions, and the functions can be corrected with Abbott's formula (Finney, 1971). However, in the remaining 6 bar detection conditions, *lapse* was always 2% or less. Therefore this putative strategy was not being consistently applied. For observer SAW (figure 5.15), the peak central *lapse* was always within ±5% of the corresponding trough central *lapse* except in 2 conditions. For the $c_1 = 10\%, \sigma_2 = 4$ condition, the peak central and trough central values of *lapse* were 8% and 2% respectively. For the $c_1 = 10\%, \sigma_2 = 6$ condition, these two values of *lapse* were zero and 8%. In both of these conditions, the data that have a lower *lapse* span the other data, which is inconsistent with the proposed guessing strategy. To conclude, it is not clear why some values of *lapse* were unexpectedly high.

The z-score plots in figure 5.14 show that SAW's polarity identification data are strikingly linear. These data were fitted with a transducer exponent of 1.11 to 1.26. A value of 1 would indicate perfectly linear performance, so these values represent nearly linear performance. This is very surprising because an accelerating non-linearity is a well-established property of low-contrast grating detection (e.g. Nachmias & Kocher (1970); Foley & Legge (1981); Meese (2004)) and the exponent is typically in the range 2 to 3, or even higher. Such a novel finding warrants further investigation. The degree of linearity in low-contrast performance forms the theme of the next chapter.

5.3 Chapter summary

A f + 3f grating was used to explore the conditions under which a thin 3f bar was detected on a clearly visible f background. The hypothesis was that the null point between seeing and not seeing the thin bar would occur at a f contrast that causes a central plateau in the luminance waveform. This expectation was approximately met for one, but not the second observer. Subsequent analysis was unable to neatly characterise the performance of both observers.

In a second experiment, the f and 3f components were replaced by wide and narrow Gaussian bars. Now the data from each observer were strikingly similar (when averaged across polarity), and approximately met the null point hypothesis. Moreover, the threshold for detection of the thin bar was the sum of the contrast for polarity identification of the thin bar (on a blank background) and the contrast required to bring the centre of the luminance waveform to a plateau.

Chapter 6

A linear or non-linear transducer at low contrasts?

6.1 Introduction

A transducer is a function that transforms the visual signal into a new form (such as the observer's internal response). One of the earliest descriptions of a non-linear transducer was made by Nachmias & Sansbury (1974), who used it to interprete the results of their classic contrast discrimination experiment. They used a 2AFC procedure with a 3cpd grating of contrast c in one interval, and a similar grating of contrast $c + \Delta c$ in the other interval. The task for the observer was to indicate which interval contained the higher contrast grating. The value of c remained fixed throughout a block of trials, while Δc was varied with a staircase procedure to find its threshold value. Subsequent blocks used different values of c. When c was zero, the task was effectively detection, so that the resulting value of Δc was the grating detection threshold.

When c was above its detection threshold, the Δc threshold increased with c (although less rapidly that Weber's Law predicts). However, for values of c between zero and its detection threshold, as c increased the Δc threshold actually decreased. In other words, a low contrast grating facilitated the detection of a small contrast increment. A plot of Δc *vs* c has a central dip, as shown in figure 6.1 (left), which is why these plots are sometimes known as 'dipper functions'. When c is not zero, it raises the contrast in both 2AFC intervals, and is therefore sometimes known as a 'pedestal'. Nachmias & Sansbury suggested that an accelerating non-linear transducer could account for the facilitation, due to the following explanation.

The right panel of figure 6.1 shows the hypothetical transducer that maps contrast onto internal response, with a linear response axes. As the stimulus contrast increases from zero, the observer is assumed to detect the stimulus when the internal response increases by a fixed amount (Δr). The contrast has increased by Δc_1. When a low-contrast pedestal is present, the response is driven higher up the transducer function. The gradient is steeper here, and so a smaller contrast increment (Δc_2) is required to obtain the same increase in response (Δr). This accounts for the facilitation. Nachmias & Sansbury obtained estimates of discriminability (d') and found that it was proportional to contrast raised to an exponent of 2.2 to 2.9. Other researchers have reported transducer exponents that fall

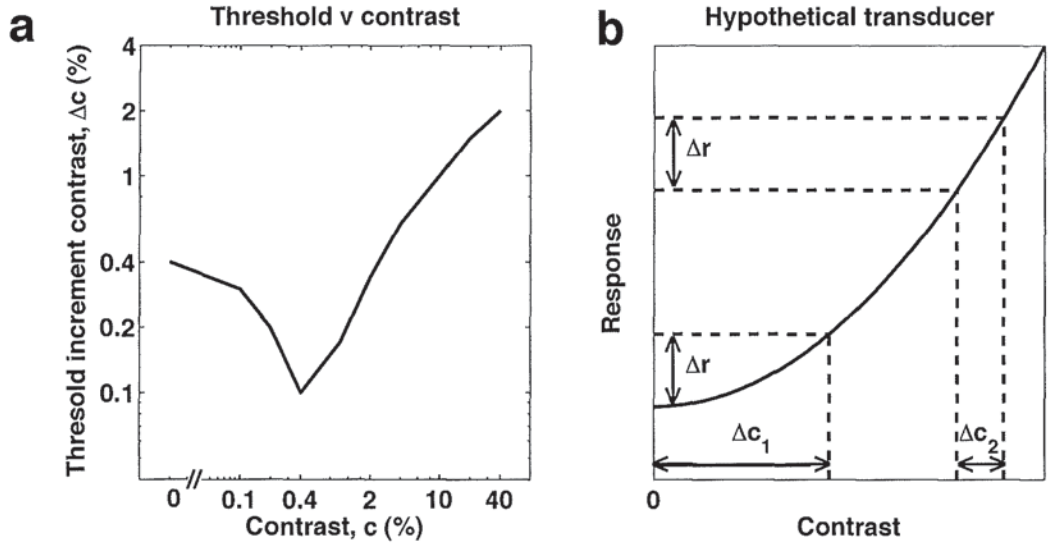

Figure 6.1: a) Contrast discrimination function that shows threshold contrast increment as a function of pedestal contrast. b) Hypothetical transducer that transforms stimulus contrast into internal response. After Nachmias & Sansbury (1974).

within, or slightly above, this range (Foley & Legge, 1981; Kontsevich et al., 2002; Meese, 2004). Table 6.1 lists the value of the exponent reported by some previous researchers. This list is not exhaustive. The values in the grand mean row are derived from all individual values in each study. Nachmias & Kocher (1970) is the only listed study where $p < 1.99$ in a certain condition. This was were where the task was detection of a luminance spot on a totally dark background. So the stimulus contrast was extremely high. In the context of a non-linear transducer for contrast, this would mean that performance was probably not being measured at the low-contrast accelerating region of the transducer, but rather higher up the transducer where it is more linear. In this respect, this condition was therefore more comparable to conventional contrast discrimination than contrast detection. Therefore the data for this condition was removed in the 'corrected grand mean' row of table 6.1. Nachmias & Kocher (1970) also presented the target spots on a background of even luminance, in other conditions, with the result $p \geq 2.02$.

There are no reports in the literature of linear performance (exponent = 1) at these low contrast levels. Therefore, it is surprising that SAW obtained a near-linear exponent value of 1.2 in experiment 7. But that experiment used polarity identification rather than the more typical contrast detection. The next experiment will attempt to replicate this atypical exponent value in order to determine whether it might be an artifact of experimental error.

6.2 Experiment 8 - Bar polarity identification with sinusoidal grating stimuli and Gaussian bar stimuli

The polarity identification data from the previous two experiments yielded conflicting results. In the f + 3f experiment (experiment 6), when the f component had zero contrast, the task was polarity identification of the 3f stimulus. These data were fitted with a power function of exponent 2 for both

Source	mean p	range of p	stimulus	method	observers	practice
Tanner & Swets (1954)	2.66	2.09 - 3.22[†]	1° disc	2AFC	3	not stated
Leshowitz et al. (1968)	2	2	1° disc	2AFC	2	not stated
Nachmias & Kocher (1970)	2.15	1.10 - 3.39[†]	1° disc or square	rating scale	12	4+ sessions
Nachmias & Sansbury (1974)	2.55	2.11 - 3.04	9cpd grating	2AFC	2	not stated
Stromeyer & Klein (1974)	4	4	9cpd grating	rating scale	1	a few trials
Foley & Legge (1981)	2.62	2.11 - 3.04	0.5-8cpd gratings	2AFC	2	not stated
Legge (1984)	2.32	1.99 - 2.63	0.5cpd grating	2AFC	6	not stated
Solomon (2007)	3.3	3.3	Gabor	2AFC	1	not stated
Meese & Summers (2009)	2.73	2.0 - 3.2	checkered 2.5cpd grating	2AFC	3	well-practiced
Grand mean $\pm$ sd	2.41 ± 0.64	1.10 - 4				
Corrected Grand mean $\pm$ sd	2.59 ± 0.5	1.99 - 4				

Table 6.1: d' exponents (p) reported in the detection literature, with experimental conditions. Grand mean is calculated from the individual values of p in every study. Corrected grand mean row excludes very high contrast data from Nachmias & Kocher (1970) (see text for rationale). [†]values extracted from plots.

observers. In the Gaussian bar experiment (experiment 7), when the wide Gaussian bar had zero contrast, the task was polarity identification of the narrow Gaussian bar. These data were fitted with an exponent of 1.2 for SAW and 2.3 for NRH. If both experiments are truly measuring the transducer's exponent in the observers' visual system, then we should expect the same exponent value for any stimulus, for a given observer. This is clearly not the case for observer SAW. It is impossible to draw conclusions about the 2nd observer in this respect, as the same one did not perform both experiments.

SAW's different performance between the previous two experiments may be due to methodological differences. Experiments 6 and 7 used two different viewing distances and image sizes, and were run several weeks apart. The increase in linearity observed in the later experiment might be due to greater practice with the polarity identification task with these bar stimuli. If practice causes an improved performance particularly at low contrast levels, then it might linearise the transducer.

SAW's apparent transducer discrepancy (between experiments 6 and 7) will now be examined by using both bar and grating images within a single experiment. Previous contrast discrimination experiments typically used feedback, which improves the effect of practice, so that performance more rapidly approaches an asymptote (Lukaszewski & Elliot, 1962). Feedback may also be expected to stabilise the observers' criterion at a level equivalent to a zero contrast stimulus, rather than allowing their bias free rein. Yet no feedback was used in the previous two experiments. Feedback will be given now.

6.2.1 Method

The method was generally the same as the parts of the previous two experiments where the stimuli were composed of a single waveform. Stimuli were either a vertical Gaussian bar (σ = 12mins) or 3.56 periods of a vertical 1cpd cosine grating, and were of two polarities ('dark' or 'light', determined by their central luminance). All images subtended 3.56 degrees at the viewing distance of 148.1cm. Each image was presented for 300ms at various contrast levels (including zero) and the task was to indicate its polarity (as 'dark' or 'light') at the image centre. A method of constant stimuli with a single interval procedure was used to obtain psychometric functions of percentage of 'light bar' responses against contrast. Grating and Gaussian bar sessions were randomly interleaved. The same two observers were used throughout the experiment and both were very experienced psychophysically. Additionally, DHB had extensive prior exposure to detection and contrast discrimination of 1cpd gratings, but not bar stimuli. SAW had a few hours prior exposure to both grating and bar stimuli. Feedback was given of the correctness of response. All other methodological details were as experiment 7.

6.2.2 Results

The percentage of 'light' responses were calculated for each condition, and are plotted as a function of contrast in figure 6.2. Positive contrast denotes a light central bar, and negative contrast denotes a dark central bar. Separate rows are for each observer and separate columns are for each image type. Error bars ($\pm 1sd$) are approximated from binomial sampling.

The curves show the best-fitting (least rms error) values that were generated by the psychometric fitting program described in section 2.5.1, and which also performed a bootstrap, using 1000 runs. The program employed a power function transducer followed by additive Gaussian noise to generate the

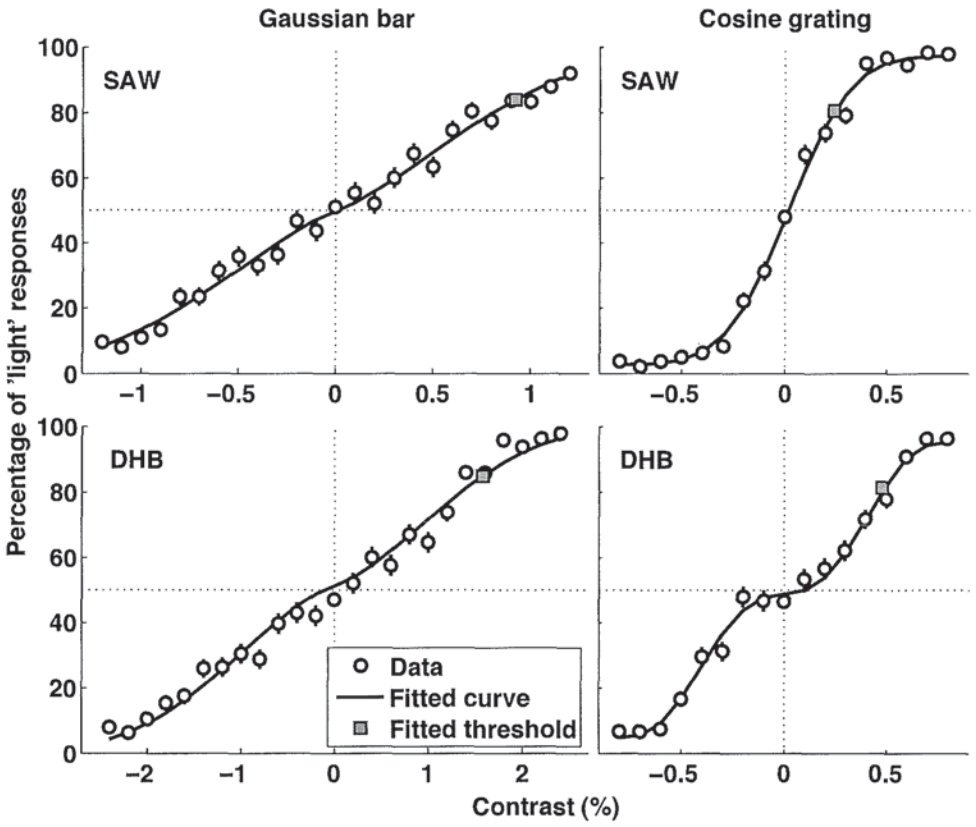

Figure 6.2: Percentage of 'light' responses as a function of contrast, with fitted curve and threshold. Positive contrast denotes a light central bar, and negative contrast denotes a dark central bar. Each datum is from 240 trials. Error bars ($\pm 1sd$) are approximated from binomial sampling.

Stimulus	Observer	*thd*	c_{50}	*lapse*	p	p 95% confidence limits
bar	SAW	0.92	0.02	0	1.23	1.09-1.42
	DHB	1.58	-0.12	0.12	1.34	1.20-1.60
grating	SAW	0.24	0.02	0.05	0.99	0.82-1.22
	DHB	0.48	0.10	0.09	2.21	1.75-2.73

Table 6.2: Parameter values for the fitted curves of figure 6.2.

fitted curve. All four parameters (c_{50}, p, σ and *lapse*) were allowed to be free independently for each plot. The fitted curves are shown as lines in figure 6.2, and the resulting parameter values are shown in table 6.2, where σ has been replaced by the threshold value (*thd*) because it was considered to be more meaningful than σ. The values for *thd* were calculated by obtaining the point on each fitted curve that was 1 d' unit above the response to the 0% contrast stimulus.

All of the values of *thd* and m and the deviance of the data from the fitted curve fell within the 95% confidence limits of the bootstrap, with one exception. The bar data for DHB fell outside the expected 95% deviance limits, which means that the data depart from the fitted curve by an amount greater than that expected by the random effect of binomial sampling. So another, currently unidentified

factor must be having an influence. From inspection of these data (figure 6.2, lower left), it appears that the residuals form no systematic pattern. This suggests that the unidentified factor has only a random influence, and will therefore not be discussed further. The high deviance means that we must exercise caution when drawing conclusions from this data set. Nevertheless, DHB's data show non-linear performance for polarity identification of the grating ($p = 2.21$) and suggest nearly linear performance with the Gaussian bar ($p = 1.34$). SAW shows linear (or nearly linear) performance for both bar and grating stimuli ($p = 0.99$ and 1.23). The 95% confidence limits for p are all about ± 0.2, except for DHB's grating data, where the limits were about ± 0.5.

6.2.3 Conclusions

The relatively wide p confidence limits for DHB's grating data are likely to arise from the dark-light response asymmetry. In figure 6.2 (lower right panel) the positive contrast data appear relatively linear. However, as contrast shifts from zero to negative values, responses initially stay close to the 50% level before falling further, which indicates non-linear performance. Thus, as the value of p decreases from the value used to generated the line (2.21), the resulting curve would form a closer fit to the positive low contrast data, but a worse fit to the negative low contrast data (and vice-versa for an increase in p). This behaviour suggests that the fitting program should incorporate separate fits to data of each polarity. However, this factor was not included in order to limit the number of free parameters. Furthermore, only 1 of the 4 data sets show the dark-light asymmetry, and it is principally caused by only 4 data points (the 2 each side of zero contrast).

The linearity of SAW's data replicates the surprising performance that was observed in the previous experiment (where Gaussian bar stimuli were also used) and extends it to the broader scale bars used here. It is particularly interesting that this linearity is now seen with 1cpd gratings, as this was not seen with the 1cpd or 1/3cpd gratings in experiment 6. The 1cpd grating stimuli used here were identical to those of experiment 6, and very similar methods were used to collect data. So it seems likely that the change in performance is due to a practice effect. Observer SAW had not used sinusoidal stimuli prior to experiment 6, except during its pilot, so was relatively unpracticed then. But in the present experiment, as the same gratings were used here, their familiarity was much greater. To make an accelerating curve linear, we can either increase low values, decrease the high values, or some combination of both. But, practice would be expected to only increase performance, and not decrease it at all.

In an attempt to determine whether SAW's improved grating performance could be due to practice, figure 6.3 shows the z-scores replotted from experiment 6 with SAW's grating data from the present experiment. The later experiment shows a clear improvement in performance, consistent with greater practice. However, an alternative explanation could be posited, based on the differences in feedback between the two experiments. Experiment 8 had feedback after every trial, whereas experiment 6 had no feedback. The lack of feedback could have caused the observer's criterion to be less stable, thus reducing performance. In order to test this hypothesis, a control experiment was carried out.

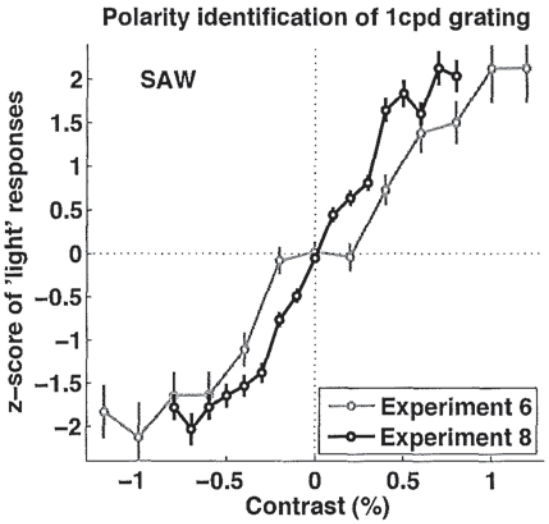

Figure 6.3: Z-scores of 'light' responses as a function of stimulus contrast for polarity identification of a 1cpd cosine grating. Error bars ($\pm 1sd$) are estimated from the expected variance, and are shorter in the experiment 8 data due to the greater number of trials.

Stimulus	Bar		Grating	
	exponent, p	95% confidence limits	exponent, p	95% confidence limits
With feedback	1.23	1.09 - 1.42	0.99	0.82 - 1.22
Without feedback	1.14	0.98 - 1.28	1.21	1.01 - 1.47

Table 6.3: Fitted exponent values for the data shown in figure 6.4.

6.2.4 Control experiment without feedback

Observer SAW repeated experiment 8, but now without feedback. If the resulting data are comparable to those of experiment 6, then that would support the hypothesis that the difference between the data sets shown in figure 6.3 is due to feedback. However, if the data are comparable to those of experiment 8, then that would suggest that feedback was not an important factor.

Figure 6.4 shows the new percentage data, in addition to the old data with feedback. There is little difference between the 2 data sets, which is indicated by an insignificant Pearson X^2 statistic (bar: p=0.99, grating: p=0.55). This statistic was calculated after correction for the false alarm rate, which has the effect of normalising the plots about the zero contrast data points. The psychometric fit program was applied to the data without feedback. The resulting exponent values, with their 95% confidence limits are shown in table 6.3, alongside the values with feedback. All 4 sets of confidence limits overlap in the range 1.09 - 1.22. There appears to be little or no effect of feedback or type of stimulus.

6.2.5 Discussion

The polarity identification data of both bar and grating stimuli suggest a transducer exponent close to 1 for observer SAW, which is surprisingly linear performance, that has not been observed in any

previous research at these low contrast levels. Given that the putative transducer is relatively linear, then SAWs previous non-linear performance with the same grating stimuli in experiment 6 ($p = 2$), cannot be due a nonlinear transducer. It may possibly be attributed to a much lower prior exposure to this image set, which caused stimulus uncertainty. The theme of stimulus uncertainty will be addressed in experiment 10.

It can be hypothesised that linear performance is due to extensive practice, and that nonlinear performance reflects a lack of extensive practice. If this is so, then we might expect to have observed nonlinear performance with the bars in experiment 8, as bars of this scale have not been used in any previous experiments. Yet the transducer exponent obtained with these bars (1.23) is close to linear. However, this nearly linear performance may reflect the benefit of practice which has transferred across scale from the use of narrower Gaussian bars in experiment 7. If the hypothesis were correct, we might expect to have observed non-linear performance in the bar data of experiment 7, as both observers had no prior exposure to Gaussian bar stimuli of any scale. Yet SAW's data were fitted with an exponent value of 1.1 to 1.25, which is nearly linear. This appears to contradict the hypothesis that nonlinear performance is due to a lack of practice. However, SAW did have prior psychophysical practice with Mach Edge stimuli (see chapter 3). Although those stimuli were not Gaussian bars, they did resemble blurred bars. Perhaps the experience with those stimuli helped to reduce stimulus uncertainty of the Gaussian bars in experiment 7. Further support for the hypothesis that nonlinear performance reflects lack of practice is provided by the 2nd observers' polarity identification data in experiments 6, 7 and 8 (grating only for experiment 8). These data were fitted with mean transducer exponents of 2.3, 3.9 and 2.2 respectively. This nonlinearity is expected, because none of the observers had prior exposure to the stimuli. The one exception to this pattern is the exponent value of 1.3 that fitted DHB's bar data in experiment 8. This observer was not experienced with bar stimuli, so it is not clear why this exponent value is so low.

To conclude, while the data suggest that non-linear performance might be an artifact of insufficient

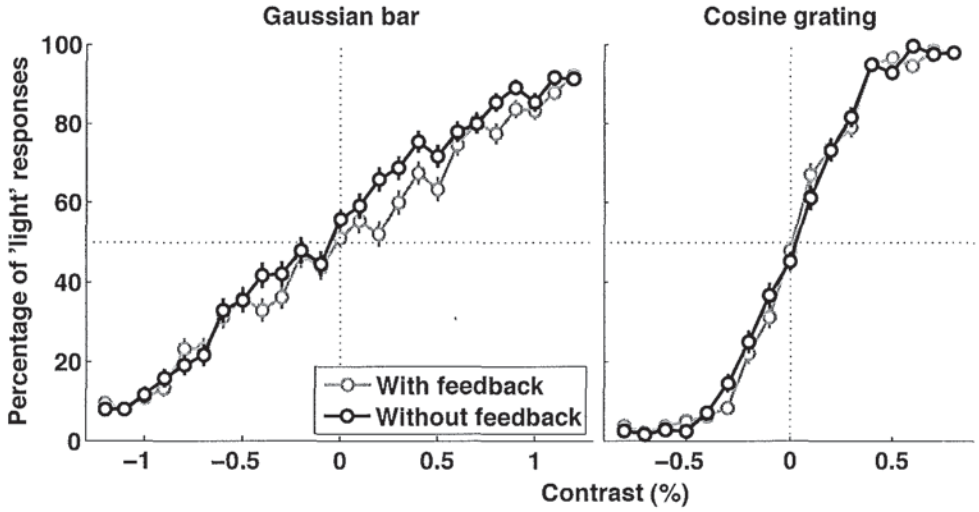

Figure 6.4: Percentage of 'light' responses as a function of contrast for Gaussian bar and cosine grating stimuli, with and without feedback, for observer SAW.

practice, the existence of linearity without extensive prior practice (SAW, experiment 7) suggests that some transfer of the practice effect can occur, or that this account is incomplete (DHB's bar data, experiment 8).

The time span required to linearise performance appears to be long. Due to program development and experimental pilots, SAW had over 8 hours of practice in each of experiments 6 and 7. Furthermore, since experiment 8 used stimuli that were essentially the same as those of experiment 6 and 7, SAW's prior practice here was over 22 hours. Whereas, mainly nonlinear performance followed approximately 20mins of practice for the 2nd observer in each of these 3 experiments.

In the psychophysical literature, there appears to be no universally-accepted criterion that specifies how much practice is sufficient. Some researchers have measured performance over many repetitions of an experimental task and observed an improvement over time. Then, when an improvement is no longer distinguishable from the data's noise, the observer is considered to have had enough practice. However, many researchers do not specify whether this, or any other, method was applied, or even the amount of prior practice (e.g. Foley & Legge (1981); Henning & Wichmann (2007); Legge & Foley (1980); Sachs et al. (1971); Solomon et al. (1997)). Notable exceptions to this trend include Badcock (1984b), who used at least 20 hours of practice, and Lukaszewski & Elliot (1962), who reported that even after 7 hours of practice in an auditory detection task, performance improved further when a monitory reward was offered.

Although the discrepancy between the linear performance observed here and the nonlinear performance in the literature could be due to an effect of extensive practice, there are methodological differences that might also have had an impact. Previous research typically used a 2AFC contrast discrimination task, whereas the relevant data from experiments 6, 7 and 8 were obtained with a single interval polarity identification method. In order to examine the effect of method, the next experiment uses a 2 x 2AFC design to examine discrimination and detection, with both bar and grating stimuli.

6.3 Experiment 9 - Detection and discrimination of a Gaussian bar and cosine grating

6.3.1 Introduction

The relatively linear nature of SAW's contrast transducer was unexpected because the current literature suggests an accelerating transducer of exponent between 2 and 3, at these low contrast levels (e.g. Foley & Legge, 1981). However, there are a number of methodological differences between the Foley & Legge study and experiments 6 to 8 above. Foley & Legge applied a 2AFC procedure with grating stimuli, and the tasks were contrast discrimination and detection. The above experiments applied a single-interval procedure with grating and bar stimuli, and the task was polarity identification. Furthermore, in the above experiments SAW had extensive practice and the stimuli were not pre-cued, whereas Foley & Legge do not mention practice at all and they presented a 100ms glimpse of the stimulus, before every trial, at 2.4× threshold (0.7% and 1.1% for the 0.5cpd and 2cpd gratings respectively). The aim of the glimpse was to reduce stimulus uncertainty. It might be thought that it would cause an after-image, which would then exert a masking effect on the next stimulus. However, Georgeson & Turner (1985) found that a 100ms presentation of a 0.5cpd grating at 2% contrast caused no measurable after-image.

Nevertheless, the glimpse may have increased temporal uncertainty when the target stimulus was in the first interval.

The present experiment aims to address some of these differences by using a 2 x 2AFC method that incorporates both detection and polarity identification. In order to maximise the chance of observing linear performance, the experiments 6 to 8 grating and bar stimuli will be reused here, and observers will be selected who have extensive experience with them. The purpose of this experiment is to determine whether the linear performance is repeatable, and to examine in more detail the conditions under which it arises.

6.3.2 Method

The grating stimuli and Gaussian bar stimuli used in experiment 8 were reused here. Stimuli were either a vertical Gaussian bar (σ = 12mins) or 3.56 periods of a vertical 1cpd cosine grating, and were of two polarities: 'dark' or 'light', determined by their central luminance. All images subtended 3.56 degrees at the viewing distance of 148.1cm and were surrounded by a full screen of mid grey. A method of constant stimuli with a 2 x 2AFC procedure was used, whereby the stimulus could appear in one of two intervals (probability = 0.5) and could be of either polarity (probability = 0.5). Each interval was of 300ms duration and was indicated by a simultaneous tone. The interval that did not contain the target stimulus contained a full screen at mid-grey luminance.

The intervals were separated by 600ms. Each image was shown at various (above zero) contrast levels and the task was to indicate the interval in which it was displayed, and then indicate its polarity. Immediate auditory feedback was given of the correctness of each response. A black central fixation spot (2 $\times$ 2 pixels) was continuously displayed, except during the two intervals of each trial.

A single run consisted of 10 repetitions of every condition for one image type (bar or grating) and took about 9mins to complete. A block of consisted of 6 runs, and was usually completed in one session, with a short break after every 2 or 3 runs. There were 2 observers, who were both experienced with these images and with psychophysical methods. They completed 4 blocks of each image type, in an ABABBABA sequence, where A represents the bar stimuli and B represents the grating stimuli. The whole experiment took about 8 hours per observer.

6.3.3 Results

The data for each observer and image type were divided into 'polarity' and 'interval' responses for each condition. The proportions of 'light bar' responses were converted to z-scores by using the inverse normal integral. These polarity data are shown in figure 6.5. Negative contrast values indicate dark polarity stimuli and positive contrast values indicate light polarity stimuli. The use of z-scores with this linear progression of contrast levels means that linear performance is indicated by data that form a straight line, with no central kinks. Such a pattern appears to be present in SAW's grating data, but not in the other three data sets. The data in all 4 panels appear to pass close to or through the origin, which shows a lack of polarity bias in the responses.

The z-scores of interval responses are shown in figure 6.6, plotted as a function of absolute contrast. The left half of each plot represents stimuli that were presented in interval 1, and the right half represents stimuli that were presented in interval 2. SAW's grating data are the most linear. His data for both

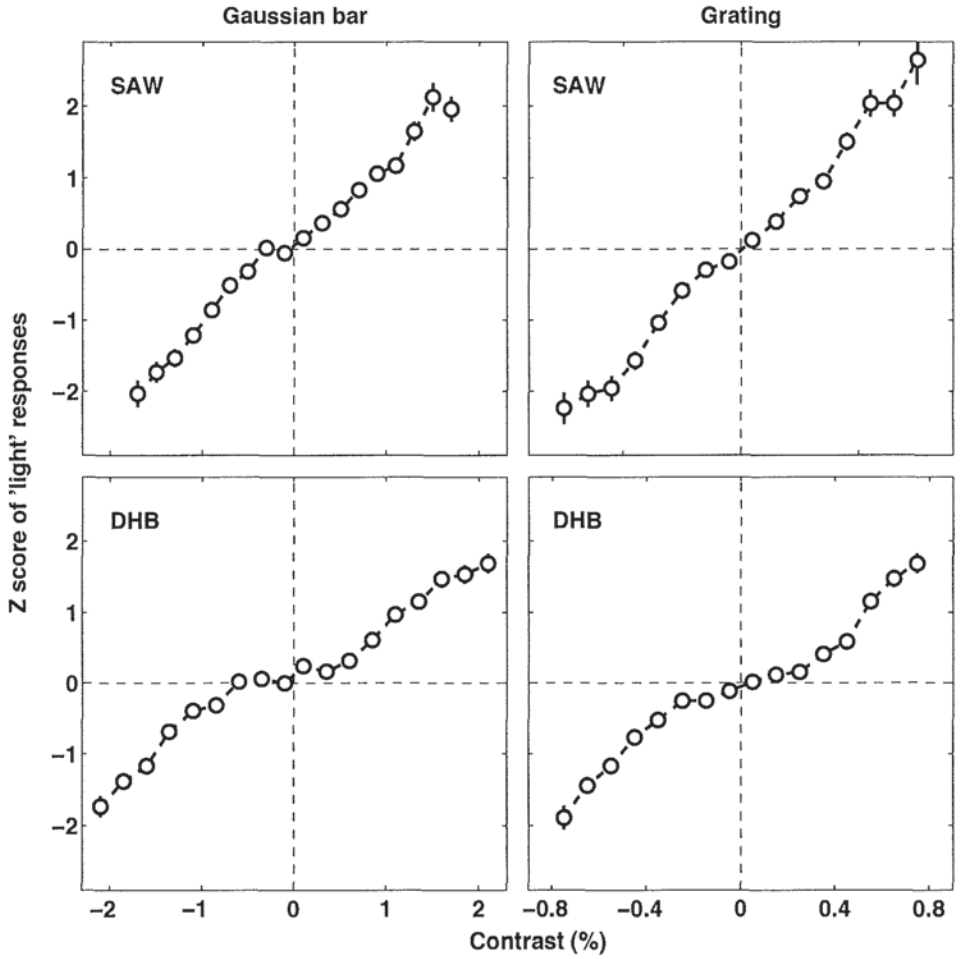

Figure 6.5: Polarity data. Z-scores of 'light' responses as a function of the contrast of bar (left column) or grating (right column) images. Each observer is shown on a separate row. Negative contrast values indicate dark polarity stimuli and positive contrast values indicate light polarity stimuli.

image types pass below the origin, which shows a small interval 1 bias. DHB's data pass slightly above the origin, which suggests a small interval 2 bias.

A Monte Carlo simulation, described in the next section, was used to generate z-scores predictions that formed the best fit (least root sum squared, inverse variance weighted) to the experimental data. The predictions were generated separately for each observer and image type, but were fitted to both interval and polarity data simultaneously. The simulation, however, assumes no bias, so the bias evident in the interval data (and to a lesser extent in the polarity data) was first factored out. The method used to achieve that will be described below, after a brief description of 2AFC criterion bias within the signal detection theory framework.

In signal detection theory, response bias is represented by a non-optimal position of the observer's criterion in decision space. The position that maximises the proportion of correct responses is at the mid-point between the means of the two distribution of responses that the observer is required to discriminate between (often termed 'signal' and 'non-signal' distributions). When the criterion is shifted away from its optimal position, performance is impaired, even though the observer may be just

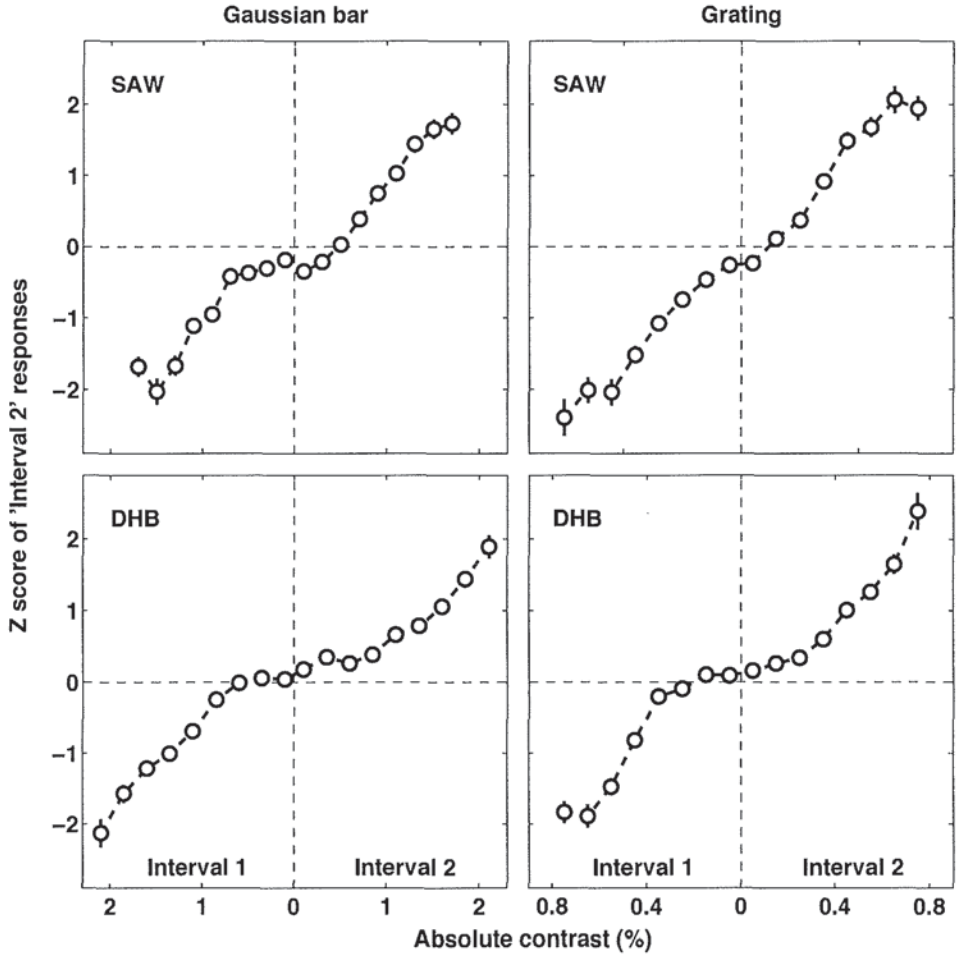

Figure 6.6: Interval data. Z-scores of 'interval 2' responses as a function of the contrast of bar (left column) or grating (right column) images. Each observer is shown on a separate row. The left half of each plot represents stimuli that were presented in interval 1, and the right half represents stimuli that were presented in interval 2.

as sensitive as before. The classic method to remove the effect of bias is to compute discriminability (d'), which is given by

$$d' = z[(p(H)] - z[p(FA)] \tag{6.1}$$

where $p(H)$ is the proportion of hits and $p(FA)$ is the proportion of false alarms. When the task is 2AFC, a 'hit' is defined as an 'interval 1' response when the stimulus was in interval 1, and a 'false alarm' is defined as an 'interval 1' response when the stimulus was in interval 2[1]. These definitions appear to be applicable to the present 2 x 2AFC task.

For the interval data, the response bias was removed as described above. For the polarity data, a hit was defined as an 'light' response when the stimulus was light, and a false alarm was defined as an 'light' response when the stimulus was dark. The removal of bias was therefore achieved by subtracting each dark contrast level response from its corresponding light contrast level response. This can be visualised as a 180 degree rotation of the dark data (left hand side) of figure 6.5 about $(0, 0)$, so

[1]The designations of 'hit' and 'false alarm' are somewhat arbitrary in 2AFC, or 2 x 2AFC, because they could equivalently be replaced by 'correct rejection' and 'miss'.

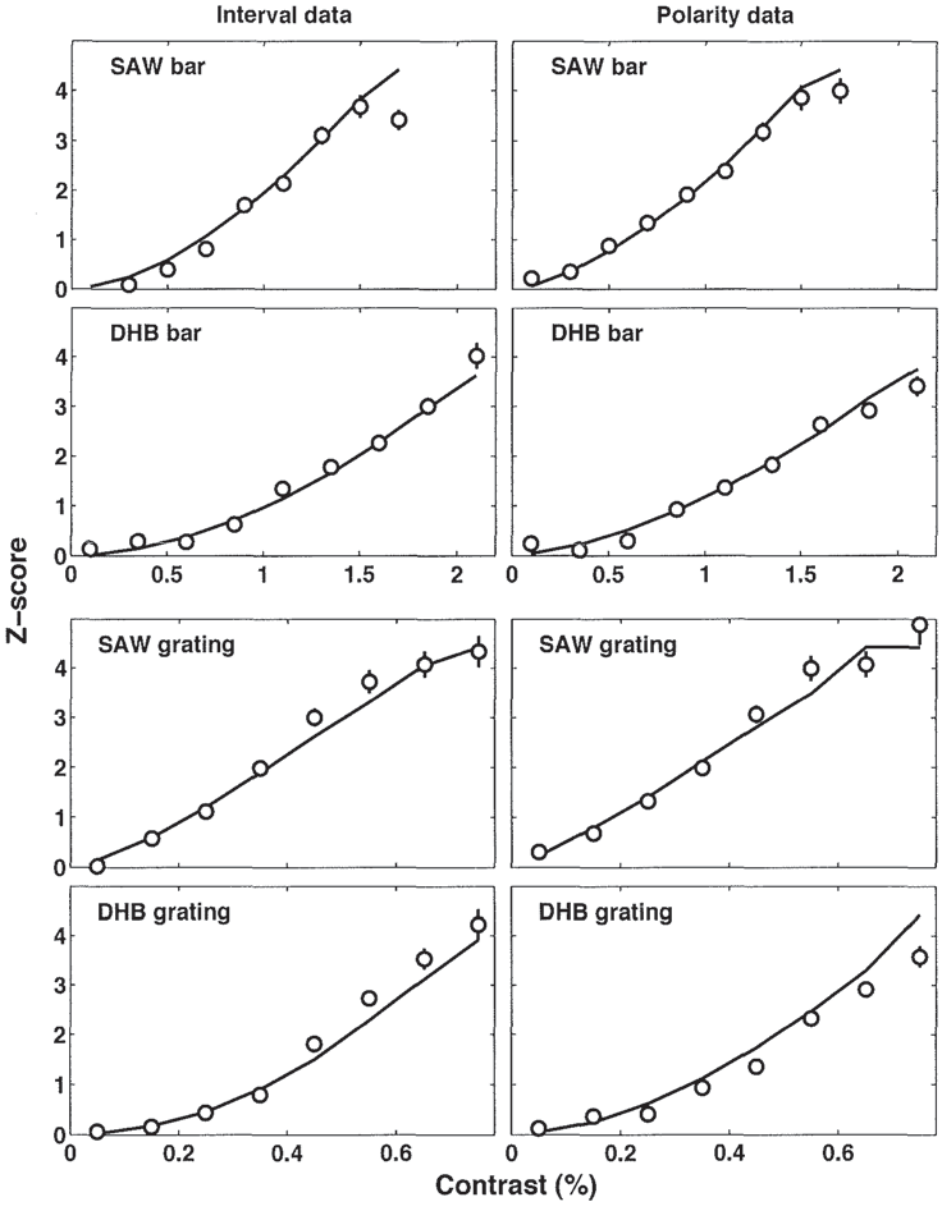

Figure 6.7: Bias-corrected z-scores (symbols) and simulation predictions (lines) as a function of stimulus contrast. The left column shows z('interval 2' responses) - z('interval 1' responses), and the right column shows z('light' responses) - z('dark' responses). Each panel is endorsed with the observer's initials and type of image. The bar data (upper 2 rows) has a larger contrast range than the grating data (lower two rows).

that it is approximately superimposed upon the light data, followed by a subtraction of dark from light. Since the dark data has negative sign, its subtraction from the light data is effectively an addition. So the net effect approximates a doubling of the light z-scores. A similar argument applies to the interval data, whereby the interval 1 z-scores are subtracted from the signal interval 2 z-scores. Bias-corrected z-scores for the experimental data are shown as symbols in figure 6.7.

The left column shows the interval data, and the right column shows the polarity data. Each panel

is endorsed with the observer's initials and type of image. The solid line shows the predictions of the Monte Carlo simulation, which were fitted to the data for each row separately. The details of the simulation are described next.

6.3.4 Monte Carlo Simulation

The simulation was performed with a custom Matlab program that was comprised of several stages. The first stage was a simple transducer that raised each stimulus contrast level (c) to a sign-preserving exponent (p):

$$r = sign(c).|c|^p; \tag{6.2}$$

This was carried out in a channel that represented the signal polarity (light or dark) and signal interval. Then Gaussian noise was added, the standard deviation of which was a free parameter. The simulation therefore had two free parameters; transducer exponent (p) and the standard deviation of the Gaussian noise.

The simulated response could be embedded in one of 4 noisy observations that represented the output of 2 polarity-specific channels in 2 observation intervals. Thus the model employed 4 parallel channels, which were

1. Signal polarity, signal interval
2. Signal polarity, nonsignal interval
3. Nonsignal polarity, signal interval
4. Nonsignal polarity, nonsignal interval

Since the signal could be in either interval and of either polarity, all 4 channels must be monitored. Thus all 4 channels are considered to be (potentially) 'relevant'.

Next, the maximum was taken over the 4 channels, and the identity of the channel that provided the maximum was noted. A correct 'interval' prediction was generated if the maximum was from one of the two signal interval channels, and an incorrect 'interval' prediction was generated if the maximum was from one of the two nonsignal interval channels.

This procedure was repeated many times (50,000 trials) for the same input contrast level. Although a constant distribution of noise values was used for all of the trials, the actual value of the noise had a different value on each trial. Then the predicted proportion of correct interval responses was generated by dividing the total number of correct responses by the total number of trials. The whole procedure was then repeated for the other contrast levels, which generated a full set of 'interval' predictions.

To generate 'polarity' response predictions, the channels were combined in one of two manners, determined by the selected decision rule. Both rules will be described next, before moving on to the next stage of the model. The first rule assumed that only the signal interval was monitored, because in the experiment, immediate feedback was given of the correctness of the interval response before the polarity response was made by the observer. Thus only 2 channels in the model were monitored here. If the response in the signal polarity, signal interval channel was greater than the response in the nonsignal polarity, signal interval channel, then the program returned a correct prediction for that

trial, and vice-versa for a incorrect prediction. This procedure was also repeated many times (50,000 trials) for the same input contrast level. The responses were combined to generate a 'polarity' response proportion, by the same method used for the 'interval' prediction.

The second rule assumed that both intervals were monitored when forming a polarity response. This rule was examined because the interval between the interval and polarity responses in most experimental trials was very short, which suggested that the observers might not have based their polarity response on the feedback given to their interval response. This rule was implemented by obtaining the maximum of all 4 channels and noting whether it came from a signal polarity or nonsignal polarity channel. As before, an overall 'polarity' response proportion was generated from many repetitions and the whole process was repeated for each stimulus contrast level.

Whichever polarity rule had been applied, the next stage of the model applied a lapse rate parameter. This was incorporated into the interval and polarity response proportions, by using

$$p(H) = p(H).(1 - lapse) + \frac{lapse}{2} \tag{6.3}$$

Where $p(H)$ is the proportion of correct interval or polarity responses and *lapse* is a parameter that was fixed at 0.00001 (i.e. 0.001%). This small lapse rate prevented the simulation from producing 0% or 100% responses that would result in infinite z-scores in the next stage of the simulation.

The proportion of correct responses were then converted to z-scores by using the inverse standard normal integral. The z-scores were then doubled, in order to simulate the correction for bias that transforms the data into the form shown in figure 6.7. An exact doubling was appropriate here because the simulation is inherently free of bias, and the terms $p(H)$ and $p(FA)$ in equation 6.1 are equivalently described as $p(correct)$ and $p(incorrect)$ here. Thus

$$\begin{aligned} z_{model} &= z[(p(correct)] - z[p(incorrect)] \\ &= z[(p(correct)] - z[1 - p(correct)] \\ &= z[(p(correct)] - (-z[(p(correct)]) \\ &= 2z[(p(correct)] \end{aligned}$$

In the final stage of the simulation, an rms error score was computed that measured the difference between the predicted and empirical z-scores for both interval and polarity responses. The error score (E) was inverse variance weighted, and was defined as

$$E = \sqrt{\sum \left(\frac{z_{model} - z_{data}}{sd_{data}} \right)^2} \tag{6.4}$$

where z_{model} is the z-score generated by the model, z_{data} is the z-score data, and sd_{data} is the expected standard deviation of the z-score data. The values for sd_{data} were obtained by applying the following formula for the expected variance of a z-score:

$$var[z(p)] = \frac{p(1-p)}{N[\Phi(p)]^2} \tag{6.5}$$

where N is the number of trials and $\Phi(p)$ is the height of the normal density function at $z(p)$ (Gourevitch & Galanter, 1967).

The best fitting parameters were used to generate the predictions that are shown as solid lines in figure 6.7. These parameters are shown in table 6.4. These values suggest that SAW's bar data and all of DHB's data are relatively nonlinear ($p \geq 1.5$). However, SAW's grating data suggest almost linear performance ($p = 1.15$).

stimulus	observer	p	*noise* (%)	error score
bar	SAW	1.50	0.65	9.10
	DHB	1.60	1.20	7.70
grating	SAW	1.15	0.20	6.68
	DHB	1.75	0.20	10.83

Table 6.4: Best-fitting parameter values for the simulation lines shown in figure 6.7. The lapse rate was fixed at 0.001%.

Further repetitions of the simulation were performed to examine the effect of the polarity rule selected. Whether the simulation used just the signal interval values (rule 1) or also the values (purely noise) from the nonsignal interval (rule 2) made little difference to the resulting fits. A switch of the rule caused no more than a ± 0.05 difference to the best-fitting value of p or *noise*, without any systematic pattern of change across the four data sets. The error score was also altered by no more than 2.5. But since the error score was slightly lower for rule 1 than rule 2 in three of the four data sets, rule 1 was adopted throughout.

6.3.5 Discussion

The aim of this experiment was to examine the conditions under which linear (or near-linear) performance occurs. In an attempt to increase the likelihood of linear performance, 2 observers were used that had many hours of previous practice with the 2 image types. Yet only one observer (SAW) achieved near-linear performance, and that was with the grating stimuli only. If a simple transducer model applies, then the exponent power (p) should remain approximately constant for each observer. While this hypothesis is supported by DHB's data ($p = 1.60$ and 1.75), SAW's data are slightly less convincing ($p = 1.50$ and 1.15). Although the Monte Carlo simulation did not return confidence limits, the values of 1.50 an 1.15 may not be significantly different from each other; the psychometric fit applied to the previous grating and bar data for SAW returned confidence limits that were about ± 0.2. Therefore a simple transducer model might be sufficient to account for his data here.

However, stimulus uncertainty also provides a good account of this experiment's data. The Monte Carlo simulation provided a very good fit to the interval and polarity data for each observer and type of image (see figure 6.7), which suggests that it correctly captures at least some of the visual system's processing of these images. The simulation assumed independence of the 4 parallel channels, so the good fit of the its predictions suggests that 'light' responses are independent of 'dark' responses, or in other words, that dark and light stimuli are processed separately by dark and light channels. This is in contrast to the notion that there is a single opponent dark-light channel that can process both dark and light stimuli. A more definitive test of these competing accounts will be provided in the next experiment.

In order to obtain similar performance, the contrast range used for the grating stimuli (0.05 - 0.75%)

was less than that of the bar stimuli (0.1-2.1%) by a factor of about 3. This demonstrates greater sensitivity to the grating images. This may reflect the combined effect of 2 factors. Firstly, different definitions of contrast that were used when generating each type of image. Michelson contrast was used for the grating, and Weber contrast was used for the bar. This approach means that when the bar and grating were of equal contrast, their maximum luminance was identical. The luminance profile of the light Gaussian bar had the same height and very similar width as the light bars of the grating. Yet the grating also contained dark bars, which were likely to act as a cue for the grating's presence. Since the grating had twice the amplitude of the bar, it might be expected that grating sensitivity would be a factor of 2 higher than bar sensitivity.

Secondly, the bar image was effectively narrower than the grating image, because it consisted of a single bar and yet the grating consisted of about 7 light and dark bars (3.5 periods), each of similar width to the Gaussian bar. Thus the effect of area summation (Legge, 1978a; Meese, 2004; Meese & Summers, 2007) is likely to increase grating sensitivity further, relative to the Gaussian bar. However, a detailed discussion of area summation is beyond the scope of this thesis.

A comparison of the left and right panels of figure 6.7 shows the strong similarity between the interval data and polarity data, which is also reflected in the simulation predictions. Both observers were as good at determining the correct polarity of the stimulus (polarity identification) as the correct interval in which it was displayed (detection). This finding is in keeping with previous research. Using a 2 x 2AFC method, as here, Tolhurst & Dealy (1975) found that the threshold for detection was only about 0.05 log units lower than the threshold for polarity identification, for both bar and edge stimuli. They suggested that the presence of polarity information at detection threshold provided evidence for 'labelled lines' in human vision, whereby the channel that transmits the stimulus is labelled with its polarity.

Such a conclusion depends partly upon the equivalence of the detection and polarity identification tasks. It might be thought that in a 2 x 2AFC design, the 2 tasks are not equivalent. The detection task is effectively 2AFC, because both possible stimuli (target and blank) are presented during each trial. However, the polarity identification task is not 2AFC, because dark and light stimuli are not presented in each trial. Since only one stimulus polarity is presented in each trial, this task is better described as single interval (although with uncertainty as to which interval contains the target). The distinction between 2AFC and single interval is important because it impacts upon how the two tasks are compared. In signal detection theory, 2AFC performance is expected to be $\sqrt{2}$ better than single interval performance (i.e. z-scores $\sqrt{2}$ higher), as there are two sources of information (2 intervals) rather than just one (Macmillan & Creelman, 2005). So 2AFC z-scores would need to be divided by $\sqrt{2}$ before a meaningful comparison could be made with single interval z-scores.

However, to apply this analysis to the 2 x 2AFC method would be incorrect. The detection and polarity identification tasks can be regarded as equivalent, if considered as follows. In detection, the observer has to identify the correct interval from 2 alternatives, but is uncertain of the target's polarity. In polarity identification, the observer has to identify the correct polarity from 2 alternatives, but is uncertain of the correct interval. Thus the amount of uncertainty in the 2 tasks is equal. The topic of uncertainty will be discussed further in the next experiment.

The 2 x 2AFC method was also used by Huang et al. (2006), who examined detection and polarity identification performance with stimuli that were Gaussian blobs, 1cpd or 0.5cpd Gabors, edges, bars or edge-bar hybrids. Performance indicators were the threshold and slope of the Weibull function fitted

to the psychometric data. Polarity identification performance was as good as detection performance, for most observers, only for the Gaussian blobs. The stimuli that were most similar to those used in the present experiment were the 1cpd Gabor and the bar. For the 1cpd Gabor, all four observers had higher thresholds for the polarity identification task than the detection task (by a factor of about 1.3 to 2.3). This is different from the present experiment which found very similar performance for the two tasks. However, Huang et al. (2006) used positional jitter between each trial and had no fixation markers, which is likely to have increased phase uncertainty. That would have a detrimental effect that is much greater for the polarity identification task than the detection task. For the bar, one of the four observers *did* perform equally well in the two tasks. But the other three observers had higher thresholds for the polarity identification task (by a factor of about 1.3 to 1.7). Since the bar was known to be either light or dark in this condition, phase uncertainty is unlikely to play a role here.

In the present experiment the data suggest that the linearity of performance depended on the image type more so for SAW than DHB. This apparent difference between observers might reflect differing amounts of prior practice with the images, if extensive practice helps to reduce stimulus uncertainty. If practice is a factor here, then a comparison of the present data with that of experiment 8 (which used identical stimuli) might be expected to reveal an increase of linearity over time. Table 6.5 shows the exponent values (p) that fitted the data in these two experiments. The right column shows the change of the exponent over time. For 3 of the 4 data sets the exponent has increased, albeit by a small amount (≤ 0.34). This is a shift in the opposite direction to that expected with practice. This decrease may be due to methodological differences between the two experiments. Experiment 8 used a single interval procedure and polarity identification, whereas experiment 9 used 2 x 2AFC. The uncertainty that is an intrinsic part of 2 x 2AFC may have caused an increase in the exponent values, that swamped the reduction (if any) caused by greater practice.

stimulus	observer	Expt 8 exponent (p_8)	Expt 9 exponent (p_9)	$p_9 - p_8$
bar	SAW	1.23	1.50	0.27
	DHB	1.26	1.60	0.34
grating	SAW	1.00	1.15	0.15
	DHB	2.21	1.75	-0.46

Table 6.5: Best-fitting exponent values for the data of experiments 8 and 9. The right column shows the change of the exponent over time.

Nevertheless, it is becoming clear from this set of results (experiments 8 - 9) that contrast transduction can be much closer to linear than previous studies have supposed. The mean exponent value taken over 2 experiments and 3 subjects is 1.36 $\pm$0.17 for the bar and 1.46 $\pm$0.45 for the grating.

The next experiment will explore the use of different methods (in addition to other manipulations) in order to explore more thoroughly the effect of uncertainty and the linearity of performance.

6.4 Experiment 10 - Single interval and 2AFC, detection and polarity identification of a Gaussian bar

6.4.1 Introduction

In complete darkness, our perception is not of total blackness, but rather a dark grey. As Hering (1964) observed, as the light level increases, the range of visible greys also increases. He noted that light areas would appear lighter, but interestingly, that dark areas appeared darker, even though the light level was *increasing*. This observation, coupled with the existence of negative after-images, led Hering to propose that opponent processes had a large part to play in human vision. He proposed the colour-opponent theory that survives to this day. This was years before electophysiological support even existed that a neurone could inhibit a second neurone. This approach suggests that both darkness and lightness are represented in the visual system, at some stage, within a single 'bipolar' channel that spans from very dark to very light (figure 6.8a).

The prevalence of bipolar cues in visual perception was noted by Klein (1985). He suggested that the perception of luminance change, Vernier alignment, displacement, depth and discrimination between edge and line could be influenced by the bipolar cues of luminance, orientation, velocity, disparity, and luminance respectively. He analysed the spot position data and detection data of Allik et al. (1982) and found that the relation between the 2 data sets fitted a bipolar model but not a monopolar model.

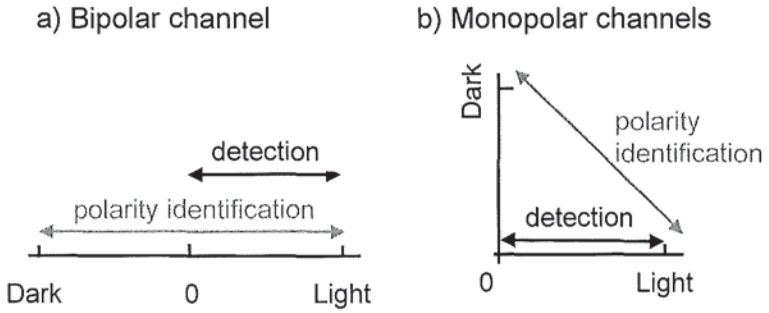

Figure 6.8: a) A single bipolar channel. b) A pair of monopolar channels.

An alternative to the bipolar approach is to consider that light and dark signals are processed by independent channels, each of which can carry the signal of only one polarity. The complete independence of 2 channels can be represented by using orthogonal axis (figure 6.8b), because a change along one axis leads to no change at all along the other axis.

Psychophysical evidence for separate channels has been provided in a number of studies. The separate processing of dark and light signals has physiological support provided by Kuffler (1953), who identified two types of retinal ganglion cell in the cat retina: 'ON-centre' cells were excited by light increments in their central region, and 'OFF-centre' cells were excited by light decrements. This segregation of ON and OFF pathways appears to continue right up to the final perception of light or dark. Schiller et al. (1986) used a neurotransmitter analogue in the monkey retina to selectively block the ON channel. The monkeys had first been trained to make saccades to target stimuli. After the ON channel was

blocked, there was a large decrease in detection performance for light squares, but not dark squares.

This method does not, however, allow us to distinguish between the monopolar and bipolar models shown in figure 6.8, because the blocking of a 'light' pathway may effectively turn the bipolar channel into a monopolar dark channel. But a comparison of detection and polarity identification *can* allow us to distinguish between the 2 models.

In signal detection theory, it is assumed that all the information that is used to make a response is combined along a single 'response axis' in 'decision space'. This is the axis shown in figure 6.8a. When a light stimulus is at detection threshold it causes an internal response that is a particular amount (such a $d' = 1$) above the response caused by a blank stimulus. This distance along the response axis is labelled 'detection' in figure 6.8a.

Let us assume that the presentation of a dark stimulus produces a mean internal response that is the same magnitude (but of opposite sign) to that produced by a light stimulus. If the task is polarity identification, then the internal response will either be light or dark. The relevant distance along the response axis is now twice as long as it was for detection (figure 6.8a). This means that polarity identification performance should be twice as good as detection performance. So z-scores for polarity identification should be double the z-scores of detection.

Now consider the pair of monopolar channels shown in figure 6.8b. Detection of light stimuli is represented by the same as before, but now dark stimuli detection is represented along the vertical axis. Polarity identification is once again indicated by the distance between the 'light' and 'dark' points, but now that the axes are orthogonal, this distance is now $\sqrt{2}$ longer than the detection distance, rather than twice as long. This means that z-scores for polarity identification should be $\sqrt{2}$ higher than the z-scores of detection. The first aim of the present experiment is to determine the extents to which the bipolar model and monopolar model provide a good account of the data, by comparing detection z-scores with polarity identification z-scores. Whether the latter are closer to a factor of 2 or $\sqrt{2}$ higher than the former will identify the model that more closely resembles human visual processing.

The second aim of this experiment is to distinguish between two alternative accounts of nonlinear low-contrast performance, known as the 'transducer model' and the 'uncertainty model', and to examine how each model can encompass linear performance. The transducer model (Nachmias & Sansbury, 1974) assumes an accelerating function that relates the sensory response to contrast (figure 6.1b) and was described in section 6.1. The uncertainty model was proposed by Pelli (1985) as an alternative to the nonlinear transducer model. It assumes that the transducer may actually be linear, but that poor performance at low contrast levels is due to the observer's uncertainty about which of many noisy channels transmits the signal. As contrast is raised, fewer irrelevant channels are monitored, leading to progressively improved performance, with an accelerating relation between d' and contrast, much like the transducer model. The Monte Carlo simulation that was applied to the previous experiment's data (and described in section 6.3.4), will be expanded so that it can operate in 2 modes; a 'transducer' mode that was described previously, and a new 'uncertainty' mode, which will be described in the next section.

The third aim of this experiment is to comprehensively explore the effect of method on performance. Both single interval and 2AFC will be used in separate conditions, and performance compared between them. As stated in the previous section, signal detection theory predicts a $\sqrt{2}$ improvement in performance when single interval is replaced by 2AFC. This is now illustrated in figure 6.9 for detection

of a light stimulus, where bipolar channels are assumed. Panel a shows single interval detection and

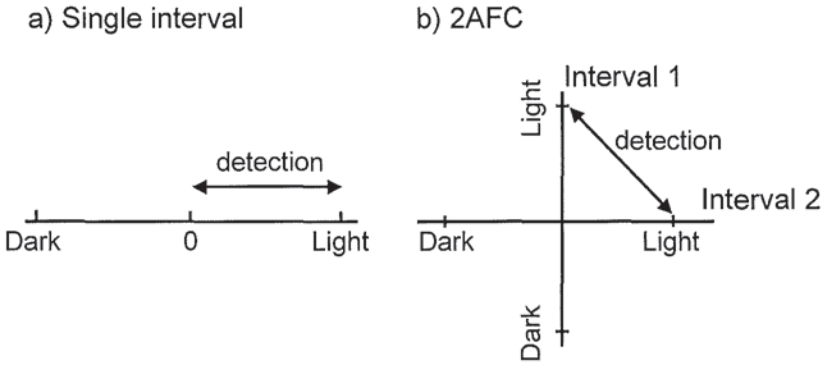

Figure 6.9: Decision space for detection of a light stimulus, using a) single interval b) 2AFC.

panel b shows 2AFC detection. In single interval detection, when the stimulus is of known polarity, the observer knows that in each trial it will either be present or absent. These two alternatives are represented by the ends of the arrow in panel a, labelled 'detection'. In 2AFC detection, when the stimulus is of known polarity, the observer knows that in each trial it will either be in interval 1 or interval 2. These two alternatives are represented by the ends of the arrow in panel b, also labelled 'detection'. Here, the axes now denote interval 1 and interval 2, and are orthogonal to show the independence of these 2 sources of information. Since the arrow in panel b is $\sqrt{2}$ longer than in panel a, the z-scores are expected to be $\sqrt{2}$ higher for 2AFC than single interval. By symmetry, a similar argument applies to detection of dark stimuli.

The $\sqrt{2}$ improvement when switching from single interval to 2AFC is also expected for the polarity identification task, if bipolar channels are assumed. Note that 2AFC (as opposed to 2 x 2AFC) means that a light stimulus is presented in one interval and a dark stimulus in the other interval. Figure 6.10a shows the decision space for single interval, and the arrow indicates the polarity identification distance. Panel b shows the decision space for 2AFC. Here, point A indicates the internal response when the order of stimulus presentation is light-dark (LD). Point B indicates a dark-light (DL) order. Note that the arrow here is $\sqrt{2}$ longer than in panel a. This figure illustrates a bipolar channel model, yet in reality a monopolar model may be more appropriate. However, a monopolar model would require 4 dimensions to illustrate in this way (one for each of interval 1 light, interval 1 dark, interval 2 light, interval 2 dark), so is impossible to draw. It is not clear whether a monopolar model would still predict the $\sqrt{2}$ relation between single interval and 2AFC. This experiment will examine this issue by employing a model that uses monopolar channels.

The present experiment will use single interval and 2AFC methods with detection and polarity identification tasks to comprehensively test the expected $\sqrt{2}$ relation. Although this relation is specified by Macmillan & Creelman (2005), it is not a foregone conclusion that empirical data will follow suit. Firstly, the relation assumes ideal observer behaviour. Yet the observers in the present experiment may not adopt an optimal decision strategy. Secondly, the memory load is heavier in single interval than 2AFC, because the observer is comparing each stimulus against his memory of the alternative stimulus. Yet in 2AFC, both alternatives are typically presented in quick succession in each trial. This

might selectively impair single interval performance. Thirdly, empirical work has produced variable values of the $d'(2AFC)/d'(single\ interval)$ ratio. Table 6.4.1 shows values of this ratio reported in a number of previous studies. While these values encompass the expected $\sqrt{2}$, the range (0.86 - 1.97) is quite wide. Furthermore, these studies were all auditory, not visual.

Source	$d'(2AFC)/d'(single\ interval)$
Swets et al. (1959)	0.96
Schulman & Mitchell (1966)	1.17
Markowitz & Swets (1967)	1.17
Leshowitz (1969)	1.19
Viemeister (1970)	1.91
Pynn et al. (1972)	1.97
Watson et al. (1973)	1.87
Jesteadt & Bilger (1974)	1.50
Jesteadt & Sims (1975)	0.86
Creelman & Macmillan (1979)	1.35
Range	0.86 - 1.97
Mean	1.40

Table 6.6: Values of $d'(2AFC)/d'(single\ interval)$ from previous auditory studies.

The final methodological manipulation that the present experiment will examine is the effect of using blocks of known-polarity stimuli compared to the use of interleaved-polarity blocks. All of the previous experiments in this thesis followed the well-established tradition of randomly interleaving stimulus polarity, in order to prevent the build up of adaptation, and to randomise the effect of after-images. However, this technique may induce a small amount of stimulus uncertainty, which could impair performance. The final aim of the present experiment is to examine this effect. Since it is likely to be small, a large number of trials will be necessary.

Using a vertical Gaussian bar stimulus, psychometric functions will be obtained for 4 tasks: detection and polarity identification, with either single interval or 2AFC presentation. The psychometric functions will be sampled very intensively (9 to 18 contrast levels at 240 or 300 trials each), which allows

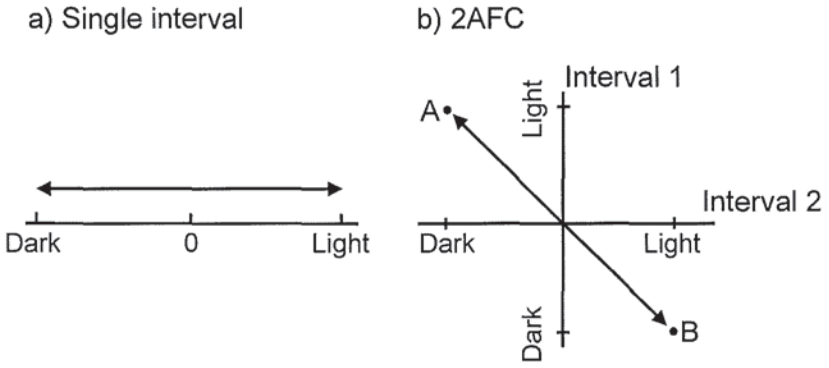

Figure 6.10: Decision space for polarity identification, using a) single interval b) 2AFC. The arrow in each panel indicates the polarity identification distance.

examination of the effect of subtle manipulations, such as the switch from blocked to interleaved trials. It also allows a close examination of the relationship between different tasks.

6.4.2 Method

The stimulus was a dark or light vertical Gaussian bar (σ = 12mins) on a mid-grey background that subtended 4.3 degrees at the viewing distance of 123.4cm (1pixel = 1degree). It was surrounded by a full screen of mid-grey. The stimulus was displayed at a number of contrast levels on a gamma-corrected screen at 14-bit resolution. The range of contrast levels was determined separately for each task and observer in a pilot study. The Weber definition of contrast was applied. A method of constant stimuli was used with the following 4 tasks:

1. Single interval detection (blocked light, blocked dark, or interleaved)
2. 2AFC detection (blocked light, blocked dark, or interleaved)
3. Single interval polarity discrimination
4. 2AFC polarity discrimination

The detection tasks each had 3 variants, according to whether the stimuli were always light ('blocked light'), always dark ('blocked dark') or of either polarity ('interleaved'). In the interleaved condition, half of the bar-present intervals contained a light bar and the other half contained a dark bar. Therefore there were a total of 8 different conditions across the 4 tasks.

In single interval detection, each trial contained either a Gaussian bar (proportion = 0.5) or a blank screen of mid grey (proportion = 0.5), and the task was to indicate whether the bar was present or not.

In 2AFC detection, there were 2 temporal intervals (duration 300ms) separated by a 600ms presentation of a mid grey screen. One interval contained the Gaussian bar and the other interval contained a blank screen of mid grey, and the task was to indicate the interval that contained the bar.

In single interval polarity discrimination, a Gaussian bar was presented on every trial, but its polarity could be either light (proportion = 0.5) or dark. The task was to report its polarity.

In every 2AFC polarity discrimination trial, a light bar was presented in one interval, and a dark bar was presented in the other interval. The task was to report the interval that contained the light bar.

A small (2 x 2 pixels) central black fixation dot was displayed continuously, except during the single interval and 2AFC intervals. The intervals were marked by a 300ms tone. Immediate auditory feedback was given of the correctness of every response. There was a minimum of 1 second between each trial. Each run consisted of 10 repetitions of a single condition and took 5-10mins to complete. A session consisted of a block of 6 runs of the same condition, and was often completed at one sitting (but with breaks between runs, as the observer saw fit).

There were two observers. One (PRM) was a 18-year old female, who had no prior practice in psychophysics. She performed 300 trials per contrast level in a randomised order of blocks. This took about 30 hours to complete. The other (SAW) was a 42-year old male, who was highly practised with these stimuli. He had previously provided single interval polarity identification data with a

Gaussian bar (σ = 12mins) in experiment 8, so that condition was omitted here. Since the design of this experiment was developed over a few months, he provided all of his 2AFC polarity identification data next. A month later, he provided all of his single interval and detection data, which was collected in randomly interleaved blocks of 6 runs. All of his data were at 240 trials per contrast level.

6.4.3 Monte Carlo Simulation

The data from all 8 conditions for a given observer were fitted with a Monte Carlo Simulation that was very similar to that used in the previous experiment (described in section 6.3.4). However, the program was amended in two ways. Firstly, the program was expanded to include a factor that represented the amount of uncertainty that was introduced by the monitoring of irrelevant channels (i.e. channels that contained no signal, only noise). Secondly, the decision rules were amended and expanded to simulate the 8 conditions used here. Each of these changes now will be described in detail.

In the previous version of the simulation there were only 4 channels:

1. Signal interval, signal polarity
2. Signal interval, nonsignal polarity
3. Nonsignal interval, signal polarity
4. Nonsignal interval, nonsignal polarity

These were chosen to model the 2 x 2AFC method, whereby the target stimulus could appear in either interval, and could be of either polarity (light or dark). Since the signal could be in either interval and of either polarity, all 4 channels must be monitored. Thus all 4 channels are considered to be (potentially) 'relevant'.

The updated version of the simulation introduced uncertainty by adding a variable number of 'irrelevant' channels. They always contained only Gaussian noise, and represented the combined effect of all sources of uncertainty. This addition was implemented by adding an extra dimension to the simulated data array whose length equalled the number of irrelevant channels. Thus each of the 4 relevant channels now had an additional number of irrelevant channels that carried its label. For each trial, the program then took the maximum across one relevant and all of its irrelevant channels to generate a final value for that channel. This maximum operation was also carried out in each of the other 3 relevant channels with their respective irrelevant channels.

The simulation could be run in one of two modes. The first was 'transducer mode', in which the transducer exponent (p) was free, but the number of irrelevant channels was zero (as used in the simulation for the previous experiment). The second was 'uncertainty mode', in which the transducer exponent (p) was fixed at 1, which made it linear. The number of irrelevant channels was free to vary, from zero to any positive integer.

The second change made to the previous version of the simulation concerned how the decision rules were applied. In the previous version, the polarity identification task was within a 2 x 2AFC design. Thus the target stimulus could be light or dark, and could appear in either interval. The polarity identification rule was to find the channel that contained the maximum amplitude, and then determine whether that channel was of signal or nonsignal polarity (data pooled across interval). If it was from

a signal polarity channel, a 'correct' response was generated, and if it was from a nonsignal polarity channel, an 'incorrect' response was generated.

In the present experiment there were 2 different polarity identification tasks; single interval, and 2AFC. In the single interval condition, the simulation subtracts the nonsignal polarity channel value from the signal polarity channel value to compute a 'difference score'. If the difference score is above zero, a 'correct' response is generated. If the difference score is below or equal to zero, an 'incorrect' response is generated. This is the same as the polarity identification process in the previous version of the simulation.

In the 2AFC polarity identification condition, a difference score was computed separately for each interval. A 'correct' response was generated if the difference score in the signal interval was greater than the difference score in the nonsignal interval. Otherwise an 'incorrect' response was generated. In a large number of simulated trials, the difference score was positive in the signal interval and negative in the nonsignal interval. This decision rule simulated the experimental procedure for this condition, where the observer knows that a light stimulus was in one interval and a dark stimulus was in the other interval, and assumes that he combined the information from each interval in this (possibly optimal) manner.

The simulation was able to replace this second decision rule with an alternative rule that also seemed plausible. In this case, the channel was identified that gave the maximum value. A 'correct' response was generated if the maximum was from either the Signal interval, signal polarity channel (Interval 1, light) or the nonsignal interval, nonsignal polarity channel (interval 2, dark). Separate runs of the simulation were used to explore the effect of decision rule selection.

The decision rules for the detection data were also added to the simulation. In the experiment, both blocked (polarity known) and interleaved (polarity unknown) conditions were explored. In the simulation, to generate blocked detection predictions, only the known polarity channels were monitored. Thus the nonsignal polarity channels were ignored (and their linked irrelevant channels). The output of this stage was 2 channels; one signal, the other nonsignal. To generate interleaved detection predictions, both signal and nonsignal polarity channels were monitored. Then the maximum was taken over the 2 polarities, so that the output of this stage was also 2 channels; one signal, the other nonsignal.

For single interval detection, a criterion was set that gave a plausible proportion of false alarms (0.35), by using the data in the nonsignal channel (which contained only noise). Then a 'correct' response (i.e. a 'hit') was generated if the value in the signal channel exceeded the criterion and an 'incorrect' response (i.e. a 'miss') was generated if the value in the signal channel was below the criterion.

For 2AFC detection a 'correct' response was generated if the signal interval channel value exceeded the nonsignal interval channel value. Otherwise an 'incorrect' response was generated.

The next stage of the simulation computed z-scores. Since the experimental data were corrected for bias, an equivalent procedure is required in the simulation. In every condition the z-score was calculated from an application of the standard formula for d' (see equation 6.1). Hits and false alarms were defined as p(correct) and p(incorrect) respectively for every task, except for single interval detection, where the false alarm rate was set at 0.35. An alteration of this value (within reasonable limits) had little or no effect on the output of the simulation, because of the correction for bias.

The final stage of the simulation was to compute the best fit to all of the experimental data for a given

observer. This was performed separately for 'transducer' and 'uncertainty' runs of the simulation. As before, the error score was an inverse variance weighted sum-of-squared differences between data and simulation predictions.

6.4.4 Results

The single interval detection data for both observers showed a substantial false alarm rate (SAW 0.35, PRM 0.29). This bias was corrected by obtaining d'. In order to be consistent across tasks, an equivalent manipulation was performed for each set of data. The definitions of 'hit' and 'false alarm' were:

Single interval polarity identification

Hit: 'Light' response to light stimulus

False alarm: 'Light' response to dark stimulus

2AFC polarity identification

Hit: '1st interval' response to light in 1st interval

False alarm: '1st interval' response to dark in 1st interval

Single interval detection

Hit: 'stimulus present' response to stimulus

False alarm: 'stimulus present' response to blank

2AFC detection

Hit: '1st interval' response to stimulus in 1st interval

False alarm: '1st interval' response to stimulus in 2nd interval

Figure 6.11 shows performance expressed as z-score differences [z(Hit) - z(FA)] of the blocked detection data (averaged across polarity) and the polarity identification data as a function of absolute contrast, for each observer (columns) and for single interval and 2AFC methods (rows). The error bars ($\pm 1sd$) are derived from the expected variance of each z-score. The lines in each panel show the detection data multiplied by $\sqrt{2}$ (solid line) or 2 (dashed line). The polarity data (triangles) are generally closer to the solid line than the dashed line, which shows that the ratio of performance between polarity identification and detection is closer to $\sqrt{2}$ than 2.

The interleaved detection data have not been plotted here because only one set of detection data is required to compare with the polarity identification data. The blocked detection data was chosen over the interleaved detection data because that task is arguably more comparable to the polarity identification on the basis of stimulus uncertainty. In single interval polarity identification, where a light or dark stimulus is presented on every trial, the stimulus can be either of 2 possibilities: light or dark. Likewise in single interval detection (blocked light), the stimulus is either light or a blank screen, which is again 2 possibilities. However, in the interleaved condition of single interval detection, the stimulus could be light, dark or blank, which is 3 possibilities. So there is greater uncertainty here, and that may impair performance.

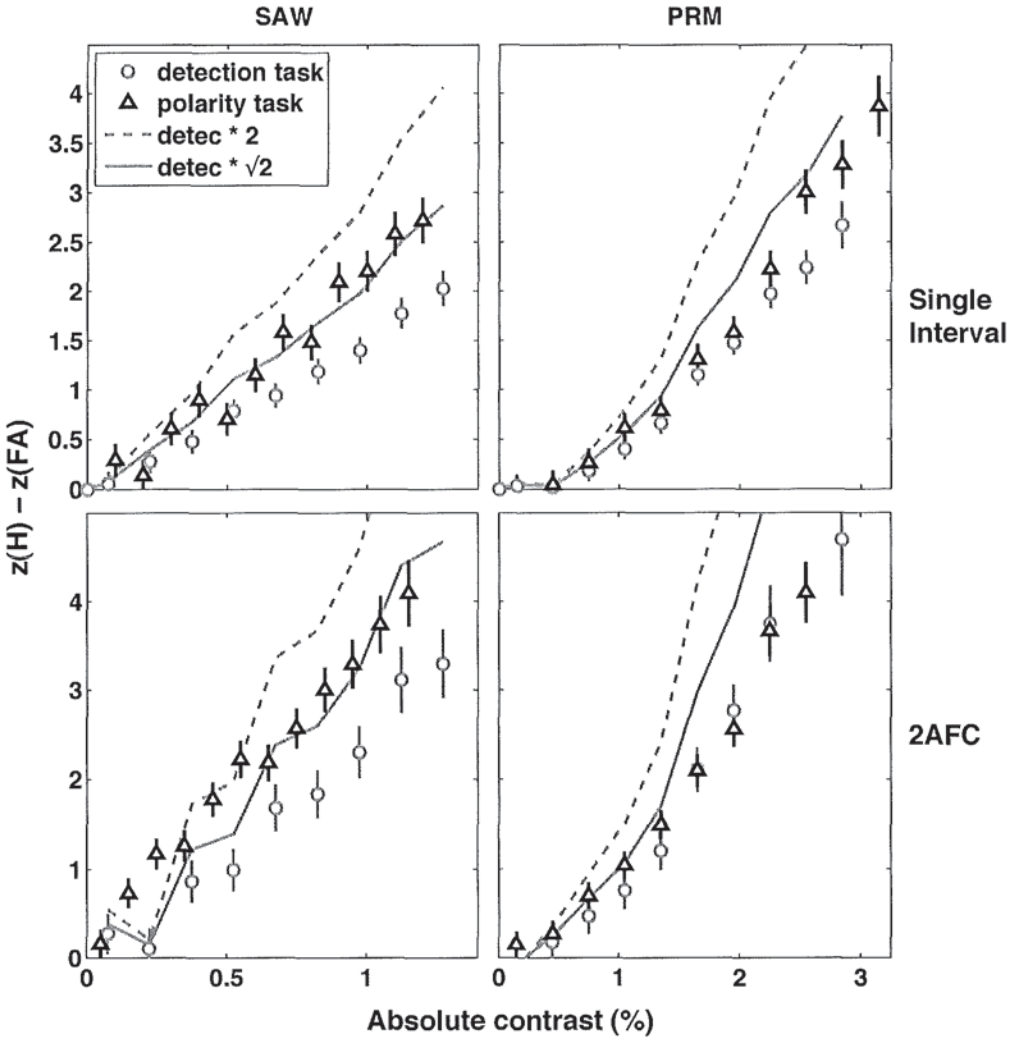

Figure 6.11: Z-scores of the blocked detection, averaged across polarity, and polarity identification as a function of absolute contrast, for each observer (columns) and for single interval and 2AFC methods (rows). The error bars ($\pm 1sd$) are derived from the expected variance of each z-score. The lines in each panel show the detection data multiplied by $\sqrt{2}$ (solid line) or 2 (dashed line).

Although figure 6.11, suggests that a $\sqrt{2}$ relation fits better than a factor of 2, PRM's data do not fit either value particularly well. Limiting the choice to just $\sqrt{2}$ or 2 appears to be too restrictive. So the data were re-plotted in a manner that allowed greater variation in the multiplier applied to the detection data.

Figure 6.12 shows polarity identification z-scores as a function of detection z-scores. Separate panels are used for each observer. Since SAW used slightly different contrast levels in the 2 tasks, his detection data required linear interpolation to generate z-scores at the contrast levels used for polarity identification. In each panel, the dashed lines show (in order of increasing steepness) a linear, $\sqrt{2}$ and factor of 2 relationship between polarity identification and detection. The solid line shows the best-fitting linear function, constrained to pass through the origin. The best fit was obtained by minimising the orthogonal distance between data and line. This method (known as 'total least-squared error') was

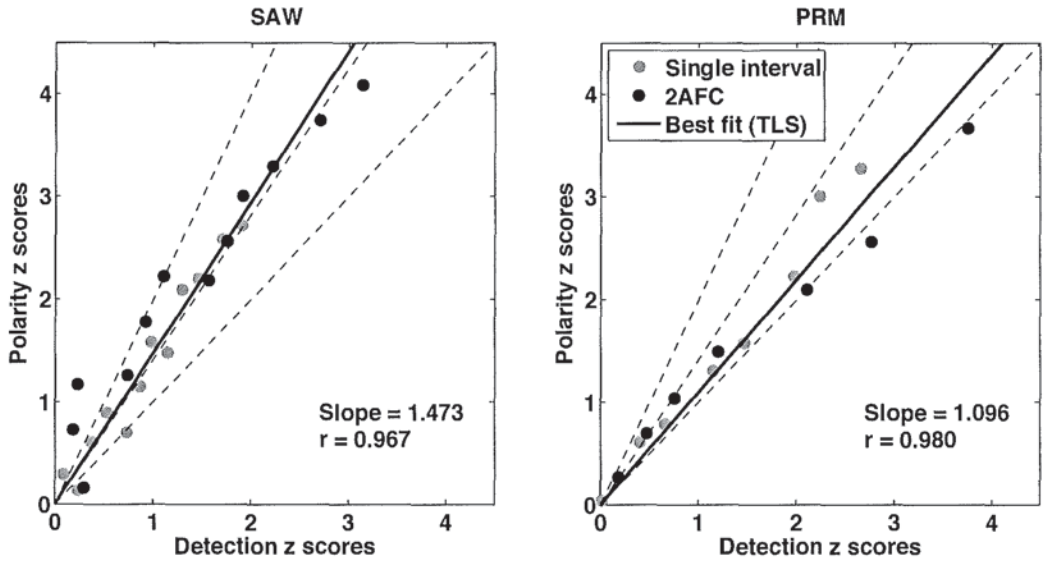

Figure 6.12: Polarity identification z-scores as a function of averaged blocked detection z-scores. Separate panels are for each observer. The dashed lines show (in order of increasing steepness) a linear, $\sqrt{2}$ and factor of 2 relationship between polarity identification and detection. The solid line shows the best-fitting (total least square) linear function, constrained to pass through the origin.

used because the x and y values were both subject to error. The slope of the line and the correlation coefficient of the data are shown in each panel. The slope of the best-fitting line is 1.47 for SAW, which is very close to the value of $\sqrt{2}$ predicted by independent light and dark channels. But for PRM, the slope is only 1.10, which show that her polarity identification performance was only very slightly better than her detection performance.

In order to examine the degree to which 2AFC performance was better than single interval, all of the data were plotted with an abscissa of single interval performance and an ordinate of 2AFC performance. The resulting plots for each observer are shown in figure 6.13. Positive and negative z-scores relate to light and dark stimuli respectively. There are more data points here than in figure 6.12 because the blocked detection data have not been averaged, and the interleaved detection data are now shown. The dashed and solid lines are as before. For both observers 2AFC performance was a factor of 1.7 better than single interval, which is higher than the value of $\sqrt{2}$ predicted from standard SDT. The high correlation coefficients (both 0.99) show the very consistent nature of this relationship, across the variations of procedure (blocked vs interleaved) and type of task (detection vs polarity identification).

Like standard SDT, the current version of the Monte Carlo simulation predicts that 2AFC performance is $\sqrt{2}$ higher than single interval performance. But the improvement observed here and elsewhere (Jesteadt & Bilger, 1974; Jesteadt & Sims, 1975) is higher than that. Jesteadt & Sims (1975) proposed that single interval performance can be poorer than expected because of a greater criterion variance. Variation in a criterion's position is equivalent to introducing an additional source of noise. So the Monte Carlo simulation was revised to include an additional parameter (k) that reduced the noise level by a fixed factor for 2AFC relative to single interval.

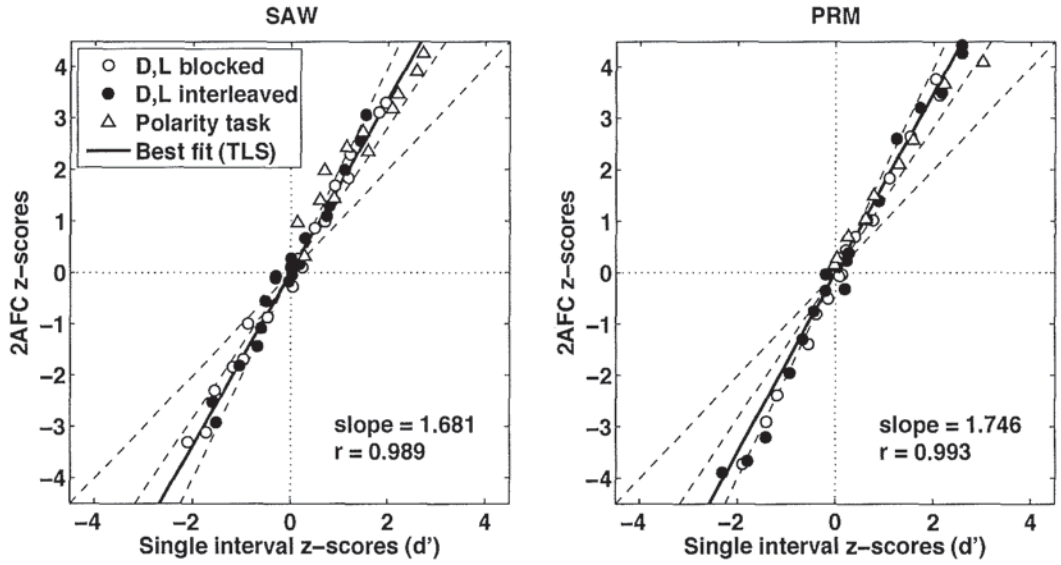

Figure 6.13: 2AFC z-scores as a function of single interval z-scores. Separate panels are for each observer. Positive and negative z-scores relate to light and dark stimuli respectively. The dashed and solid lines are as figure 6.12.

The Monte Carlo simulation was fitted to all of the data, separately for each observer. Simulations were run in either 'transducer mode' (number of irrelevant channels fixed at zero; transducer exponent free) or 'uncertainty mode' (number of irrelevant channels free; transducer exponent fixed at 1). Many runs of the simulation were performed, each with different values of k (at 0.01 increments) in order to find the value that resulted in the lowest error score. This was 0.85 for observer SAW for both uncertainty and transducer versions of the simulation. This value means that the 2AFC noise has a standard deviation that is 0.85 less than the single interval noise. The error scores were 34.4 for the uncertainty version and 33.5 for the transducer version.

For PRM, the lowest error score (60.3) was obtained with $k = 0.92$ for the uncertainty version and with $k = 0.85$ for the transducer version (error score = 61.9). In order to find the optimum single value of k that applies to both versions of the model, the error scores were summed across models version for each value of k. The lowest resulting score (123.1) was obtained with $k = 0.89$. All further runs of the simulation incorporated $k = 0.85$ for SAW and $k = 0.89$ for PRM. The resulting best-fitting (lowest error score) values of the parameters are shown in table 6.7. For each observer, the error scores

Observer	SAW	SAW	PRM	PRM
Model version	transducer	uncertainty	transducer	uncertainty
Error score	33.4	34.4	61.9	60.3
No. of irrelevant channels	(0)	1	(0)	15
Transducer exponent	1.20	(1)	1.70	(1)
Noise sd	1.35	1.10	4.30	1.45
2AFC multiplier	0.85	0.85	0.89	0.89

Table 6.7: Best-fitting parameter values for the Monte Carlo simulation. Brackets indicate a fixed parameter.

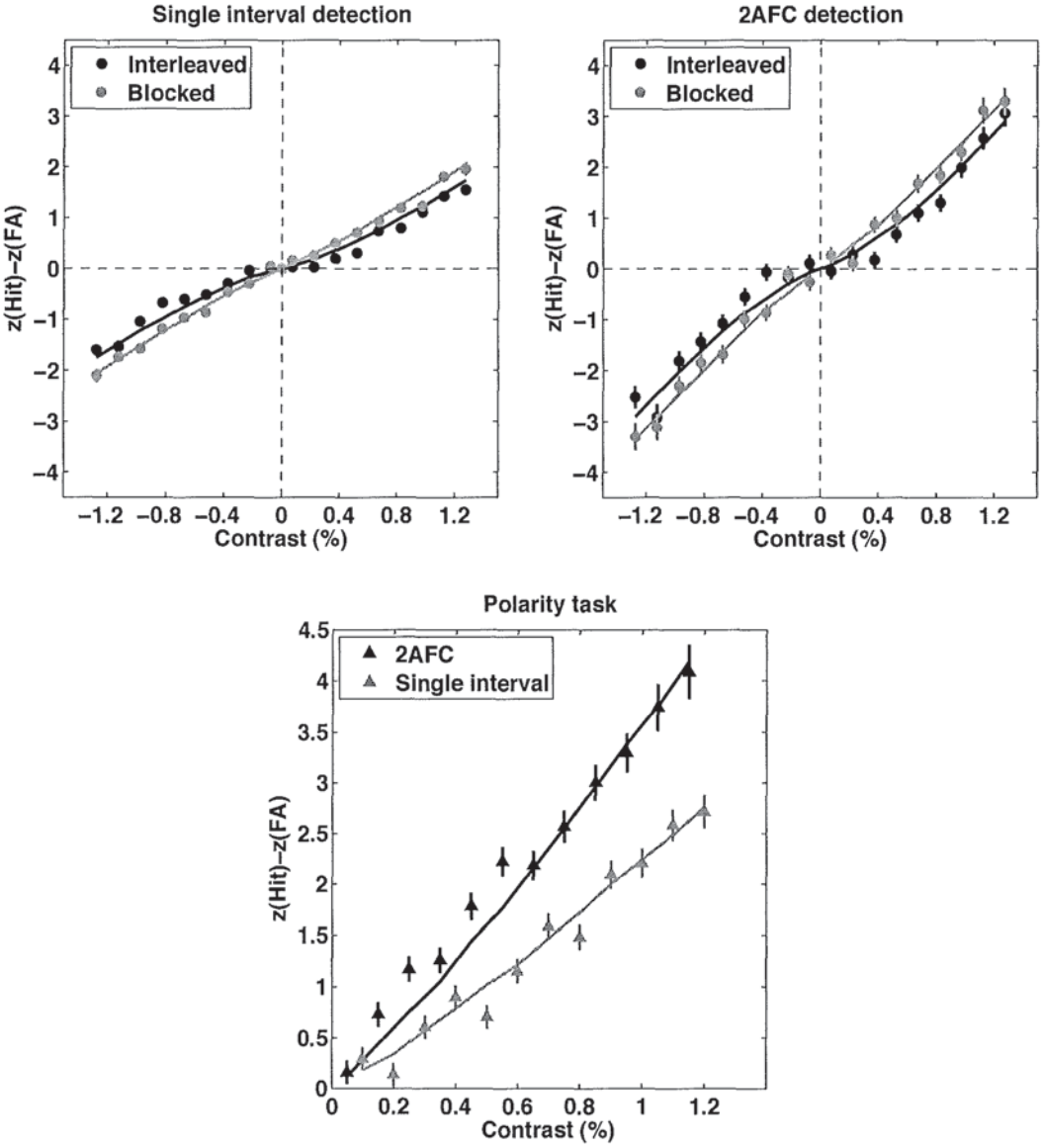

Figure 6.14: Data (symbols) and uncertainty simulation predictions (lines) for observer SAW. Negative and positive contrast levels indicate dark and light stimuli respectively. Error bars ($\pm 1sd$) are derived from the expected variance of each raw z-score.

show that the transducer and uncertainty models fit equally well. When the transducer version of the model was used, the fitted transducer exponent was nearly linear for SAW (1.2), but relatively nonlinear for PRM (1.7). When the uncertainty version of the model was used, the same difference between observers was captured by a PRM's greater uncertainty (PRM: 15 irrelevant channels; SAW: 1 irrelevant channel; table 6.7).

The close fit of the uncertainty simulation's predictions to all of SAW's experimental data is shown in figure 6.14. The predictions from the transducer version of the simulation were practically identical, so are not shown separately. The top 2 panels show the detection data for single interval (left) and 2AFC (right). The interleaved and blocked data are in different shades. Negative and positive contrast levels indicate dark and light stimuli respectively. Although the blocked data are all shown in grey, each of

these 2 panels contains 2 sets of blocked data; blocked dark and blocked light (left side and right side of each plot). The z-scores are slightly lower for interleaved data than blocked data at almost every contrast level. This small but systematic effect of extrinsic, stimulus uncertainty is also captured nicely by the simulation. Error bars ($\pm 1sd$) are derived from the expected variance of each raw z-score. They were calculated as follows.

The expected variance of each z(Hit) and z(FA) was obtained by application of the approximation provided by Gourevitch & Galanter (1967), which is specified above (equation 6.5). Note that the variance reduces as the number of trials increases. At each contrast level, the 2 variances were summed, and then the square root was taken to obtain the expected standard deviation of z(Hit)-z(FA). The error bars at a given z-score are smaller for single interval detection than 2AFC detection, because of the much larger number of trials designated as non-signal trials (interleaved single interval = 4320; interleaved 2AFC = 240). On these linear axes, linear performance is indicated by the degree to which the data fall on a straight line through the origin. Linear performance is evident for SAW's blocked detection data, especially for single interval. The interleaved data are less linear.

The lower plot in figure 6.14 shows SAW's polarity identification data for both single interval and 2AFC. Note that the scale used on the z-score axis is larger than in the upper plots. Both polarity identification data sets suggest linear performance. The uncertainty version of the simulation again provides a good fit to the data.

All of observer PRM's data, with uncertainty simulation predictions, are shown in figure 6.15. The same format as figure 6.14 is used here. The predictions from the transducer version of the simulation were practically identical, so are not shown separately. The two detection plots show little or no difference between the blocked and interleaved data. This is accurately captured by the simulation predictions. The central kink indicates nonlinear performance, which is consistent with the fitted transducer exponent of 1.7 or with a higher level of intrinsic uncertainty (table 6.7).

The lower panel shows polarity identification z-scores. Performance is non-linearly related to contrast. But the simulation predicts z-scores that are too high when the contrast is above 1%. This is the main reason why PRM's error scores (62, 61) were higher than SAW's (33, 34). It appears from figure 6.15 that PRM's polarity identification was somehow impaired relative to her detection performance. During data collection, observer PRM complained of negative after-images, although SAW did not. This may have impaired her polarity identification performance, because both polarities would be perceived when only one was presented. Such after-images would not, however, be expected to impair detection performance, because that does not require polarity information to be used.

The simulation was revised to incorporate a negative after-image factor (ai) by introducing a proportion of the signal into the non-signal polarity channel. Repeated runs of the simulation were performed, each at a given value of ai (increment = 0.01), in order to determine the value that provided the lowest error score, with all other parameters fixed as in table 6.7. The best fit was provided when $ai = 0.45$ for both transducer and uncertainty versions of the simulation. This means that 45% of the signal's amplitude was passed to the nonsignal polarity channel. The resulting error score was 43, which is a large reduction from the previous value of 63. The new predictions are shown, with the experimental data, in figure 6.16. A comparison between this figure and figure 6.15 shows the improved fit of the polarity identification predictions, which is entirely due to the addition of the negative after-image factor to the simulation. This improvement has not been accompanied by an impairment of the fits to the detection data. The detection predictions for the interleaved and blocked conditions are now

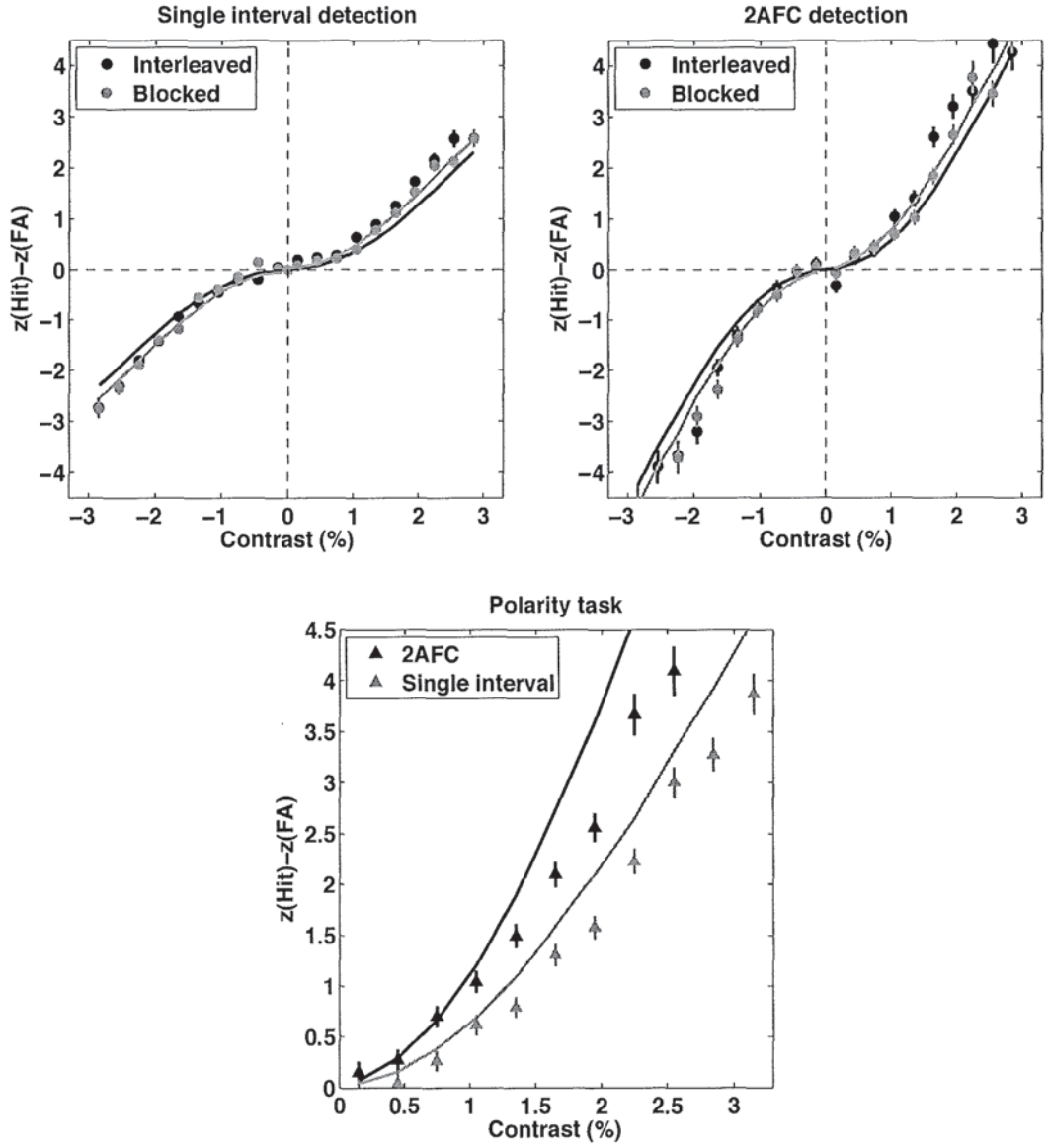

Figure 6.15: Data (symbols) and uncertainty simulation predictions (lines) for observer PRM. Negative and positive contrast levels indicate dark and light stimuli respectively. Error bars ($\pm 1sd$) are derived form the expected variance of each raw z-score.

slightly closer together than in figure 6.15, and are now almost superimposed. This more closely matches the pattern of the data in these conditions.

6.4.5 Discussion

Using a vertical Gaussian bar stimulus, psychometric functions were obtained for 4 tasks: detection and polarity identification, with either single interval or 2AFC presentation. The psychometric functions were sampled very intensively (9 to 18 contrast levels at 240 or 300 trials each), which allowed examination of the effect of subtle manipulations, such as the switch from blocked to interleaved trials. It also allowed a close examination of the relationship between different tasks.

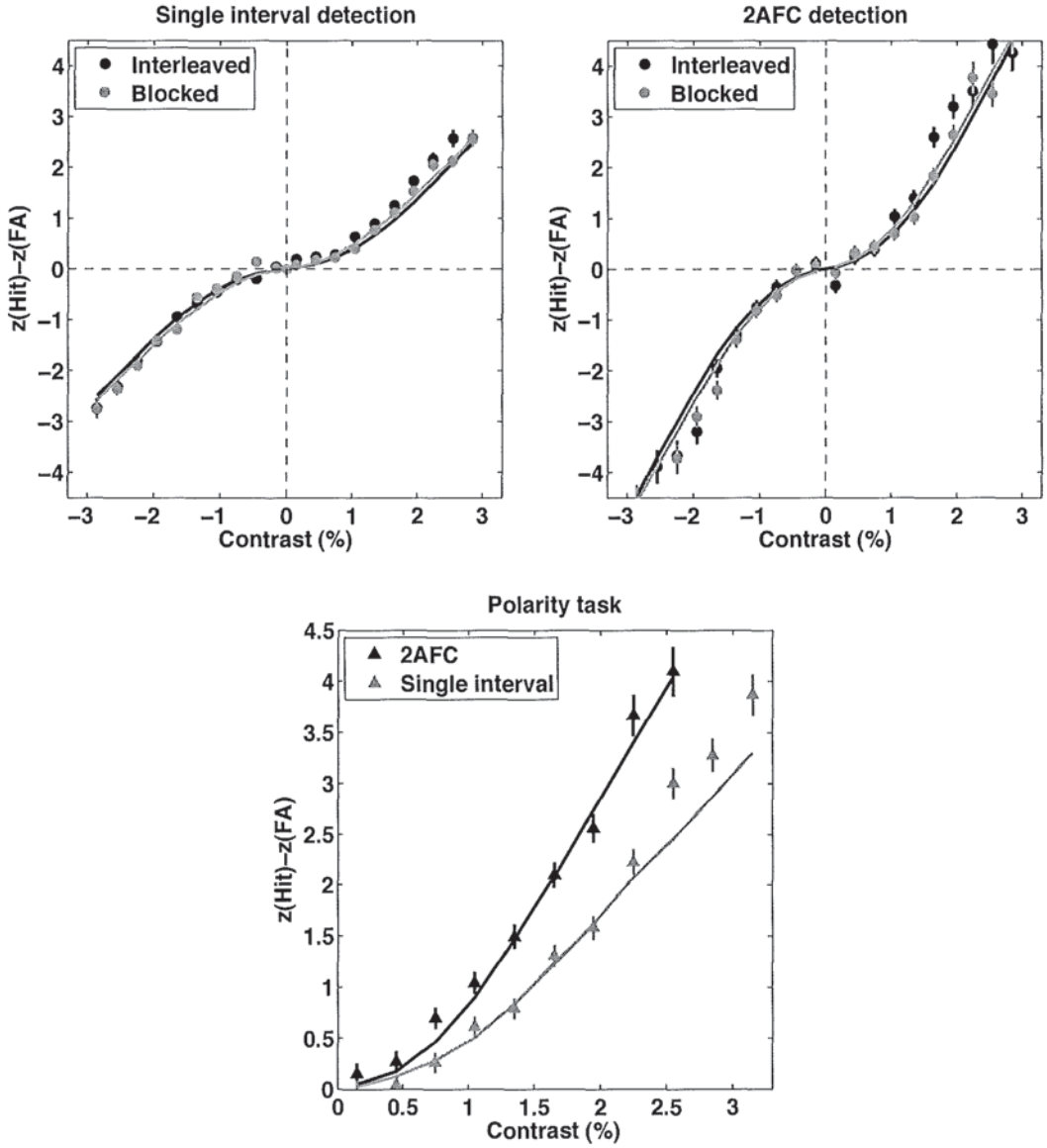

Figure 6.16: Data (symbols) and simulation predictions with negative afterimage (lines) for observer PRM ($ai = 0.45$). Negative and positive contrast levels indicate dark and light stimuli respectively. Error bars ($\pm 1sd$) are derived form the expected variance of each raw z-score.

The first aim of this experiment was to determine whether the observers' decision space contains separate independent channels for dark and light (monopolar model), or a single dark-light channel (bipolar model). The monopolar model predicts a $\sqrt{2}$ increase in performance when detection is replaced by polarity identification, whereas the bipolar model predicts the increase to be factor of 2. An initial examination of the experimental data (figure 6.12) showed that polarity identification z-scores were higher than detection z-scores by a factor that was closer to a $\sqrt{2}$ than 2. This supports the monopolar model. However, much of PRM's polarity identification performance was little different from her detection performance, and some of SAW's polarity identification data fell above the expected level.

The plots of polarity identification z-scores vs detection z-scores (figure 6.12) showed an improvement

in performance when the task was polarity identification. The multiplier applied to the detection data was 1.47 for SAW, which is close to the value of 1.41 that is expected from signal detection theory when monopolar channels are used. This provides support for the monopolar model, and the assumptions made in signal detection theory, such as a common standard deviation of noise in both detection and polarity identification tasks. Construction of a bipolar version of the simulation was not attempted because the results clearly favour the monopolar approach.

The multiplier applied to the detection data was only 1.10 for PRM, which demonstrates that her polarity discrimination performance was poorer than expected. This is not necessarily evidence against the monopolar model, and certainly does not favour the bipolar channel which predicts a ratio of 2. The Monte Carlo simulation (which assumed monopolar channels) provided a very good fit to her data when the after-image factor was introduced. Its introduction improved the fit to her polarity identification data without worsening the fit elsewhere. It was not needed with SAW's data, as the simulation provided a very good fit to his data without it. This difference between observers is consistent with their subjective reports; PRM complained of after-images, whereas SAW did not. It may be that SAW's greater psychophysical experience enabled him to ignore the after-images that PRM found so distracting.

The Monte Carlo simulation provided a very good fit to all of the data, separately for each observer. These data arise from a large and diverse selection of methods. If the 3 different conditions for detection (blocked light, blocked dark, interleaved) are counted as separate methods, there are 8 different methods in total. For each observer, the same parameter values were applied to fit the data of all 8 methods. The sheer size of each data set (SAW: 29,760 trials; PRM: 37,800 trials) lends strength to the conclusions drawn.

The second aim of this experiment was to examine whether the transducer or uncertainty models would be favoured by such a comprehensive set of data. The transducer and uncertainty versions of the simulation fitted the data equally well, so neither model can be rejected. This result is not entirely unexpected. When the uncertainty model was first applied to 2AFC contrast detection (Pelli, 1985), it was suggested as an alternative account to the nonlinear transducer model. Each value of transducer exponent could be mapped onto a specific number of irrelevant channels; the higher the transducer exponent, the higher the number of irrelevant channels. The transducer model applies a power function to the signal before noise is added; hence d' is a power function of contrast. The uncertainty model, in which the observer monitors irrelevant or non-informative channels, also leads to a very similar nonlinear relation between d' and contrast.

Given the equivalent predictions of the transducer and uncertainty models, the question of which one is more likely to fit the present experimental data may appear redundant. Yet this was one experimental aim. However, prior to the simulation, it was unclear how well each model would capture the effect of blocking vs interleaved trials. Furthermore, it was not clear how well the models would fit such a comprehensive set of data. Although both simulations fitted equally well, the uncertainty model may be regarded as more parsimonious, because differences in uncertainty between observers and between conditions can account for the data in a unified fashion. The transducer model, however, must appeal to both transducer nonlinearity and stimulus uncertainty to account for performance in blocked vs interleaved trials.

The uncertainty and transducer versions of the simulation are identical when the transducer exponent is 1 and the number of irrelevant channels is zero. This is a special case common to both models,

and represents linear transduction combined with no uncertainty. Observer SAW is very close to this very particular model ($p = 1.2$, or irrelevant channels = 1), although PRM is not ($p = 1.75$, or irrelevant channels = 15). But SAW had many more hours of previous practice than PRM. Perhaps intensive practice decreases the number of irrelevant channels monitored. The number of irrelevant channels obtained here are lower than those estimated by Pelli (1985) when fitting the data of previous researchers with an uncertainty model. He fitted the 2AFC grating detection data of Nachmias & Sansbury (1974) and Legge (1978b) by using 800 and 17 irrelevant channels respectively. He also fitted the data from 6 different conditions of a contrast detection rating-scale experiment (Nachmias, 1981) by using 10 - 1000000 irrelevant channels. Furthermore, Tanner (1961) concluded that 100 irrelevant channels would account for the steepness of his luminance discrimination psychometric functions.

This experiment also examined the relationship between single interval and 2AFC performance. Signal detection theory predicts that 2AFC performance should be $\sqrt{2}$ better than single interval. The present experimental data (figure 6.12) showed that the multiplier was 1.7 rather than 1.4, which indicates poorer single interval performance than expected. Jesteadt & Sims (1975) also found relatively poor single interval performance for frequency discrimination of a pure tone, and proposed that it may arise because of the greater memory load of the single interval procedure. In single interval, the observer is effectively comparing the stimulus with his memory of the alternative stimulus, which has not been seen since the previous trial (or longer). But in 2AFC, the comparison is made over a shorter time interval within one trial.

The simulation incorporated relatively poor single interval performance by reducing the noise by 11% or 15% in the 2AFC conditions. The good fit of the resulting predictions vindicate this manipulation, and show that the predictions of signal detection theory apply even in the presence of task-specific uncertainty. The uncertainty here could represent the greater memory load of the single interval procedure, which could manifest itself as criterion variance.

An other aim of this experiment was to examine the effect of using blocked or interleaved sets of stimuli. Observer SAW provided slightly better performance in the blocked than interleaved condition, and this was nicely predicted by the simulation. The effect of blocking may be due to the greater uncertainty in the interleaved condition, where the observer does not know whether to expect a light or dark stimulus. PRM showed little or no effect of blocking. But her data were fitted with a larger number of irrelevant channels (15), consistent with internal uncertainty. Perhaps her very small blocking effect was overwhelmed by the large internal uncertainty. On the uncertainty model, SAW had only 1 irrelevant channel, which suggests that having much less internal uncertainty did not overwhelm the influence of external uncertainty (the effect of interleaving light and dark trials).

The near-linear performance of SAW is a novel and very unusual finding. It was particularly striking in his data from single interval blocked detection, single interval polarity identification and 2AFC polarity identification. The data from these 3 conditions are re-plotted in figure 6.17, with the best-fitting linear function to each data set, constrained to pass through the origin. Each data set is accompanied by its correlation coefficient, which is high.

Such linear performance at these low contrast levels has not been reported in the literature. Rather, the exponent value is more typically reported as within the range of 2 to 4 (see table 6.1). At present, it is not clear why these researchers did not observe linear performance. The next experiment will examine the discrepancy between previous research and the linear performance reported here, and will compare contrast detection with contrast discrimination.

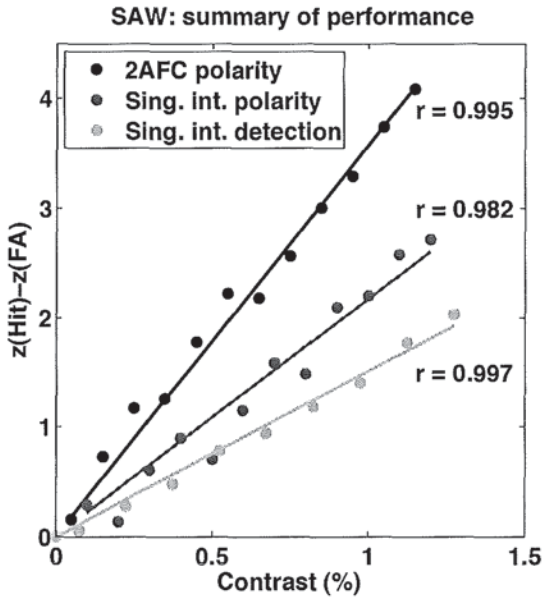

Figure 6.17: Z-scores from observer SAW (symbols) for 2AFC and single interval polarity identification and single interval blocked detection. Detection data is averaged across polarity. The lines are the best-fitting linear function to each data set. r is the correlation coefficient.

6.5 Experiment 11 - Contrast discrimination of a Gaussian bar

6.5.1 Introduction

In a contrast discrimination research, a low-contrast pedestal has been found to facilitate the detection of a target stimulus (Nachmias & Sansbury, 1974). In other words, the contrast increment required to discriminate 2 otherwise identical low-contrast stimuli was actually less than the stimulus detection threshold (see this chapter's introduction for a more detailed account). This facilitation was attributed to a transducer that is accelerating at low contrast, as shown in figure 6.18b. However, SAW's data have been fitted with a transducer that is nearly linear ($p = 1.2$). If the transducer was completely linear ($p = 1$), then no facilitation would occur at all. The line in figure 6.18b would have a constant gradient, and so Δc_1 and Δc_2 would have the same length. This means that the contrast increment required to discriminate 2 otherwise identical low-contrast stimuli would equal the stimulus detection threshold. In other words, the presence of the pedestal would have no effect on the amount of contrast required to detect the stimulus. In this situation, the discrimination-threshold function shown in figure 6.18a would therefore have no dip, and would instead be flat, at least in the low contrast range. For SAW, his near-linear performance would suggest that facilitation will barely be present. The present experiment aims to test this prediction, with the most commonly-used method for contrast discrimination (2AFC).

6.5.2 Method

The stimulus was the same vertical Gaussian bar ($\sigma = 12$mins) that was used in experiment 10, and was viewed under the same conditions. Contrast discrimination experiments typically use multiple levels of pedestal in different conditions, so that a full description can be obtained of how pedestal contrast

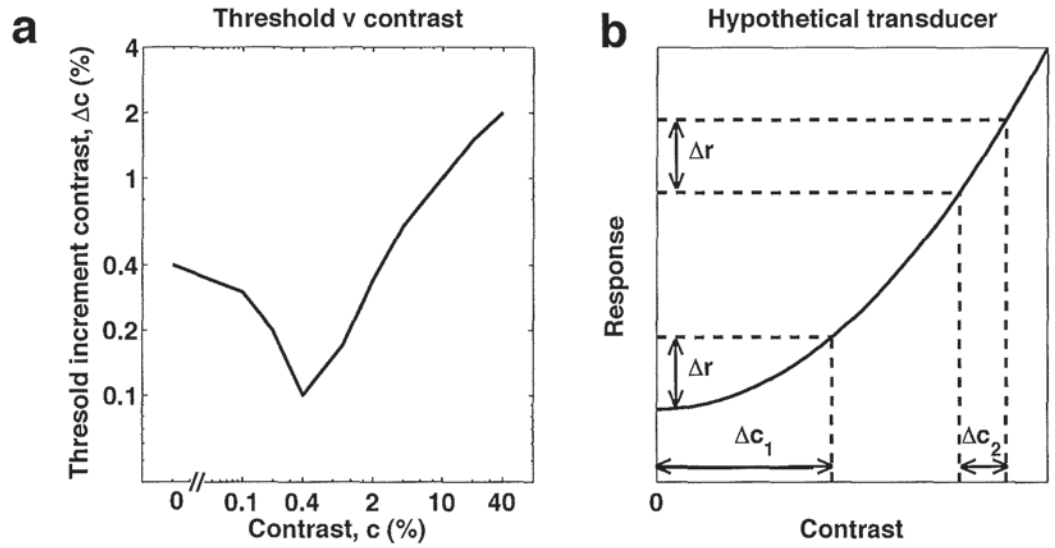

Figure 6.18: A copy of figure 6.1. a) Contrast discrimination function that shows threshold contrast increment as a function of pedestal contrast. b) Hypothetical transducer that transforms stimulus contrast into internal response (in the low-contrast range only). After Nachmias & Sansbury (1974).

affects the threshold increment contrast required for discrimination (figure 6.18a). However, the present experiment merely aims to determine whether facilitation occurs or not. Therefore, only 2 pedestal contrast levels were used here; zero and the contrast expected to cause maximum facilitation. Figure 6.18a shows that Nachmias & Sansbury (1974) found that maximum facilitation occurred when the pedestal was at the contrast equal to the stimulus' detection threshold. This finding is well established for gratings that are sinusoidal (Legge & Foley, 1980; Meese, 2004; Pelli, 1985), Gaussian derivatives (Bowen, 1995), or embedded in broadband noise (Henning & Wichmann, 2007) and also for Gaussian and rectangular bars (Legge & Kersten, 1983). SAW's detection threshold for the current Gaussian bars was obtained from his data in experiment 10. If threshold performance is defined as $d' = 1$, then SAW's threshold contrast level can be obtained by inspection of the top left panel of figure 6.11, which shows his blocked single interval detection data (averaged across polarity). By this method, the non-zero pedestal contrast level was set at 0.75%. Given the broad width of the dip in a typical dipper function, the exact contrast level chosen was not critical.

A 2 interval forced choice procedure, with method of constant stimuli, was used to determine contrast discrimination performance with the Gaussian bar. Both intervals contained the pedestal (Weber contrast = 0% or 0.75%), but only one interval also contained the target Gaussian bar. The bar was presented at a range of contrast levels, which had been selected from a pilot study. Its polarity always matched the pedestal polarity. In separate conditions, the bar and pedestal polarity was either always light (blocked light), always dark (blocked dark), or of either polarity (interleaved). Each of these 3 conditions was presented with each pedestal level, so there were 6 conditions in total.

There was only one observer (SAW), as only he had consistently shown near-linear performance. He was very experienced with psychophysical methods and with this set of images. The task was to indicate which interval appeared to be of higher contrast. When the pedestal contrast was zero, the task was detection, but when it was 0.75%, the task was contrast discrimination.

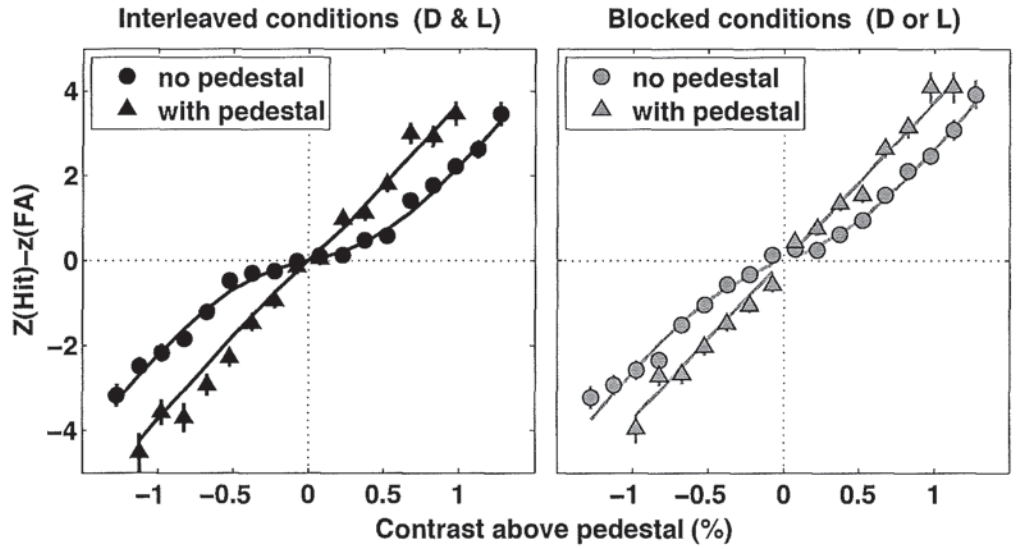

Figure 6.19: Z-scores of discrimination performance (symbols) as a function of contrast above pedestal, with simulation predictions (lines). Left and right panels show interleaved and blocked conditions respectively. Negative contrast refers to dark stimuli, and the sign of those z-scores has been inverted.

Model version	transducer	uncertainty
Error score	19	19
No. of irrelevant channels	(0)	5
Transducer exponent	1.40	(1)
Noise sd	1.10	0.75

Table 6.8: Best-fitting parameter values for the simulation predictions shown in figure 6.19. Brackets denote a fixed parameter.

A single block consisted of 60 presentations of every contrast level in the same condition. The observer completed 4 blocks for each condition. The order of blocks was randomised. The experiment took about $6\frac{1}{2}$ hours to complete.

6.5.3 Results

Following the procedure used in experiment 10, z-scores were corrected for bias. A Hit was defined as an 'interval 1' response to an interval 1 stimulus, and a False Alarm was defined as an 'interval 1' response to an interval 2 stimulus. Discrimination performance was taken as z(H) - z(FA). The resulting z-scores are plotted as a function of contrast above pedestal in figure 6.19. The left panel shows the 2 conditions where the dark and light stimuli were interleaved within every session. As before, negative contrast refers to dark stimuli and positive contrast refers to light stimuli. The z-scores for dark stimuli have had their sign inverted, so that in most cases they are now negative. Before this manipulation was performed the functions were U-shaped. Now, linear performance is indicated by data that form a straight line. The right panel shows the 4 conditions where stimulus polarity was blocked. One or two

data points at the highest contrast levels (both dark and light) are omitted from the 'with pedestal' data sets, because the responses were at 100%, which caused infinite z-scores.

The circles show the detection performance (pedestal absent), which now appears to be less linear than previously (*cf* figure 6.14, top right). The triangles show discrimination performance (pedestal present). The absolute z-scores are higher with the pedestal than without it, which suggests facilitation. The with-pedestal z-scores also appear more linear.

The Monte Carlo simulation that was used in experiment 10 (see section 6.4.3) was fitted to these data. The simulation was run exactly as before, except that the pedestal contrast was passed through the nonsignal interval, signal polarity channel, and the pedestal plus increment contrast was passed through the signal interval, signal polarity channel. The uncertainty and transducer versions of the model once again fit equally well. Best-fitting parameter values are shown in table 6.8 and the corresponding simulation predictions are shown as lines in figure 6.19.

The fitted transducer exponent is 1.40 and the number of irrelevant channels is 5. Both of these values suggest modestly nonlinear performance, the existence of which is consistent with the observed facilitation. In order to determine whether the same representation of signal and noise was used in the detection and discrimination tasks, the z-scores were combined in the manner specified by Foley & Legge (1981). They specify that the z-score of the discrimination of 2 contrast levels equals the difference between the z-scores for detection of each contrast level. This property is sometimes known as 'z additivity'. This is equivalently expressed, in the context of the present experiment, as that at each contrast level, the z-score for detection should equal the sum of the z-score for discrimination and the z-score for detection of the pedestal.

The z-scores for detection of a 0.75% contrast stimulus were obtained from the simulation fits, separately for the blocked data and interleaved data. Since the stimulus contrast levels did not include this particular value, the z-scores were calculated from the average of the detection ('no pedestal') z-scores for the stimuli of 0.675% and 0.825% contrast. This resulted in z-scores of 1.73 for the blocked condition and 1.39 for the interleaved condition. These values were added to the discrimination ('with pedestal') z-scores, and plotted as squares in figure 6.20. The detection data are shown as circles (directly copied from figure 6.19). Note that the abscissa now shows contrast, rather than contrast above pedestal. The left panel shows the interleaved data, and the right panel the blocked data. In both plots the 2 sets of data have 4 positive (light) and 4 negative (dark) overlapping contrast levels, which is shown by overlapping circular and square symbols. The close similarity between the z-scores from each data set shows the expected amount of z-additivity. This suggests that the same representation of signal and noise is used in the pedestal present and pedestal absent tasks.

The simulation is also able to provide a test of the degree of facilitation observed in this experiment. It can be used to fit only the detection data, and then the resulting parameters used to perform a fit to the discrimination data. If the latter fit is good, then support is provided to the hypothesis that amount of facilitation is consistent with the exact amount of non-linearity observed at low contrast levels. The simulation was fitted to the detection data alone. The resulting parameter values were identical to those obtained when the discrimination data was also included, which clearly supports the hypothesis.

In order to examine the effect of blocked vs interleaved trials, the data shown in figure 6.19 were re-arranged with different pairings of data, and re-plotted in figure 6.21. The left panel shows the

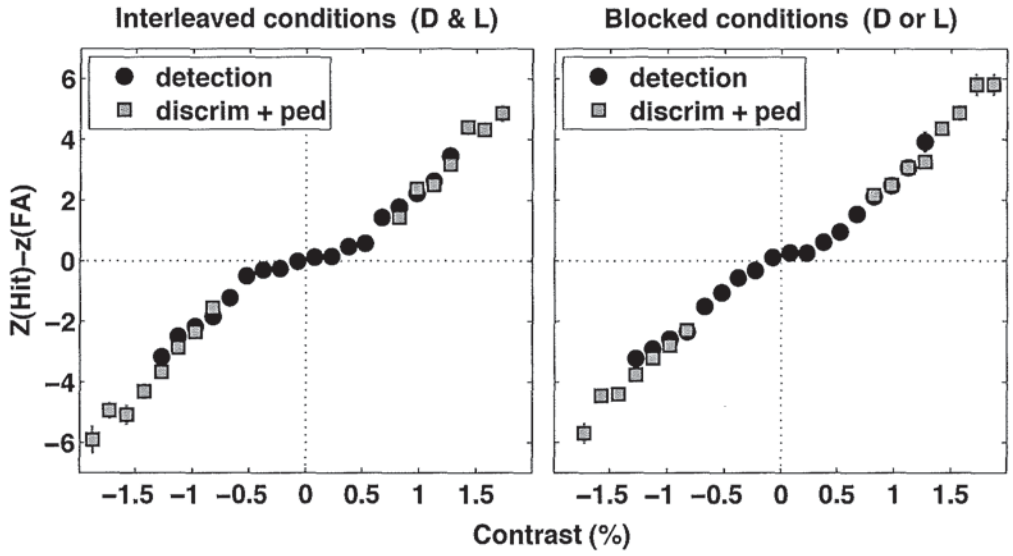

Figure 6.20: Z-scores of detection as a function of stimulus contrast, plotted with z-scores of discrimination summed with the z-score of detection of the pedestal.

no pedestal conditions. Switching from blocked to interleaved trials has caused a small increase in the absolute z-scores. The right panel shows the conditions with pedestal. Switching from blocked to interleaved trials has little or no effect here. The simulation predictions capture both of these patterns in the data. In the right panel, the same predictions are generated for both data sets.

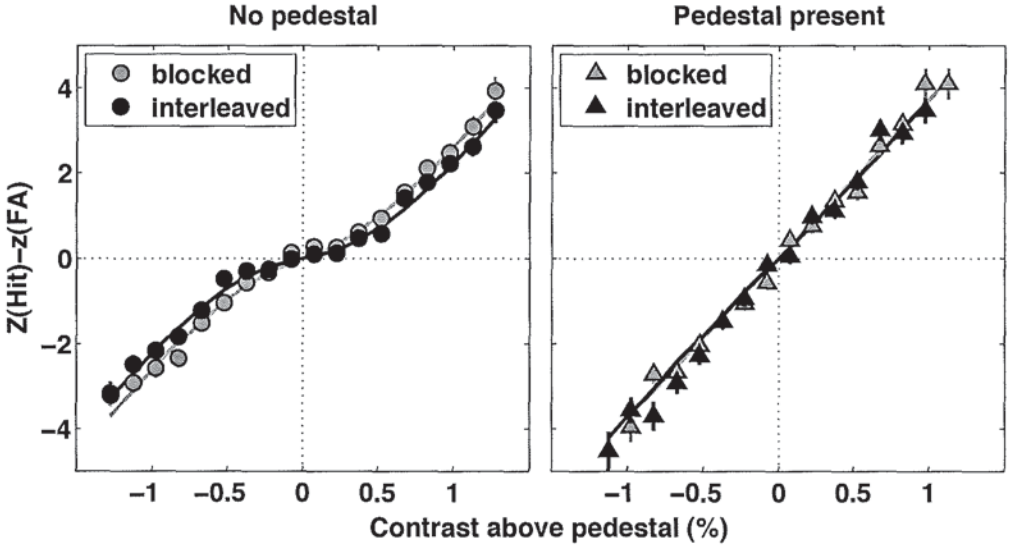

Figure 6.21: The data (symbols) and simulation predictions (lines) of figure 6.19, re-arranged here to show the effect of blocked vs interleaved trials. In the right panel, the same predictions are generated for both data sets.

6.5.4 Discussion

The aim of this experiment was to examine whether SAW's near-linear performance would lead to a near absence of facilitation in low-contrast discrimination. When the transducer version of the simulation was fitted to the data, the resulting exponent value (1.4) was similar to the value of 1.3 obtained in experiment 10. When the uncertainty version of the simulation was used, the resulting number of irrelevant channels was 5, whereas it was only 1 in experiment 10. This shows a very small decrease in the linearity of SAW's performance. But he is still more linear than typically observed in the literature. Nevertheless, the no pedestal data shown in figure 6.19 clearly show a central kink, which is consistent with nonlinear performance.

When the pedestal was added, the z-scores increased, which indicates facilitation. Furthermore, the data became more linear, which is consistent with performance being driven up a nonlinear transducer function, from an accelerating region to a linear region. But it is also consistent with a reduction in stimulus uncertainty, because the pedestal would reduce uncertainty about stimulus spatial and temporal extent (and polarity in the interleaved condition). Thus once again the data do not distinguish between the transducer and uncertainty models.

The amount of facilitation was found to be consistent with the pattern of detection performance (figure 6.20), which suggests that detection and discrimination are tapping into a common mechanism. It also supports the assumptions behind z-additivity, such as a constant standard deviation of noise across contrast levels; therefore precluding the possibility that the nonlinear nature of the z-score functions arise purely from a progressive decrease in noise variance (Foley & Legge, 1981).

When the data were re-plotted to show the effect of blocked vs interleaved trials (figure 6.21), it was seen that in the detection ('without pedestal') data, polarity uncertainty slightly decreased performance. This is the same pattern that was seen in experiment 10, and again suggests an effect due to small change in stimulus uncertainty. In the discrimination ('with pedestal') data, polarity uncertainty with pedestal did not change performance. This shows that at these higher contrast levels, the switch from blocked to interleaved trials did not increase stimulus uncertainty. However, the transducer version of the simulation also predicted this blocking effect, because it employed 2 polarity-specific channels. Nevertheless, the existence of the blocking effect suggests that detection was not based on an energy-based decision variable (Morrone & Burr, 1988), because the process of squaring and summing loses all polarity-specific information. Thus an energy model would predict identical responses for blocked and interleaved conditions.

To summarise, a monopolar model that assumed a simple non-linear transducer, or a linear transducer and several noisy channels, was able to provide a good fit to the discrimination data, while simultaneously fitting the detection data. The model correctly captured the small blocking effect in the detection data and its disappearance when the pedestal was added. As in experiment 10, the close fit of the model predictions to the data suggests that the model's assumptions are correct (such as the assumption of monopolar channels). The generality of that finding is now extended to the discrimination paradigm used here.

6.6 Experiment 12.1 - A pilot to examine the effect of a noise mask

6.6.1 Introduction

The previous 2 experiments used a Gaussian bar as the target stimulus. The intensively sampled psychometric functions for a range of tasks and conditions were simultaneously well-fitted by predictions from a Monte Carlo simulation. The success of the simulation suggested that it captures some of the processing that is undertaken by the human visual system at very low contrast. However, a bar is unlikely to be the only feature in a 'primitive' representation of a visual scene. Marr (1982) also proposed the features of edge, blob and termination. The dominant theme in the literature is that edges and bars are the 2 prime candidates for visual primitives (Georgeson & Freeman, 1997; Georgeson, 2006; Georgeson et al., 2007; Hesse & Georgeson, 2005; Tolhurst & Dealy, 1975; Watt & Morgan, 1985) although some work has suggested a single prime candidate of bars (Huang et al., 2006) or edges (Stromeyer & Klein, 1974).

The final aim of the present series of experiments (12.1 to 12.3) is to determine whether the highly lawful pattern of Gaussian bar data (experiment 10) could be repeated with an edge stimulus. In other words, whether polarity identification performance would be $\sqrt{2}$ better than detection, whether 2AFC performance would be about 1.7 times better than single interval, and whether blocked polarity stimuli would lead to slightly better performance than interleaved polarity stimuli. An additional aim was to determine whether the data would be fitted equally well with the transducer and uncertainty versions of the Monte Carlo simulation that was applied to the bar data in experiment 10. In other words, whether the best-fitting model would be one that used a non-linear transducer with just 2 (blocked polarity conditions) or 4 (interleaved polarity conditions) channels, or one that employed linear transduction with several noisy channels.

However, the use of an edge stimulus throws up the interesting problem of how to display an isolated edge. If the stimulus is simply a region of light on one side and a region of dark on the other, then there will also be unwanted stimulus cues at the borders of the stimulus and its background. This pilot experiment explores some different possible solutions to this problem, and aims to arrive at an optimal stimulus design.

Previous researchers have adopted different strategies to preclude border effects. Barbieri (2000) used luminance ramp stimuli with central 'blips' of gradient that sometimes caused perceived edges. To prevent the occurrence of Mach bands at the ends of the ramps, the stimuli were surrounded by binary noise of element size 1.7 × 1.7mins. Similar binary noise was also used by Hesse & Georgeson (2005) and May (2003) to mask stimulus border edges.

Tolhurst (1972b) limited the horizontal extent of a vertical edge stimulus by employing an image similar to those that generate the Craik-Cornsweet illusion (Cornsweet, 1970). The reason that this stimulus was employed was not to limit stimulus border edges, but rather to limit the role of differential light adaptation over different parts of the retina. Nevertheless, the borders at each side of the stimulus were at background luminance, so precluding edge polarity cues here. A later variation of this design was a vertical step edge enveloped by a horizontal Gaussian, which allowed greater control of the stimulus width (Shapley & Tolhurst, 1973).

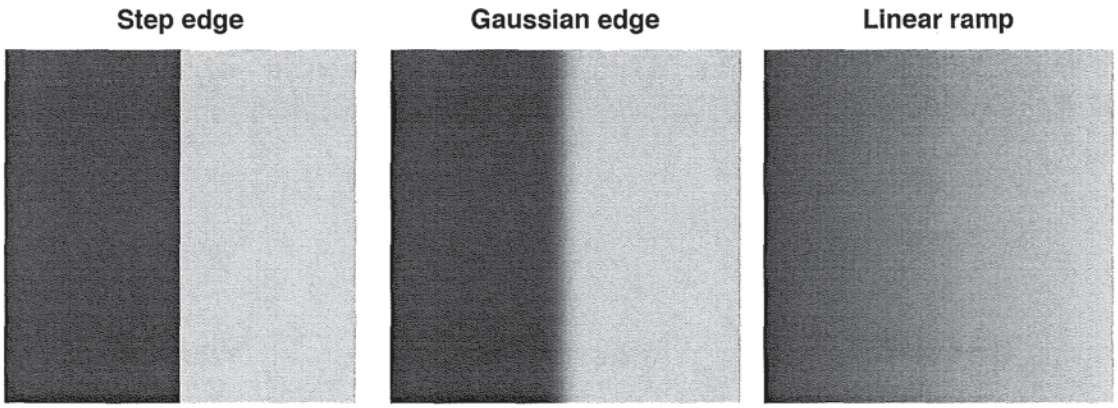

Figure 6.22: High contrast example of the dark-light target images.

A similar approach was taken by Huang et al. (2006). Their edge stimuli also resembled a relatively sharp edge enveloped by a Gaussian, but the Gaussian envelope was also applied vertically so that all stimulus borders were smooth. Unfortunately, the horizontal fall-offs of the vertical edge were so steep that the edges resembled a dark bar adjacent to a light bar, which perhaps accounts for their conclusion that human vision uses only cosine phase (bar) mechanisms. Nevertheless, the use of an isotropic Gaussian envelope provides a possible solution to the problem of stimulus border artifacts.

In an experiment that examined probability summation of two separate edges, Meese & Williams (2000) used an edge stimulus that was windowed by a circular function consisting of a central plateau surrounded by a sigmoidal ramp. Thus the outer borders of the stimulus faded gradually to background luminance. Nachmias (1981) used a 'bipartite field, which was a full-screen presentation of an edge that was light on one side and dark on the other. While this pushed the outer border of the stimulus away from the fovea, the outer border of the light region would still be visible to observers.

The aim of this experiment was to examine the efficacy of a binary noise mask at preventing the observer from using stimulus border cues when the intended stimulus is a central edge. The following experiment will explore the use of a smooth envelope.

6.6.2 Method

A method of constant stimuli was used with a 2AFC detection paradigm. The stimulus consisted of a target image and a binary noise mask. The target image was either a step edge, a Gaussian edge or a linear luminance ramp. High contrast examples of dark-light stimuli are shown in figure 6.22. Light-dark stimuli were also employed. The Gaussian edge was the integral of the Gaussian function whose σ was 5.4mins. These images formed a coherent set that ranged from a step edge, through a blurred edge, to the most blurred edge possible (linear ramp). The target image subtended 3.48×3.48 degrees at the viewing distance of 148cm and was displayed at the centre of the monitor screen. It had 6 contrast levels in a logarithmic progression between 0.125% and 4% (Michelson contrast) and 2 polarities (dark-light or light-dark). The noise mask was a central 6.95×6.95 degrees display of randomly arranged binary elements of size 4×4 pixels (3.25×3.25 mins). It had the 2 contrast levels

of 0% (no mask) or 32%. When the target image and noise mask were presented simultaneously, the noise was only in a square annulus surrounding the target.

Each trial consisted of two 300ms intervals separated by 600ms. Each interval contained either the target stimulus or a blank region at mid-grey luminance, surrounded by the noise mask, and was indicated by a 300ms duration tone. Each interval was flanked by 150ms presentations of the full noise mask. Thus the sequence of events was: 150ms noise, 300ms interval 1 stimulus surrounded by noise border, 150ms noise, 300ms blank, 150ms noise, 300ms interval 2 stimulus surrounded by noise border, 150ms noise. This sequence is illustrated in figure 6.23 for an interval 1 Gaussian edge. It shows the time in ms as the trial progresses. Each panel depicts the full monitor screen.

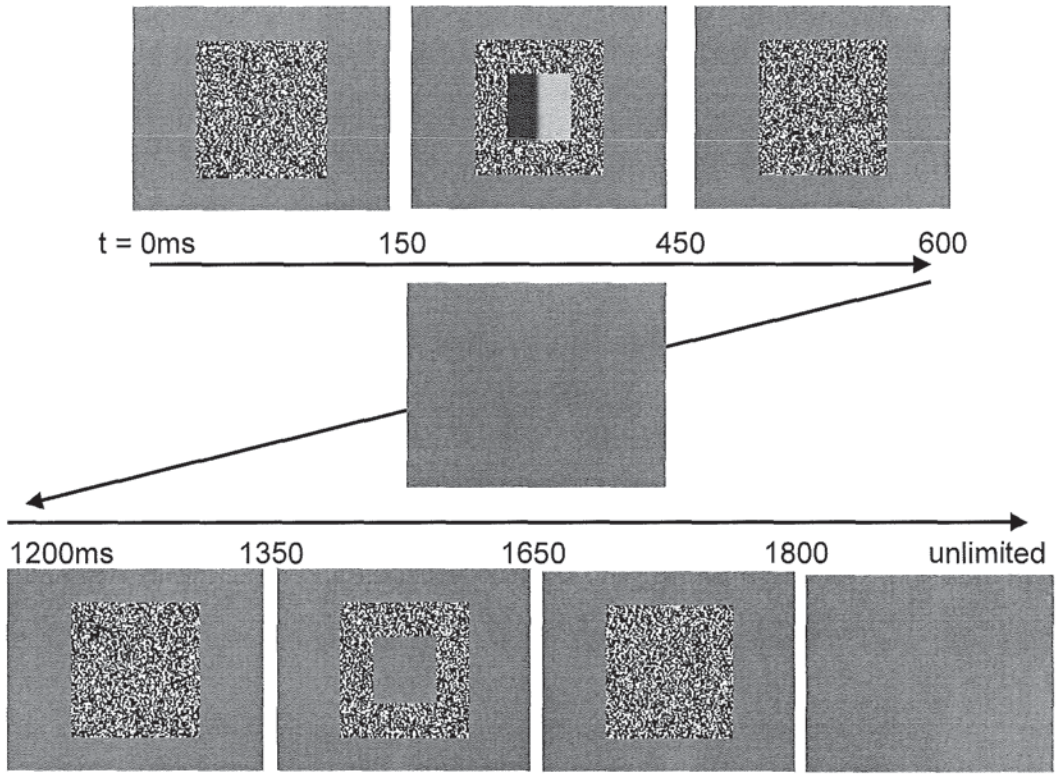

Figure 6.23: High contrast example of the sequence of events in a single trial, where the Gaussian edge was in interval 1. Each panel depicts the full monitor screen.

A different noise mask was randomly selected (from a bank of 50 masks) for each of the 6 events that contained noise. This sequence of events was used to mask any temporal luminance transients that would otherwise act as a cue. Thus the design used both spatial and temporal masking. A central black fixation spot (2×2 pixels) was displayed continuously except during the 2 observation intervals.

There was one observer. The task was to indicate the interval that contained the target image. The observer had unlimited time to respond. Immediate auditory feedback was given after each response. The interval between trials was at least 1 second. One session consisted of 20 presentations of all of the images for a single condition in a randomised order. Each of the 6 conditions (3 target image types, 2 noise contrasts) was presented in a randomised order.

6.6.3 Results

For each condition, the percentage of correct responses were averaged across polarity and plotted as a function of target image contrast. These plots are shown in figure 6.24. Separate panels are for the

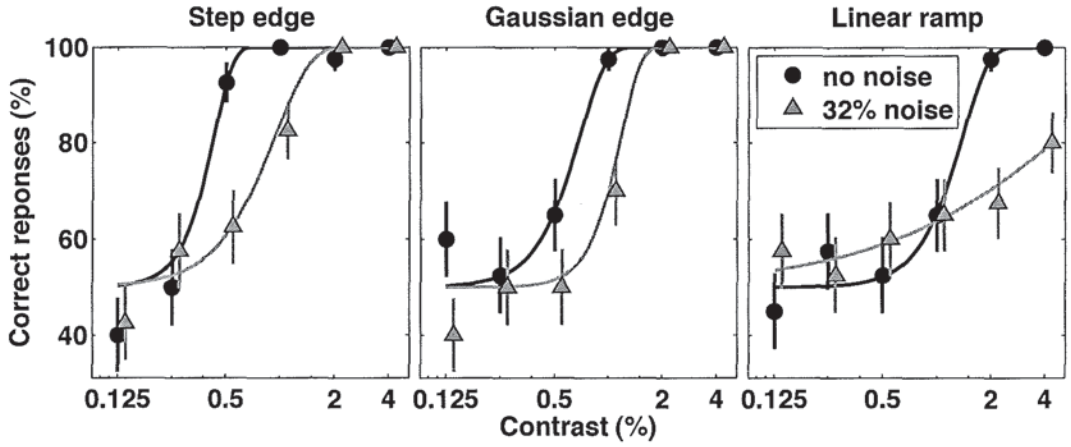

Figure 6.24: Detection performance averaged across polarity as a function of target image contrast. Separate panels are for separate image types. Error bars (±1sd) are derived from the expected variance in binomial sampling. For clarity, the 32% noise data have been shifted slightly rightwards. The lines are the best-fitting Weibull function of slope ≤ 4. Weibull parameters are shown in table 6.6.3.

image types of Gaussian edge, step edge and linear ramp. Error bars (±1sd) are derived from the expected variance in binomial sampling (equation 3.4) and are relatively long due to the small number of trials per datum (40). For clarity, the 32% noise data have been shifted slightly rightwards. The lines are the best-fitting Weibull function of slope ≤ 4.

A Weibull is a commonly-used function that has the formula

$$p_1 = 1 - exp\left(-\left(\frac{c}{\alpha}\right)^{\beta}\right) \tag{6.6}$$

where c is the stimulus contrast level, and α and β are parameters that denote the Weibull's horizontal position and its slope. The values of p_1 range from 0 to 1. But 2AFC functions range from 0.5 to 1. So the proportion (p_1) from equation 6.6 was then converted to this smaller range by

$$p = 0.5 + \frac{p_1}{2} \tag{6.7}$$

The value of α now indicates the contrast level where p_1 is 0.82, and α is sometimes termed the 'threshold'.

The ceiling on the permitted slope was imposed, because in its absence, two of the fitted functions had very large slopes (Gaussian edge with noise: slope = 155; step edge no noise: slope = 17). These steep slopes arose as a consequence of the sparse sampling of the psychometric functions. In both cases there was only one datum between the extremes of 50% and 100% correct responses. The ceiling was set at 4, because this is the typical maximum reported in the literature for 2AFC detection (Mayer & Tyler,

Image type	noise (%)	threshold, α (%)	slope, β
Step edge	0	0.43	4
	32	0.94	2.25
Gaussian edge	0	0.69	2.98
	32	1.18	4
Linear ramp	0	1.39	3.01
	32	5.00	0.71

Table 6.9: Values of α and β for the Weibull fits shown in figure 6.24.

1986; Nachmias, 1981; Pelli, 1985; Williams & Wilson, 1983).

The best-fitting values of α and β were obtained by using maximum likelihood estimation. This technique maximises the log likelihood of obtaining a particular set of data, assuming that it was derived from the Weibull, and was described in detail in section 2.5.1. The best-fitting values of α and β are shown in table 6.6.3.

Figure 6.24 shows that in the absence of the noise, performance is generally at about chance (50%) for all image types at or below 0.25% contrast. However performance is markedly improved as the image reaches a contrast level of 0.5% for the step edge, 1% for the Gaussian edge and 2% for the linear ramp. The addition of the noise mask has shifted the psychometric functions rightwards, which indicates raised thresholds for all three image types; a pattern also seen in table 6.6.3. The table and figure 6.24 also show that the reduction in the Weibull's slope (β), caused by switching from no noise to 32% noise, is marked for the linear ramp, but not for the other two image types.

6.6.4 Discussion

The aim of this experiment was to examine the efficacy of a binary noise mask at preventing the observer from using stimulus border cues. Of the 3 image types used, the linear ramp had the lowest luminance gradient, so was most likely to be detected by way of the high contrast edges of its border. The noise mask reduced β for this image much more than the step or Gaussian edges, which indicates a large loss of sensitivity. This change is consistent with the proposal that different mechanisms are involved in the no noise and 32% noise conditions for the linear ramp. In the no noise condition, the observer appears to have made substantial use of the image border, whereas in the 32% noise condition he would have only the luminance ramp itself. To conclude, the linear ramp data demonstrate the effectiveness of the noise at masking the image border.

However, if the noise acts only to mask the image border, then it would be expected to have little or no effect on detection of the step and Gaussian edges, because the contrast at the image centre (i.e. at fixation) was always twice the contrast at its border. This expectation was not met for the step edge, because the slope decreased by almost a factor of 2 when the noise was added. This suggests that a different mechanism underlies each condition, and therefore that the observer was using the image border in the no noise condition. It might be expected that the pattern of Gaussian edge data would fall somewhere between the step edge and linear ramp data patterns, because the Gaussian edge's blur falls between the very sharp step edge and the very blurred edge that is the linear ramp. However, the Gaussian edge slope increased (not decreased) when the noise was added, which shows that the noise improved sensitivity to changes in the edge's contrast. This unexpected finding may, however,

be an artifact of the undersampled psychometric functions. For all 3 image types, the noise did shift the psychometric functions rightwards. This suggests that the noise masked the target stimulus, and is consistent with the phenomenon of surround suppression (Cannon & Fullemkamp, 1991; Yu et al., 2002) and the subjective impression of the stimuli, which was that the noise mask was very distracting. The present experimental design does not allow the spatial location of the masking effect to be identified. In other words, it does not distinguish between masking of only the image border and masking across the whole image.

To conclude, although the noise mask appears to successfully mask the image border, it may also mask the image itself. This is a substantial drawback, because performance in the forthcoming main experiment (experiment 12.3) needs to be measured at very low contrast levels in order to determine whether performance is truly linear. Furthermore, experiment 12.3 is intended to be a simple extension of experiment 10, which replaces bar stimuli with edge stimuli. But the noise mask has considerably altered the stimuli's appearance. Given these reservations, an alternative method of reducing border cues is explored next.

6.7 Experiment 12.2 - A pilot to examine the effect of a temporal and spatial soft envelope

6.7.1 Introduction

Tolhurst (1972b) used edge stimuli similar to a Cornsweet edge in a detection and polarity identification experiment. His objection to the use of spatially extensive edge stimuli was due to the possibility of differential light adaptation over different parts of the retina. However, his concern may have been unwarranted, because performance at detecting the absolute luminance of a large uniform area is relatively poor. Knau (2000) found that when the luminance of a Ganzfeld (a full featureless field of even luminance) was slowly changed over time, the discrimination threshold was about 0.2 log units. To put this value in context, grating contrast discrimination thresholds can be as low as 0.04 log units (Nachmias & Sansbury, 1974).

Instead of assuming that the regions of high and low luminance that flank an edge confound edge detection, a better approach would be to measure the strength of the confounding effect under various conditions and then select only those conditions where the effect is minimal. This experiment aims to perform this measurement by comparing detection of a Gaussian edge with discrimination of a luminance change in a uniform field. If the detection threshold for the Gaussian edge is at a contrast level which only provides chance performance for the uniform field, then it can be assumed that edge detection does not rely upon the regions of high and low luminance that are remote from the edge itself.

A spatial soft window was applied to the edge stimulus, in order to prevent image border effects. A temporal envelope was also applied to the images, in order to prevent the observer's use of temporal transients (a sudden increase and decrease of luminance at the image's onset and offset). It might be thought that a half-cosine function would be optimal here, as it is entirely smooth. However, this intuition is based on the assumption that it is sharp 'corners' in the temporal function that must be avoided. Yet this assumption is not supported by the literature. A biphasic temporal response function

has been observed in the cat LGN (Cai et al., 1997) and macaque simple cells (De Valois & Cottaris, 1998), and this profile has received psychophysical support in humans (Burr & Morrone, 1993). A biphasic operator is one that has a peak of activity followed by a trough (or vice versa). In other words, it approximates a 1st derivative operator. This means that it will respond maximally to peaks of gradient, and not to the peaks of curvature that are sharp 'corners'. Therefore to minimise the temporal response of a stimulus, its temporal envelope should have the lowest gradient possible. This will be achieved with a linear ramp. If a half-cosine ramp and a linear ramp span the same luminance range, then the half-cosine will have a greater central gradient than the linear ramp, as shown in figure 6.25. This means that it may be expected to perform less well in preventing the perception of temporal transients, so leading to lower detection thresholds. This experiment aims to test this prediction.

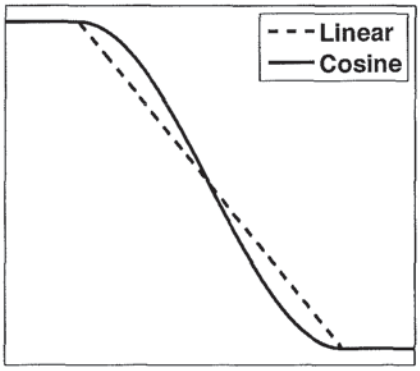

Figure 6.25: Linear and half-cosine ramps that span a common amplitude.

6.7.2 Method

The experiment used a 2AFC detection paradigm with method of constant stimuli. The stimulus was either a luminance plateau or a vertical Gaussian edge (σ= 5.4mins). Unlike the previous experiment, the image now spanned the full screen of the monitor (13.9 × 10.4 degrees). A spatial envelope was applied to the image to smooth its borders very gradually to zero contrast. The smoothing was applied with a half-cosine whose width was half the height of the screen. Thus for the luminance plateau, the luminance profile through the vertical mid-line was 1/2 period of a sinusoid. The profile through its horizontal mid-line was a plateau flanked by 2 half-cosines of same width as the vertical half-cosines. The Gaussian edge had mid-grey luminance at its centre, so its vertical luminance profile was 1/2 period of a sinusoid only to each side of centre, away from the transition from dark to light. High contrast examples of the images are shown in figure 6.26, with their horizontal mid-line luminance profiles. Each panel represents the whole monitor screen. Both image polarities were used, so a dark luminance plateau and a light-dark Gaussian edge were also part of the image set.

The temporal envelope consisted of a high plateau flanked by a pair of ramps that were both cosine or linear (figure 6.25). The ramps consisted of 8 levels of attenuation that ranged from zero (no contrast attenuation) to 1 (contrast reduced to zero) and therefore consisted of 7 steps. The 'off' ramp was a left-right transposed version of the 'on' ramp. Each temporal ramp duration was 165ms (14 frames) and the plateau duration was 450ms. Stimulus duration, defined from the mid-points of the on and off

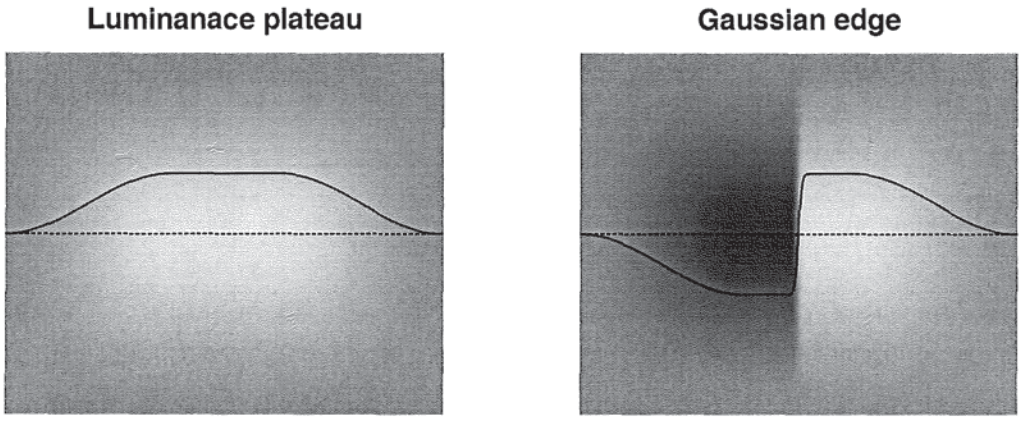

Figure 6.26: High contrast examples of the 2 types of image, with their horizontal mid-line luminance profiles. Each panel represents the whole monitor screen.

ramps, was 615ms.

There were 3 temporal conditions; the temporal soft window was either absent (stimulus has abrupt onset and offset) or had a half-cosine or linear ramps. Due to the evolution of the experimental design over time, these 3 conditions were presented in this order. The stimuli were presented at 6 Weber contrast levels that ranged from 0.125% to 4% in a logarithmic progression, for the no temporal ramp condition. Where the temporal ramp was present, the contrast levels were 0.25% to 8%.

A 2AFC procedure was used, whereby the target stimulus was presented in one interval, and a blank screen at mid-grey luminance was presented in the other interval, with an inter-stimulus interval of 600ms. There was one observer. The task was to indicate which interval contained the target stimulus. A single block consisted of all 6 contrast levels for both polarities of a given stimulus type (Gaussian edge or luminance plateau) and a given temporal ramp condition. A session consisted of 20 blocks, with the trials presented in a randomised order. There were 2 sessions for each image type, presented in an ABBA order, where A is the luminance plateau and B is the Gaussian edge.

6.7.3 Results

In a given condition, psychometric functions for each stimulus polarity were very similar, so results are presented pooled across polarity. Figure 6.27 shows the percentage of correct responses as a function of contrast, plotted separately for each temporal ramp condition (separate panels) and image type. Error bars ($\pm$1sd) are derived from the expected variance in binomial sampling. For clarity, the luminance plateau data have been shifted slightly rightwards. The lines are the best-fitting Weibull function. Weibull parameters are shown in table 6.10.

In each temporal ramp condition, the luminance plateau thresholds are higher than the Gaussian edge thresholds. The thresholds have increased by a factor of 3.6, 4.8 and 6.2 for the no ramp, cosine ramp and linear ramp conditions respectively.

The Gaussian edge threshold in the absence of the temporal ramp (0.69%) was slightly lower than

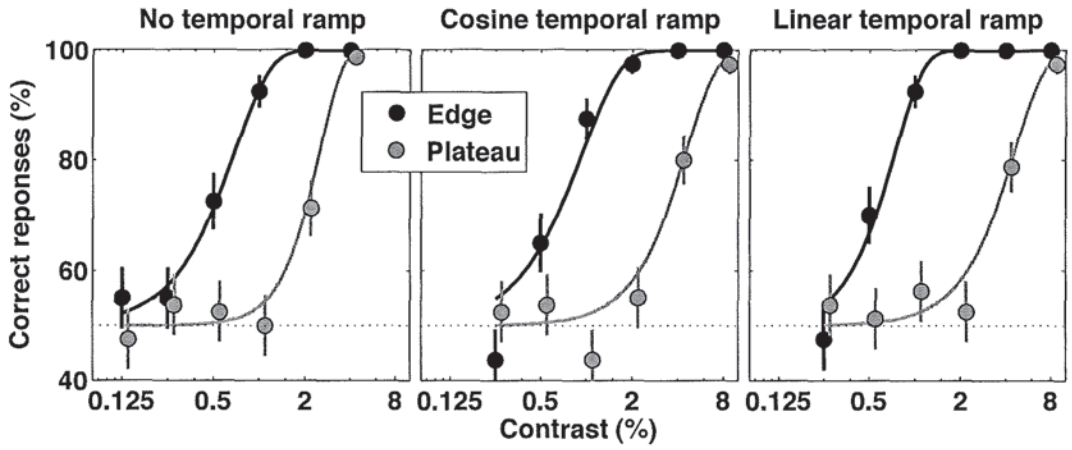

Figure 6.27: Detection performance averaged across polarity as a function of image contrast. Separate panels are for separate temporal ramp conditions. Error bars (±1sd) are derived from the expected variance in binomial sampling. For clarity, the luminance plateau data have been shifted slightly rightwards. The lines are the best-fitting Weibull function. Weibull parameters are shown in table 6.10.

Temporal ramp	image type	threshold, α (%)	slope, β
None	edge	0.69	1.81
	plateau	2.51	2.88
Cosine	edge	0.95	1.70
	plateau	4.56	2.12
Linear	edge	0.74	2.28
	plateau	4.60	2.07

Table 6.10: Values of α and β for the Weibull fits shown in figure 6.27.

thresholds when a ramp was present (0.95% and 0.74%). However, for the luminance plateau, the threshold in the absence of the temporal ramp (2.51%) was almost a factor of 2 smaller that the thresholds when a ramp was present (4.56% and 4.60%).

6.7.4 Discussion

Inspired by Tolhurst's (1972b) reluctance to use spatially extensive edge stimuli, one experimental aim was to determine the extent to which edge detection is governed by the regions of high and low luminance that are remote from the edge itself. In all three temporal conditions, the threshold for detection of the Gaussian edge was at a contrast level where performance for the luminance plateau was still at chance. This suggests that the edge was detected by the central rapid change in luminance of the edge itself, and not by using the adjacent regions of high and low luminance, and not by using the remote borders.

Both image types had lower detection thresholds in the absence of the temporal ramp than when it was present. This threshold difference suggests that luminance transients were providing a cue that was used by the observer. This conclusion is supported by the relatively high threshold difference for

the luminance plateau (compared to the Gaussian edge), because here the observer does not have the central edge cue, and therefore would rely on luminance transients more heavily.

The second aim of this experiment was to test the hypothesis that the cue of temporal transients would be progressively minimised as the conditions change from no ramp, to cosine ramp and finally linear ramp. If this hypothesis is true, then the detection thresholds should also progressively increase, due to the gradual removal of this cue. Although the luminance plateau thresholds do follow this predicted pattern, the Gaussian edge data do not. Their thresholds were 0.69%, 0.95% and 0.74%, for no ramp, cosine ramp and linear ramp respectively. This lack of a progressive increase may, however, be due to the lack of randomisation between the temporal conditions. All of the no ramp data were collected first, followed by all of the cosine ramp data, followed by all of the linear ramp data. Thus the low threshold in the linear ramp, Gaussian edge condition may represent a practice effect, whose absence would have lead to a threshold higher than in the cosine ramp condition.

Since order effects may be present, a more appropriate comparison can be made between the Gaussian edge and luminance plateau data within each temporal condition. The switch from Gaussian edge to luminance plateau caused an increase in threshold by a factor of 3.6, 4.8 and 6.2 for the no ramp, cosine ramp and linear ramp conditions respectively. To conclude, the majority of the data support the hypothesis that temporal transients would be progressively minimised as the conditions change from no ramp, to cosine ramp and finally linear ramp. Thus the linear temporal ramp is the favoured temporal window for the main experiment.

The present experiment used a spatial soft window and a threshold of 0.69% was obtained for the Gaussian edge without a temporal soft window. The previous experiment also obtained a threshold of 0.69% for the Gaussian edge that had an abrupt change to background luminance at its border. Notwithstanding a possible order effect, this similarity in thresholds suggests that the spatial soft window and the abrupt border had equivalent border effects, or more plausibly, that the observer was attending purely to the central edge. If the latter explanation is true, then the raised threshold caused by the noise mask in the previous experiment (1.18%) would appear to represent a masking effect of the noise on the central edge. To conclude, the spatial and temporal soft window is the preferred method of preventing image border cues.

6.8 Experiment 12.3 - 2AFC detection and polarity identification of a blurred edge, with a spatial and temporal soft window.

6.8.1 Introduction

The aim of the present experiment is to determine whether the highly lawful pattern of Gaussian bar data (experiment 10) could be repeated with an edge stimulus. In other words, whether polarity identification performance would be $\sqrt{2}$ better than detection, whether 2AFC performance would be about 1.7 times better than single interval, and whether blocked polarity stimuli would lead to slightly better performance than interleaved polarity stimuli. An additional aim was to determine whether the data would be fitted equally well with the transducer and uncertainty versions of the Monte Carlo simulation that was applied to the bar data in experiment 10. In other words, whether the best-fitting

model would be one that used a non-linear transducer with just 2 (blocked polarity conditions) or 4 (interleaved polarity conditions) channels, or one that employed linear transduction with several noisy channels.

The previous two pilot experiments have established that the best edge stimulus, of those considered, was a full screen image with a linear temporal soft window and a cosine spatial soft window. That stimulus will be used here. In order to reduce the number of hours of data collection to a manageable number (from 60 hours in experiment 10), single interval methods will not be used here. Furthermore, the psychometric functions will be sampled slightly less finely.

6.8.2 Method

The stimulus was the Gaussian edge used in the previous experiment that had a linear temporal soft window and a cosine spatial soft window. The edge was of polarity dark-light (DL) or light-dark (LD). A 2AFC procedure was applied with a method of constant stimuli. The tasks were polarity identification, or detection. The detection task had 3 conditions: blocked DL, blocked LD or interleaved polarity. All other methodological details were as experiment 10, except that a different second observer was used, who was very experienced psychophysically, but still had little prior practice with this particular image set. Furthermore, there were 240 presentations of every stimulus for each observer, which represented a total of 20 hours of data collection.

6.8.3 Results

Following the experiment 10 analysis, bias was corrected by obtaining z(H)-z(FA). As before, the definitions of 'hit' and 'false alarm' were:

2AFC polarity identification

Hit: '1st interval' response to light in 1st interval

False alarm: '1st interval' response to dark in 1st interval

2AFC detection

Hit: '1st interval' response to stimulus in 1st interval

False alarm: '1st interval' response to stimulus in 2nd interval

Data were fitted with the same Monte Carlo simulation described above. The '2AFC multiplier' parameter (k) was fixed at 0.85 and the after-image parameter (ai) was fixed at zero for both observers. These are the values that provided the best fit to SAW's data in the previous experiment. Since the whole experiment was 2AFC, the value of k serves only to reduce the value of the *noise* parameter to a level comparable with the previous experiment. Data and simulation predictions for both observers are shown in figure 6.28. Separate columns are for each observer. Upper row shows all detection data. Lower row shows all polarity discrimination data with blocked detection data averaged across polarity. The whole pattern of data is very similar to SAW's data in the previous experiment, which used a Gaussian bar stimulus. Both observers show a small increase in (absolute) z-scores for detection when interleaved presentation was replaced by blocked presentation. Furthermore, both observers show the expected improvement in performance when detection was replaced by polarity discrimination.

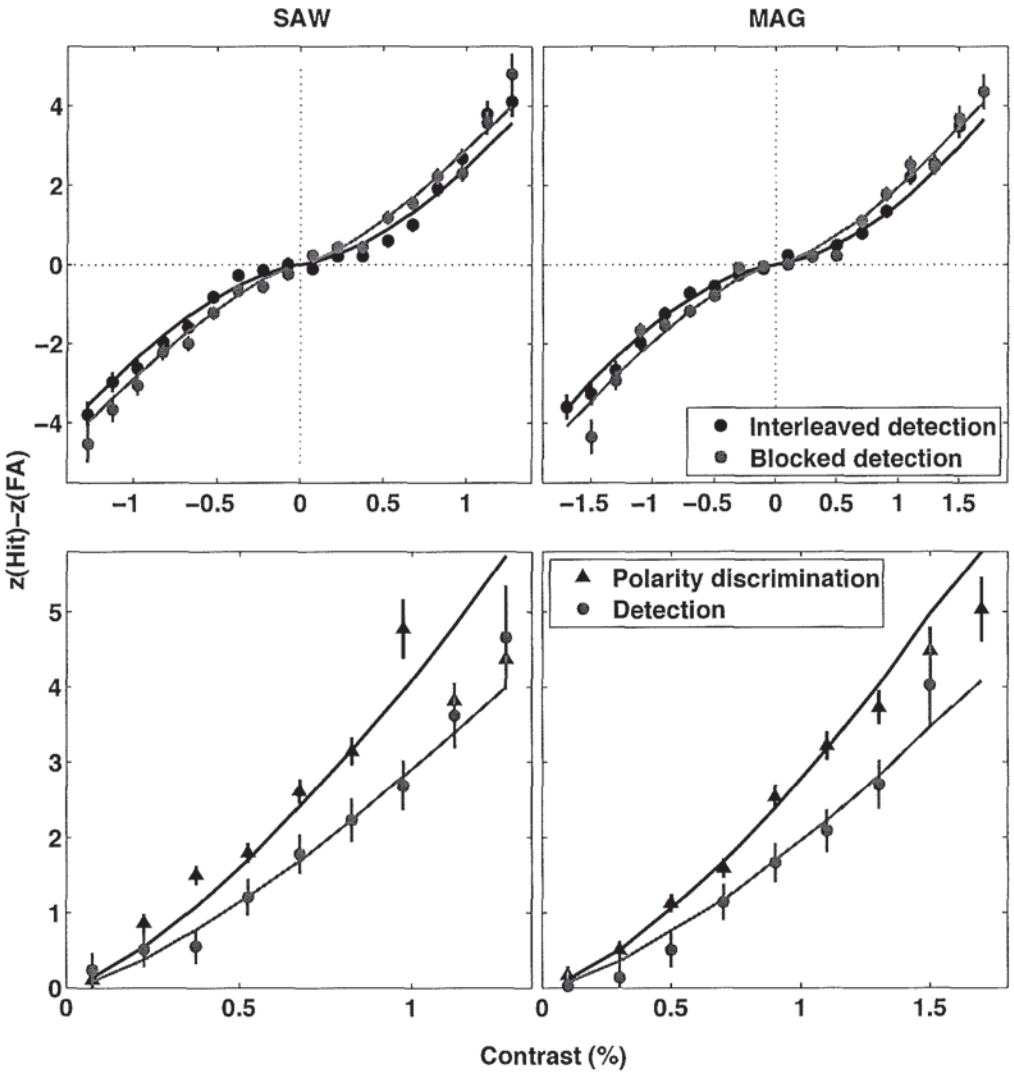

Figure 6.28: Data (symbols) and simulation predictions (lines) for detection and polarity identification of a Gaussian edge with a spatial and temporal soft window. Separate columns are for each observer. Upper row shows all detection data. Lower row shows all polarity discrimination data with blocked detection data averaged across polarity.

The whole set of data is captured nicely by the simulation predictions. The simulation's best-fitting parameter values are shown in table 6.11. For each observer, the transducer and uncertainty versions of the simulation fit equally well. SAW's transducer exponent of 1.35 once again shows relatively linear performance. But now the second observer also has a low exponent value, of 1.40.

Both observers provided a subjective report of the edge's appearance that was at odds with its luminance profile. The luminance profile through the edge's horizontal mid-line contained plateau each side of the edge, which extended to 1.75 degrees on each side. Yet the subjective impression, when the edge was barely visible, was of one or two bars that were only about 1/2 degree wide. This will be discussed below.

6.8.4 Discussion

The aim of the experiment was to determine whether the highly lawful pattern of Gaussian bar data (experiment 10) could be repeated with an edge stimulus, and whether the same Monte Carlo simulation would provide an equally good fit to the data, with similar parameter values. The good fit of the simulation predictions to the data, shown in figure 6.28, suggest that the first two parts of this aim have been met. The final part of this aim is supported by the similarity of the SAW's best-fitting parameter values in this experiment to those of the previous two experiments. Here, the number of irrelevant channels and the transducer exponent were 2 and 1.35 respectively. These values were 1 and 1.20 in experiment 10, and 5 and 1.40 in experiment 11.

The similarity of SAW's bar data in experiments 10 and 11 to his edge data in the present experiment also suggests that these two features are processed in a similar manner. The simulation assumed independent channels for light-dark and dark-light edges. Its good fit therefore provides support for independent processing for separate edge polarities in human vision. However, it cannot be concluded that separate mechanisms exist for edges and bars, because in the present experiment, each edge might have been detected with an adjacent pair of bar detectors; one light and the other dark. If edges were detected in this way, then it might be assumed that edge detection performance would be better than bar detection performance, due to probability summation. However, this assumption cannot be tested in the present series of experiments, because the edge and bar stimuli were not matched in any way (e.g. contrast, area). Furthermore, the subjective appearance of the low-contrast edge was of a narrow pair of opposite polarity bars (or a single narrow bar, located to the left or right of fixation), which suggests that putative bar detectors may be more sensitive than edge detectors at these near-threshold contrast levels.

A direct comparison between SAW's edge data and his bar data in experiments 10 and 11 may also not be appropriate because of the different stimulus configurations. Here, a spatial and temporal soft window was applied, but in experiments 10 and 11 the stimulus had an abrupt spatial and temporal window. The next experiment will address this discrepancy by applying a spatial and temporal soft window to a bar stimulus.

Observer	SAW		MAG	
Model version	transducer	uncertainty	transducer	uncertainty
Error score	19.7	19.7	14.7	14.9
No. of irrelevant channels	(0)	2	(0)	5
Transducer exponent	1.35	(1)	1.40	(1)
Noise sd	1.15	0.90	1.70	1.10

Table 6.11: Best-fitting parameter values for the simulation predictions shown in figure 6.28. Brackets denote a fixed parameter.

6.9 Experiment 13 - 2AFC detection and polarity identification of a Gaussian bar, with a spatial and temporal soft window.

6.9.1 Introduction

Within the current experimental design, a direct comparison between edge data and bar data may shed some light on the degree to which edges and bars are processed in a similar manner in the human visual system. If the same simulation is able to provide a best fit to bar and edge data with the same parameter values, then it can be concluded that the visual system processes these two features similarly. Although SAW's bar simulation parameters were similar to his edge simulation parameters, they were not identical. The difference might reflect experimental noise, or the different stimulus configurations used in the edge and bar experiments. A spatial and temporal hard window was used with the Gaussian bars, yet the windows were soft for the edge. In order to rule out stimulus differences the present experiment will use identical soft windows to those in experiment 12, with similar Gaussian bars to those used in experiments 10 and 11.

6.9.2 Method

The stimulus was a vertical Gaussian bar ($\sigma = 12$mins) similar to that used in experiments 10 and 11, except that it was not restricted to a central region of the monitor screen. It extended to the full height of the screen, but was enveloped by the same spatial and temporal soft windows that were used in experiment 12. Although the stimulus was theoretically of full-screen width, the low value of σ meant that it was only visible to about 1/2 degree each side of centre. All other methodological details were as experiment 12, except that there was only one observer (SAW), who was very experienced with these images.

6.9.3 Results

Data were fitted with the same Monte Carlo simulation described above. As before, '2AFC multiplier' parameter (k) was fixed at 0.85 and the after-image parameter (ai) was fixed at zero. Although now that single interval was not used, the value of k had no impact on the fitted curves. A decrease of k served only to decrease the fitted value of the noise parameter. Data and simulation predictions are shown in figure 6.29 and the best-fitting parameter values are shown in table 6.12.

Model version	transducer	uncertainty
Error score	17.5	18.0
No. of irrelevant channels	(0)	1
Transducer exponent	1.05	(1)
Noise sd	1.45	1.25

Table 6.12: Best-fitting parameter values for the simulation predictions shown in figure 6.29. Brackets denote a fixed parameter.

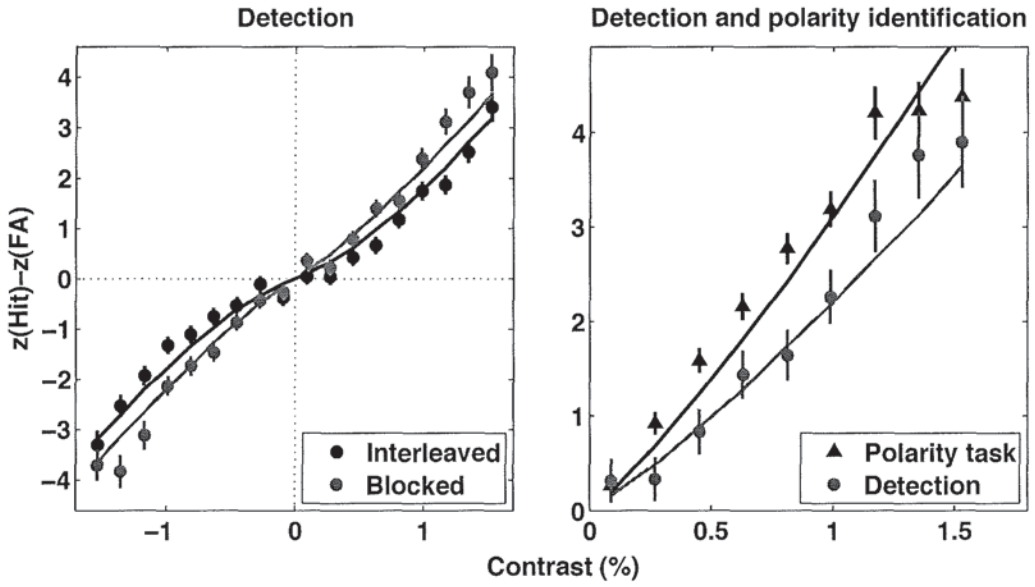

Figure 6.29: Data (symbols) and simulation predictions (lines) for detection and polarity identification of a Gaussian bar with a spatial and temporal soft window. Left panel shows all detection data. Right panel shows all polarity discrimination data with blocked detection data averaged across polarity.

As before, the transducer and uncertainty versions of the simulation provide equivalent fits. Also, the data and simulation predictions show a small effect of blocked vs interleaved trials, and an approximately $\sqrt{2}$ improvement in performance when detection is replaced by polarity identification. The fitted transducer exponent of 1.05 indicates performance that is almost perfectly linear. In order to further examine this linearity the data are re-plotted in figure 6.30 with the best-fitting (least rms) linear function to each data set, constrained to pass through the origin. The values of the the correlation coefficient (r) are high.

6.9.4 Discussion

The good fit of the simulation to the data once again suggests that it correctly captures some of the processing undertaken by the visual system, such as the use of monopolar channels, additive Gaussian noise and (for the transducer version) the dominant source of noise being after the non-linear transducer (due to Birdsall's theorum (Cohn & Lasley, 1975)). The best-fitting parameter values show performance that is slightly more linear than in the previous 3 experiments. Table 6.13 lists the best-fitting parameter values for experiments 10 to 13 for observer SAW. The aim of this experiment was to determine whether the simulation's best-fitting parameters for the Gaussian bar used here were similar to the parameters obtained with the edge stimuli in experiment 12. Table 6.13 shows a small difference of 0.3 for the exponent and 1 for the number of irrelevant channels. Given the fluctuation in these parameter values over experiments 10 to 13, this small difference appears to be random noise. Alternatively it might represent a practice effect, given that the spatial and temporal soft window was first used in experiment 12. Since a comparison across experiments is confounded by order effects, a definitive comparison between edge and bar performance should use both stimuli within one experiment.

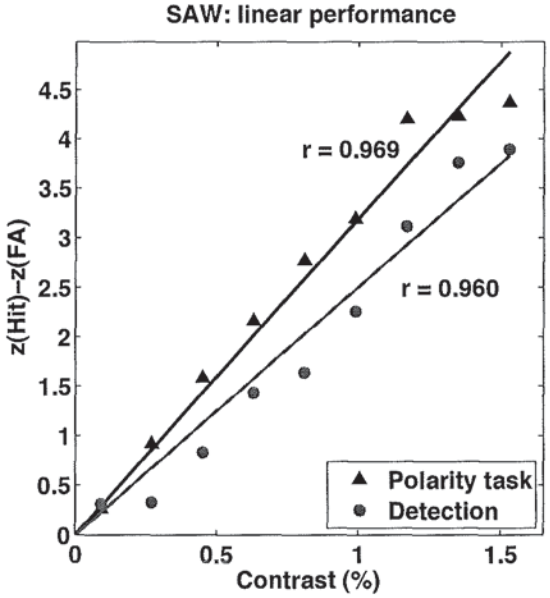

Figure 6.30: Z-scores for 2AFC polarity identification and blocked detection of a Gaussian bar with a spatial and temporal soft window. Detection data are averaged across polarity. The lines are the best-fitting linear function to each data set. r is the correlation coefficient.

Stimulus	Transducer version		Uncertainty version	
	exponent	noise	irrelevant channels	noise
Bar, hard window (expt 10)	1.20	1.35	1	1.10
Bar, hard window (expt 11)	1.40	1.10	5	0.75
Edge, soft window (expt 12)	1.35	1.15	2	0.90
Bar, soft window (expt 13)	1.05	1.45	1	1.25

Table 6.13: Best-fitting parameter values for Monte Carlo simulation fits to SAW's data from experiments 10 to 13.

Table 6.13 shows that a decrease in the transducer exponent is accompanied by a reduction in the number of irrelevant channels. This relationship is in general agreement with that derived by Pelli (1985) for his uncertainty model. He used a Monte Carlo simulation to generate psychometric functions and explored how d' changed as uncertainty was increased, where d' was defined as contrast raised to an exponent, p. The resulting values of p are plotted as a function of uncertainty in figure 6.31 (line). Pelli's definition of uncertainty (M) has been adopted here, which is the total number of channels. Since Pelli assumed that there was one relevant channel, the values of M shown here are one more than the number of irrelevant channels. The figure also shows the values of p and M obtained from experiments 10 to 13. They generally conform to Pelli's predicted relationship.

Previous research also conforms to this relationship, but often at much higher values of M and p. Pelli (1985) obtained values of M that ranged from 18 to 1000, when fitting experimental data. Nachmias & Sansbury (1974) obtained a value of 800 with 9cpd gratings, and Foley & Legge (1981) reported values that ranged form 296 to 2032 for 0.5cpd and 8cpd gratings. Thus the results of experiments 10 to 13 extends confirmation of the predicted relationship to the low levels of uncertainty used here, and demonstrates the equivalence of the transducer and uncertainty models. These two models can

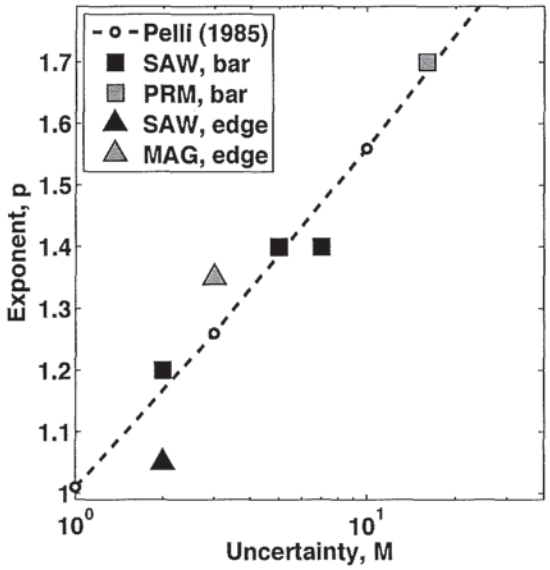

Figure 6.31: Transducer exponent as a function of uncertainty. The line shows the prediction of Pelli (1985). The symbols show the values obtained in experiments 10 to 13 for each observer.

be regarded as the extreme cases of a continuum of models, as some degree of uncertainty with some degree of non-linearity could also account for a given set of non-linear data. Hence it is difficult to reject either model when the data are nonlinear.

However, the situation is different when performance is linear, because the two models converge to a singularity. In experiment 13, SAW's performance was very close to that expected from an 'ideal observer', who has no uncertainty and a linear transducer (irrelevant channels = 0, and transducer exponent = 1). SAW also demonstrated performance that was near-linear in experiment 10 (irrelevant channels = 1) and exactly linear in experiment 8 (grating: $p = 1$). The existence of such linear performance suggests that the non-linear performance reported in the literature may not necessarily represent the limit of visual performance *per se*, but rather the limit of performance given the amount of uncertainty inherent in the experimental design.

In order to examine the pattern of exponent values over all of the experiments in this chapter, table 6.14 shows the value obtained for each observer and type of stimulus. It is unclear from this table whether a pattern exists in the values of p, so they are plotted in figure 6.32 a function of observer, image type and experiment number. The same data are shown in each panel. The image types refer to hard-windowed images unless denoted by 'soft'. Panel a shows that the range of p values is smaller for SAW than DHB, which may reflect SAW's greater experience with these images, which would reduce the level of uncertainty. This conclusion is supported by his generally low values of p. Panel b shows that grating p values are consistently higher for observer DHB than SAW. This suggests that DHB has greater uncertainty than SAW. But panel a shows that DHB attained lower p values for the hard-windowed bar. Thus he is not simply more uncertain about all visual stimuli. Panel c shows the progression of the experiments, which is generally in time order. Values of p greater than 1.5 were obtained only in the earlier experiments, which suggests a long-term effect of practice. However, experience with psychophysical stimuli generally may also be a factor; panel a shows a higher value of

Experiment	Stimulus	Observer	Exponent, p
8	Bar	SAW	1.23
8	Grating	SAW	1.00
8	Bar	DHB	1.23
8	Grating	DHB	2.21
9	Bar	SAW	1.50
9	Grating	SAW	1.15
9	Bar	DHB	1.60
9	Grating	DHB	1.75
10	Bar	SAW	1.20
10	Bar	PRM	1.70
11	Bar	SAW	1.40
12	Edge, soft window	SAW	1.35
12	Edge, soft window	MAG	1.40
13	Bar, soft window	SAW	1.05

Table 6.14: Best-fitting exponent values from experiments 8 to 13.

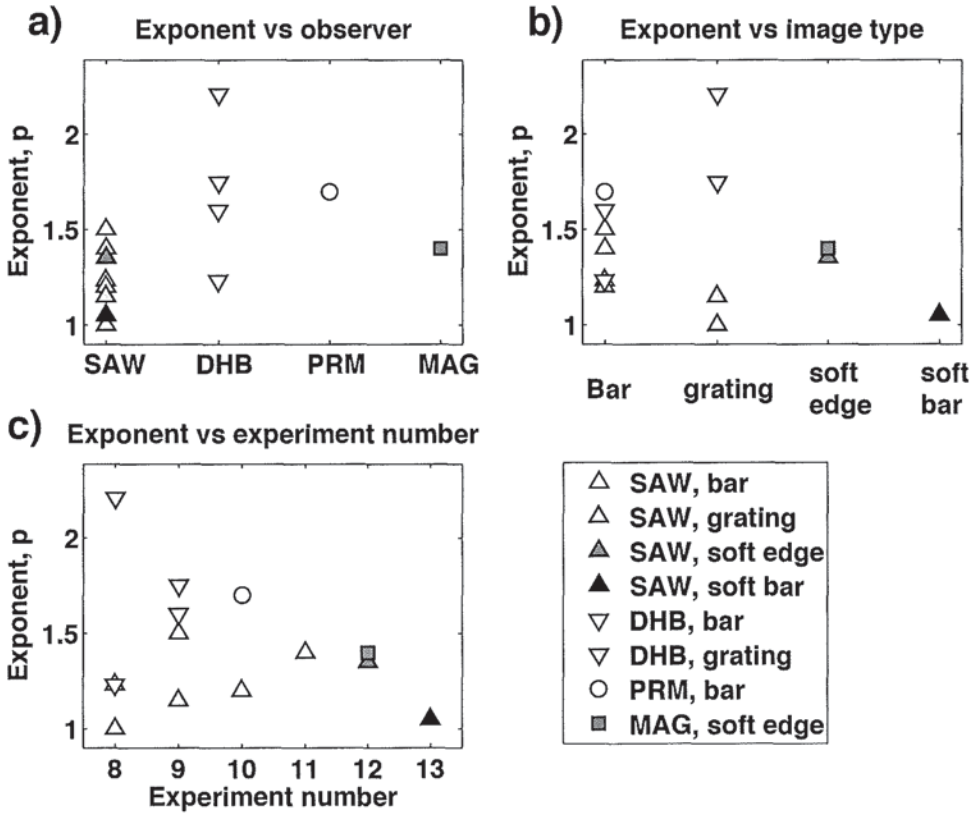

Figure 6.32: Exponent values shown in table 6.14 plotted as a function of a) observer, b) image type and c) experiment number. The same data are shown in each panel.

p for PRM (psychophysically naive) than MAG (very experienced). To conclude, the data show some variability across observers and type of image, but generally show low values of p. The data also hint at a long-term practice effect, which would be expected to reduce uncertainty over time.

An alternative explanation can be posited for the unexpectedly linear performance demonstrated throughout this chapter, based on criterion variance. The following explanation shows how a nonlinear transducer can masquerade as a near-linear transducer. The experiments in this chapter were typically long and the data that formed each psychometric function were collected in several sessions, spread over several days. Thus it seems likely that a given observers criterion would vary over that time. A shift in the criterions position would cause that sessions percentage psychometric function to shift sideways, but would not alter its slope. But when a number of shifted psychometric functions are averaged, the result is a single psychometric function that is abnormally shallow. This shallow slope would be interpreted as low intrinsic uncertainty or a relatively linear transducer. But this conclusion would not be accurate.

In order to investigate whether this explanation accounts for SAWs near linear transducer, his data from two experiments were re-examined. The chosen experiments were the ones that resulted in the most linear performance: experiments 8 (grating data only) and 13. The data from each experiment were placed in time sequence, and then split into 2, 3 or 4 equal portions. Separate psychometric functions were plotted for each portion. This allowed examination of any shift in data over time. All of the psychometric functions from a single experiment approximately superimposed (not shown). No lateral shift was visible. If a lateral shift had been visible, then the above explanation regarding criterion variance would have been supported. However, the fact that no lateral shift was visible does not preclude that explanation completely. The criterion might have varied within the sessions that were contained in one portion, or even within any individual session. This could feasibly be examined by splitting the data into a greater number of portions. But then each psychometric function would be generated from a very low number of trials, resulting in wide confidence intervals. Thus it would not be possible to observe a convincing lateral shift. To conclude, criterion variance between groups of sessions does not provide an alternative account for SAWs near-linear performance.

Although some researchers consider the uncertainty and transducer accounts of nonlinear performance to be equally likely (Foley & Legge, 1981; Nachmias & Sansbury, 1974; Pelli, 1985), the uncertainty model was rejected by Bowen (1995) on the basis of its inability to predict polarity-specific performance. He performed a 2AFC contrast discrimination experiment with pedestal, where the target stimulus was a small Gabor-like pattern derived from the 6th derivative of a Gaussian, and the pedestal was a large cosine grating of matching spatial frequency. The target and pedestal were of either dark or light polarity. When target and pedestal were of the same polarity, the target threshold vs pedestal contrast (TvC) function had the typical 'dipper' shape. But when target and pedestal were of opposite polarity, the function rose monotonically. This monotonic effect was taken to imply that uncertainty was actually increasing over the same range of pedestal contrasts that supposedly reduced uncertainty for same-polarity stimuli. Yet an uncertainty model would predict similar performance in both conditions, because the only difference between them is that the added contrast signal is either decremental or incremental. Thus the uncertainty approach was rejected.

The transducer account was also favoured by Kontsevich & Tyler (1999b). They obtained detection thresholds, using both in-phase and counter-phase sustained pedestals, for a positive DoG (central region was light) and a negative DoG (central region was dark). When the pedestal was in phase

with the target stimulus, the resulting threshold vs contrast (TvC) functions had the expected 'dipper' shape. However, the TvC curves for negative pedestals were markedly different. At small pedestal magnitudes the threshold increased, then suddenly dropped, and subsequently increased in accord with a power law. This function shape is sometimes known as a 'bumper'.

The data were fitted with a nonlinear transducer model with additive noise, which had 12 parameters. So many parameters were used because the TvC curves were not symmetrical across pedestal polarity. Thus the model required separate parameters for each pedestal polarity. Furthermore, the model fitted the upper region of the curves (in addition to the lower region), which corresponds to the upper compressive part of the transducer. For the positive DoG, the fitted transducer exponent (p) was 33 for one observer and 138 for the other. For the negative DoG, p was 1.6 and 3. This large difference in the value of p for each DoG was taken as evidence against the channel uncertainty account of nonlinear performance, because since the DoG stimuli of both polarities had the same size, the degree of channel uncertainty in their detection would be expected to be the same. Therefore, if the accelerating nonlinearity was solely due to uncertainty, the nonlinearity powers would be identical for both conditions. Note that this explanation does not preclude an uncertainty process completely, merely that it must be combined with a nonlinear transducer to account for the results.

Legge et al. (1987) reported that a transducer model was consistent with their data, whereas an uncertainty model was not. They measured contrast-increment thresholds as a function of spectral density of added binary noise, for near-threshold and suprathreshold pedestal contrasts. An expression was derived, which predicted that the threshold signal energy would increase in direct proportion to the noise spectral density for an ideal observer. Thus a plot of threshold signal energy vs noise spectral density would be a straight line of slope 1 that passed through the origin. When switching to a real observer, this predicted function was expected to shift leftwards and/or increase in slope. The amount of horizontal shift was assumed to indicate the amount of equivalent noise in the noisy signal, and the increase in slope indicated the reduction of sampling efficiency.

Legge *et al.* stated that the uncertainty model would predict improved performance when a pedestal is included in a contrast discrimination task, as demonstrated in experiment 11, above. In other words, performance would be expected to be closer to that of an ideal observer. They suggested that this analysis can be extended to the case where various levels of external noise are added to a constant pedestal. Low external noise should lead to good performance, and high external noise should lead to impaired performance. The impaired performance would be expected to increase the slope of a threshold signal energy vs noise spectral density function, relative to the slope obtained with low external noise. Although their experimental data did show a higher slope for high external noise than low external noise, the increase was not significant. This was taken as evidence against the uncertainty model. However, they did not attempt to model the amount of change in slope expected from the uncertainty model, so it is not clear if they really are justified in rejecting it.

Legge *et al.* also found that the psychometric slopes for detection were lower in external noise than in no noise. This was taken as further evidence against the uncertainty model, because the model predicted that the slopes of detection psychometric functions should be the same in the presence or absence of external noise. The reasoning behind the prediction was not stated. This is a pity, because an increase in noise would surely increase stimulus uncertainty, which is known to increase psychometric slopes (Pelli, 1985). Legge *et al.* found that their data did, however, support a transducer model.

The work of Lasley & Cohn (1981) supported the uncertainty model and not the transducer model.

They applied Birdsall's theorem to a 2AFC luminance discrimination task. According to this theorem, when the dominant source of noise is before a nonlinear transducer, the resulting psychometric function will be linear. This behaviour occurs because the noise distribution and the signal-plus-noise distribution are affected in the same way by the nonlinearity. Thus the signal to noise ratio is unchanged. This theorem means that if a large amount of external noise was added to a signal, then nonlinear performance that is due to a transducer would become linear. However, Lasley & Cohn argued that if nonlinear performance was due to uncertainty, then the addition of external noise would have no effect on performance. When these predictions were tested, the data showed that the addition of external noise caused no significant change in psychometric slope, which led to rejection of the transducer model.

Cohn & Lasley (1985) continued to test the transducer account of non-linear performance by performing a luminance discrimination experiment with a pedestal, where the stimulus was dichoptic[2]. The results showed that the pedestal caused nonlinear performance to become more linear. The conclusion reached was that the source of nonlinearity must be located at a stage after the signals from the two eyes are combined. Furthermore, an application of Birdsall's theorum suggested that the dominant source of noise must be after that stage. That seemed unlikely to Cohn & Lasley as they suggested that the dominant source of noise was likely to be before that stage, and due to "quantal fluctuations, pupil and accommodative fluctuations, dark light, and synaptic and membrane noise of the early elements". Thus they once again rejected the transducer model.

The uncertainty model was favoured by Solomon (2007) because it could fit the data from a two-response 4AFC detection experiment, whereas a transducer model could not. In this experimental design the stimulus is presented in one of 4 intervals. The observer's first response indicates which interval appeared to contain the stimulus. The observer then makes a second response, which indicates which of the remaining 3 intervals is most likely to contain the stimulus if their first response was incorrect. High threshold theory predicts that incorrect first responses are due to the stimulus level falling below a fixed threshold and the observer guessing incorrectly. Thus the probability of a correct 2nd response would be exactly 1/3. However, experimental data provided by Swets et al. (1961) showed that the proportion of correct 2nd responses systematically increased with contrast, which allowed them to reject high threshold theory. They applied signal detection theory, but found that it provided a poor fit to the data until the usual equal variance assumption was replaced by the assumption that the variance of the signal increased with its mean.

Solomon (2007) produced the psychometric functions predicted by signal detection theory for a 2AFC experiment, but found that when equal variance was assumed, the slope of the psychometric functions were too shallow to fit his experimental data. When he then applied Swets *et al.*'s amount of increasing variance, the resulting slopes were even shallower. He then attempted to improve the fit by incorporating a nonlinear transducer. A exponent value of 3.3 provided a good fit. However an uncertainty model (without nonlinear transducer) provided an equivalently good fit with 280 channels. Interestingly, this amount of uncertainty also fitted Swets *et al.*'s results, without requiring any additional assumptions. But a transducer model was unable to fit these results without the additional assumption of increasing variance.

An alternative to the transducer and uncertainty models was proposed by Kontsevich & Tyler (1999a). Their 'distraction model' was designed to emulate an observer who had total certainty of the stimulus'

[2]'Dichoptic' means, in this context, that the pedestal was presented to one eye, and the luminance increment to the other eye.

time and place, but whose attention was sometimes misdirected to irrelevant channels. It was modelled with a version of Pelli's uncertainty model that had an additional stage that followed the final max operator. This stage vetoed the output of the previous stage if attention was elsewhere in a given 2AFC interval. For example, if distraction occurred in interval 1 only, then any amount of signal in interval 2 would result in an 'interval 2' response, even if the interval 2 signal was generated purely by noise. On these occasions the model would be incorrect 100% of the time. Likewise, if distraction occurred during the non-signal interval, then the model would always select the signal interval, and so would be 100% correct. If distraction occurred in both intervals, then the model responded at chance. If there was distraction in neither interval, then the interval with the largest response was selected.

Kontsevich & Tyler used a simulation to show that the model provided predictions that were broadly consistent with Pelli's uncertainty model, when it used 1 relevant channel and a given number of irrelevant channels. However, it produced markedly different predictions when the pool of available channels was fixed at 100,000 and the proportion of channels that were relevant was varied. Kontsevich & Tyler did not run the uncertainty model under the latter conditions, so it is not clear whether the distraction and uncertainty models would differ here. The similarity of the models' predictions may be due to the attention algorithm used in the distraction model. It directed attention to the strongest monitored signal. This appears to repeat the max operator that is the final stage of Pelli's model, with the result that this additional 'attentional' stage is actually transparent.

Uncertainty was an important component of a model proposed by Meese & Summers (2009). They assessed summation of contrast across eyes and area at detection threshold. In one condition, the stimulus was a sine-wave grating that was enveloped by a plaid, producing a 'checkered' pattern. In another condition the stimulus was a full grating had the same spatial extent as the checkered pattern. So a comparison of detection thresholds across these two conditions allowed an examination of area summation that was not confounded by a change in stimulus spatial extent. A comparison of monocular *vs* binocular presentation allowed examination of the effect of summation across eyes. Meese & Summers considered 192 model architectures that were composed of the following 4 elements, in all possible orders: (i) linear summation or a MAX operator across eyes, (ii) linear summation or a MAX operator across eyes, (iii) linear or accelerating contrast transduction, (iv) additive Gaussian noise. The most successful four-element model was linear summation across eyes, followed by nonlinear transduction, linear summation across area, and late noise. However, the model fitted better when additional nonlinear transducers were placed before each summation stage. The final transducer and late noise in this model were then replaced by using multiple parallel noisy channels, which created intrinsic uncertainty. Although the change from transducer to uncertainty versions of the model resulted in no change in its predictions, the uncertainty version was preferred by Meese & Summers, because the resulting model structure more closely resembled that proposed in earlier work (Meese & Summers, 2007).

Lu & Dosher (2008) presented evidence that favoured the nonlinear transducer approach. They examined 5 different observer models and employed two methodological procedures to constrain the models. In one procedure, termed 'triple-TvC, observers judged the orientation of a Gabor that was presented in 2D noise, at a range of contrast levels for target and noise. The resulting threshold contrast was plotted as a function of noise contrast (TvC function) at three different criterion levels (65%, 75% and 85% correct). The three TvC functions were then used to obtain two threshold ratios. The ratio between the 65% and 75% performance levels was greater than the ratio between 75% and

85%, which was consistent with a nonlinear transducer.

The second procedure employed by Lu & Dosher was the double-pass method. The sequence of stimulus presentations was recorded, and then the experiment was repeated in an identical manner, thus forming a second pass over the stimuli. Thus, for each condition there were two sets of data, one for each pass. Since the stimuli in the two passes were identical, any variation in response between the two passes could be attributed entirely to the observers internal noise. If the internal noise was extremely high, the proportion of agreement (P_A) of decisions made on the two passes would be at chance (0.5 for 2AFC). But if the internal noise was negligibly low, then the P_A would be 1. A plot of proportion of correct responses vsP_A was an approximately linear function with positive gradient, and with P_A values from about 0.75 to 0.9. When the data from both procedures was fitted to 5 different models, the best fit was with a perceptual template model that incorporated additive noise, multiplicative noise and a nonlinear transducer. In contrast, models that incorporated uncertainty performed less well.

To summarise this previous research, although many studies do not distinguish between the uncertainty and transducer models, there are a few that favour one or the other model. Bowen (1995) and Kontsevich & Tyler (1999b) rejected the uncertainty model because of its inability to predict masking when target and pedestal were of opposite phase. Yet the uncertainty model can only ever account for facilitation. Once the pedestal has reduced stimulus uncertainty by the full amount (where the observer monitors only the relevant channel(s), and no irrelevant channels), then further increases in pedestal contrast would not alter the number of channels monitored. If uncertainty was the only influence on performance then as pedestal contrast increased from zero, detection thresholds for the target would first decrease, and then level off (Solomon, 2007). It seems likely, therefore, that facilitation and masking represent different processes (Georgeson & Georgeson, 1987), and that the uncertainty model accounts for facilitation only. Bowen (1995) and Kontsevich & Tyler (1999b) were attempting to apply the uncertainty model to masking data, which may be outside of this model's domain. Therefore their rejection of it may be unwarranted.

Perhaps the strongest evidence that counters the uncertainty model was provided by Goris et al. (2008). They used a 2AFC task to measure contrast detection performance for a 7cpd Gabor (in one interval) in the presence of static 2D white noise (in both intervals). When the noise power was high, it raised the detection threshold, but when it was low it facilitated detection of the Gabor. This facilitation effect, which was termed 'stochastic resonance', had been observed previously in the tactile (Collins et al., 1996) and auditory (Zeng et al., 2000) domains. It was not attributed to a reduction in stimulus uncertainty, for 2 reasons. Firstly, the experimental design minimised spatial and temporal uncertainty. The Gabor was surrounded by a clearly visible red square, whose onset and offset coincided with the start and end of each interval. Furthermore, the interval between trials was fixed at a constant duration (even if the observer had failed to respond), which gave the experiment a predictable rhythm. Secondly, the noise covered a larger area than the target Gabor, so it would not reduce spatial uncertainty.

Goris et al. (2008) fitted the detection data with an 8-parameter model that had the stages of external noise, linear filter, rectifier, Naka-Rushton transducer and dominant late internal noise. The model predicted the observed facilitation because rectification caused the external noise to increase the mean filter response to both a weak Gabor-plus-noise and a noise stimulus. Since these 2 responses were then passed through the accelerating portion of the Naka-Rushton transducer, the difference between them increased, which lead to improved performance. Thus a model that included a non-linear transducer was able to fit these low-contrast data, whereas an uncertainty account appeared unfeasible.

To conclude, the uncertainty and transducer models both provide a good account of low contrast performance in the experiments of this chapter and in much previous research. However, the linear performance of observer SAW suggests that his transducer is actually linear, and that nonlinear performance reflects the monitoring of irrelevant channels (perhaps due to a lack of extensive practice) and/or experimental noise. The uncertainty model provides a more complete account of some results in this chapter than the transducer model, because the variation in performance between observers is generally consistent with their psychophysical experience. This provides a nice explanation that is in keeping with the uncertainty model. A transducer model, by contrast, can only state that different observers have different transducer exponents, without any explanation regarding why. However, given that both models fitted the data in this chapter equally well, and both have support in previous literature, the final conclusion is that both are viable models of human low contrast vision.

6.10 Chapter summary

The direction of this chapter was was inspired by the unexpected observation of nearly linear, low contrast detection performance in experiment 7 of chapter 5. Although contrast detection and discrimination have been extensively examined in previous research, such linear performance has not been reported before. Discriminability (d') is typically reported to be a power function of contrast, where the value of the exponent is about 2 to 3 (or even higher). Thus the novel near-linear performance warranted further examination.

Experiment 8 used a single interval polarity identification method and showed that near-linear performance was not specific to observer SAW and to Gaussian bar stimuli. It was also observed in the second observer's performance with Gaussian bars, and in SAW's performance with a sinusoidal grating (where performance was totally linear: exponent, $p = 1$). In order to determine whether near-linear performance was task-specific, the experiment 8 grating and bar stimuli were presented in a new experiment (experiment 9) that used a 2 x 2AFC design, whereby polarity identification and detection could be observed concurrently. The interval and polarity responses were very similar within each condition. Thus SAW's near-linear performance was shown to extend to the detection task within this paradigm, for the grating stimuli. However his bar data and the second observer's data for grating and bar stimuli were a little less linear.

Experiment 10 was a very comprehensive examination of performance with single interval and 2AFC presentations where the tasks were detection (with blocked or interleaved stimulus polarity) and polarity identification. The results established that monopolar channels were used rather than bipolar channels, and demonstrated that best performance was obtained when external uncertainty was minimised. The whole set of data for a given observer was well-fitted with a Monte Carlo simulation. 'Uncertainty' and 'transducer' versions of the simulation fitted equally well, although the uncertainty account was considered a more parsimonious explanation of the whole pattern of data.

Experiment 11 tested the prediction that near-linear performance should lead to an absence of facilitation that is typically observed with low-contrast pedestals (i.e. the absence of a dip in a threshold vs contrast function). However, performance was slightly less linear than previously observed, so facilitation was observed. The magnitude of facilitation was consistent with the exact degree of nonlinearity in the detection data.

Experiment 12 extended the work from a Gaussian bar to a single Gaussian edge (with a spatial and temporal soft window), with the finding that performance was of equal linearity. When the same spatial and temporal soft window was applied to a Gaussian bar (experiment 13), performance approached total linearity ($p = 1.05$). Thus near-linear performance was maintained over several experiments, for one observer. Furthermore, the data from the other 3 observers were fitted with low transducer values ($p = 1.23 - 1.70$). This provides clear evidence that the transducer exponent, for low-contrast detection, is not inevitably constrained to the values that have been reported previously ($2 - 4$).

Chapter 7

Summary and conclusions

This chapter summarises the findings of each experimental chapter and suggests the direction of future research.

7.1 Mach edges (chapter 3)

In experiment 1, a new phenomenon was introduced, termed 'Mach edges', which are perceived edges where there are no peaks in the luminance profile's 1st derivative or zero-crossings in its 2nd derivative. Using blurred triangle-wave stimuli, Mach edges were found to be located at 3rd derivative peaks, where models based on the 1st or 2nd derivative predicted no edges. Predictions of a scale-space, non-linear 3rd derivative model provided a good fit to the data. In experiment 2, a Yes/No method was used to test the model further, with a new set of stimuli that were blurred triangle-waves with an added luminance ramp. The addition of the ramp can be equivalently described as the addition of a vertical offset to the 1st derivative profile, followed by integration to form the luminance profile (see figure 3.17). Mach edges were still perceived under most conditions tested.

The 1st derivative offset had no effect on the 2nd derivative profile, which always had a box function profile whose base was at zero, and whose sides were at the marked Mach edge locations. So an alternative hypothesis for Mach edge detection could be proposed, without recourse to the 3rd derivative. Mach edges might occur at the junction of zero and non-zero values in the 2nd derivative. This hypothesis seems plausible, given that zero-bounded regions of activity in the 2nd derivative (albeit pooled across scale and half-wave rectified) are the basis of the MIRAGE feature detection model. In order to test the hypothesis, future work could apply a vertical offset to the 2nd derivative, before integrating it twice to form the luminance profile. If Mach edges are no longer perceived under these conditions, then the hypothesis would be supported, and role of the 3rd derivative in Mach edge detection would be somewhat undermined. Conversely, the presence of Mach edges under these conditions would support the 3rd derivative approach.

Modelling of the Mach edge data of experiment 2 required the addition of a DOG pre-filter, whose properties were based on known ganglion cell physiology. Unfortunately, this addition introduced the DOG + 1st derivative as a viable alternative to the DOG + 3rd derivative when modelling the data from experiments 1 and 2. In experiment 3, a feature-marking method was used to examine the position

and strength of Mach edges in the image set of experiment 2. The resulting data were consistent with both DOG + 1st derivative and DOG + 3rd derivative models. In experiment 4, high spatial frequency stimuli of short duration were then used to attenuate and by-pass the putative pre-filter's inhibitory surround. Mach edges were still perceived, and they were predicted equally well by the 3rd derivative model, and the DOG + 3rd derivative model, but not by the 1st derivative model, nor the DOG + 1st derivative model. The conclusion reached was that the 3rd derivative approach was superior to the 1st derivative approach as an account of Mach edge perception.

Taken together, the results of this chapter might be thought to be inconsistent, because the DOG pre-filter was not required in the modelling of the results when the waveform was a blurred triangle-wave (experiments 1 and 4), yet it was required when the luminance ramp was added (experiments 2 and 3). It was proposed that the pre-filter emulated the effect of retinal processing, so it would surely be expected to apply to all stimuli. However, when applied to the waveforms of experiment 1, the predictions of the single-scale 3rd derivative model were very similar to the predictions when the DOG was added (figure 3.34). This suggests that DOG + 3rd derivative filtering does apply across all of the stimuli described in chapter 3. Thus there is no inconsistency.

The 3rd derivative model computed the 3rd derivative in 2 stages, with half-wave rectification after each stage. If the input waveform was a blurred edge, then the peak in the model's output was narrower than the peak generated by a 1st derivative filter applied to the same edge. This means that a 3rd derivative model would be expected to have better resolving power than a 1st derivative model. This property has not been explored. A future psychophysical experiment could employ 2 neighbouring, same-polarity edges at a range of separations (while holding edge blur constant) to determine the threshold separation that produces a perception of two separate edges, rather than a single edge. Furthermore, in different conditions, the blur of one or both edges could be varied, in order to fully explore resolving power in scale-space. When the threshold luminance waveforms are passed through 1st derivative and 3rd derivative models, 2 separate edges might be predicted by the 3rd derivative model, whereas only 1 blurred edge might be predicted by the 1st derivative model. If so, then this would provide further evidence in support of the role of 3rd derivative filters in edge detection.

7.2 Mach bands (chapter 4)

The perception of Mach bands was examined by using a novel set of edge stimuli that spanned a range of Mach band detectabilities. The waveforms ranged from a slightly blurred Mach ramp, through a Gaussian edge, to a very blurred edge. In experiment 5.1, a single-interval Yes-No method was applied with a Mach band detection task, and in experiment 5.2 a feature-marking method was applied. The resulting data were matched against the predictions of 5 competing models that all used 2nd derivative filters, or approximations thereof. The best-fitting model was a scale-space model that used peaks in the normalised 2nd derivative to detect bar locations, and then used the bar-edge amplitude ratio at these locations to determine the likelihood of Mach band perception. It was concluded that this model gives a good account of the processing of Mach bands in the human visual system.

The 1st defeated model computed the 2nd derivative at a single fine scale, and its predictions were too sensitive to the scale of the experimental images. In contrast, the detection responses for each image scale were similar, which demonstrated a greater degree of scale-invariance than predicted.

The 2nd defeated model was the single-scale 3rd derivative with DOG pre-filter. This was an edge detection model that had provided a good fit to the data in chapter 3. It was incapable of producing bar predictions directly, but did produce predictions for the edges of the Mach bands, which were a good fit to the Mach band detection data. However, this model was rejected because it predicted Mach band widths that were up to a factor of 2 times wider than their perceived widths. It is perhaps unsurprising that this edge detection model performed less well than a model that was specifically designed to detect bars.

The 3rd defeated model was the energy model. It was designed to detect bars and edges concurrently. It assumed that features were located at peaks in local energy, whose profile was obtained by squaring and summing the output of a quadrature pair of filters. Where it did predict Mach bands, it provided a reasonable fit to the data. But it was rejected on the basis that it predicted the absence of Mach bands in 6 of the 19 experimental images where they were actually perceived. This finding is consistent with previous reports of its failure to predict a sufficient number of perceived features in a number of 1D images (Georgeson & Freeman, 1997; Hesse & Georgeson, 2005).

The 4th defeated model was MIRAGE. It was similar to the successful scale-space edge-bar model, in that it derived bar locations from the combined output of 2nd derivative filters of various scales. But it differed in the normalisation factor applied to filter outputs, and in the manner in which the outputs were combined across scale. MIRAGE was rejected because its predictions for Mach band detection were over-sensitive to image scale; an increase in image scale by a factor of 4 caused a predicted reduction of Mach band strength by a factor of 80. Yet the proportions of Mach bands detected were relatively similar at the 3 image scales tested.

Although the scale-space edge-bar model was better than 4 other models in accounting for Mach band perception, this may not be the only viable model. There may be other, currently unconsidered models that fit the data just as well, or even better. Future research could examine the effect of systematic alterations to the scale-space edge-bar model, in order to determine whether a revised version would fit better, and the bounds of successful performance.

7.3 Luminance bars (chapter 5)

The previous chapters had examined the perception of 'illusory' edges and bars. This chapter extended the work to 'real' luminance bars. The stimulus was either an f + 3f grating or a narrow Gaussian bar on an inverted wide Gaussian bar background. The experimental hypothesis was that the null point between seeing and not seeing the thin central bar would occur at a small Gaussian or f contrast that caused a central plateau in the luminance waveform. For the f + 3f image set, this hypothesis was supported for one, but not the second observer. But for the Gaussian bar image set, the empirical support was more consistent. Furthermore, the threshold for detection of the thin Gaussian bar was found to approximately equal the sum of the contrast for polarity identification of the thin bar (on a blank background) and the contrast required to bring the centre of the luminance waveform to a plateau. These results suggested that the background bar or grating was acting as a simple luminance background, and did not interact with the target bar in a facilitatory or inhibitory manner.

However, an additional analysis of the experimental data suggested some kind of facilitatory effect, as a 'dipper-like' function was obtained. This apparent discrepancy was resolved by appealing to the

different reference level in each case. Facilitation was observed when the bar detection thresholds were referenced to the thresholds that produced a 50% performance level (c_{null}), and the lack of facilitation was observed when the reference level was the contrast that produced a luminance plateau.

However, this analysis is incomplete. Firstly, if the analyses that suggested facilitation and no facilitation truly represent just 2 different viewpoints of the same data set, then it seems likely that the approach taken by one analysis could be used to produce predictions that are in the format of the other analysis. If there was a match, then the notion of 2 viewpoints would be supported, but if the match was poor, then the comparison between the 2 original analyses would require more thought. Secondly, a new line of thought could be followed, whereby the raw psychometric functions (or perhaps their z-scores) are manipulated in such a way as to make them all superimpose. The manipulations that succeed in making the functions superimpose could then be assumed to partly describe the visual processing of the experimental images. Future work could address these issues.

7.4 Transducer linearity (chapter 6)

It is widely accepted that contrast detection and discrimination performance can be characterised, in part, by the action of a non-linear transducer that transforms stimulus contrast into internal response. The putative transducer is typically described as accelerating at low contrasts, compressive at high contrasts, and (depending on the particular model) approximately linear at moderate contrast levels. At low contrast levels, discriminability (d') is typically reported to be a power function of contrast, where the value of the exponent (p) is about 2 to 4. Unexpectedly, nearly linear, low contrast detection performance had been observed in experiment 7 of chapter 5. The remainder of the experiments in this thesis examined this exciting phenomenon further.

The results of experiment 8 demonstrated totally linear or near-linear polarity identification performance when the stimulus was a Gaussian bar or a sinusoidal grating. In experiment 9, the same images were again examined at low contrast, but now a 2 x 2AFC method was employed, whereby polarity identification and detection could be observed concurrently. Near-linear performance was shown to extend to the detection task, for the grating stimuli.

Experiment 10 employed only Gaussian bars, and used single interval and 2AFC methods with tasks of detection (with blocked or interleaved stimulus polarity) and polarity identification. Again, near-linear performance was observed ($p = 1.20$), although performance for one observer was less linear ($p = 1.70$). Polarity discrimination z-scores were $\sqrt{2}$ higher than detection z-scores, which supported the assumption of separate, independent, monopolar channels for dark and light bars. 2AFC z-scores were about 1.7 times higher than single interval z-scores, which was consistent with the predictions of signal detection theory (with criterion variance), and implied that sensitivity (d') was the same for both methods. Moving from interleaved trials to trials blocked by polarity caused a small but distinct improvement in z-scores when the observer's intrinsic uncertainty was low. But when intrinsic uncertainty was higher, the effect of blocking was hidden. All of these effects were captured simultaneously by a Monte Carlo simulation that employed several 'irrelevant' channels and a linear transducer (uncertainty version) or it employed no irrelevant channels and a non-linear transducer (transducer version). Uncertainty and transducer versions of the simulation fitted equally well.

It is well-established that a low-contrast 'pedestal' grating facilitates grating detection, when the pedestal contrast is close to detection threshold. This facilitation effect is typically ascribed to a transducer that is accelerating at low contrast levels. But if the transducer was near-linear, then the facilitation would be much reduced. Experiment 11 tested this prediction with a pedestal of a single contrast level and with the observer that had shown most linear performance so far. Performance was slightly less linear than previously ($p = 1.4$), and the amount of facilitation was consistent with the degree of non-linearity.

Experiments 12 and 13 applied a spatial and temporal soft window to Gaussian edge and bar stimuli, and again, near-linear performance was observed. The data were again well-fitted by the Monte Carlo simulation. When the experiments of this chapter were considered together, the transducer values ($p = 1 - 1.70$) were all lower than the values that have been reported previously ($p = 2 - 4$) for low contrast detection. Near-linear performance appears to reflect low uncertainty, which may be partly due to extensive practice and experimental conditions that minimise stimulus uncertainty.

In experiment 10, the comparison of polarity discrimination z-scores with detection z-scores led to the conclusion that independent monopolar channels existed for dark and light bars. A similar method could be used in future research to determine the relationship between edge and bar channels, rather than dark and light channels. However, it is not possible to make a direct comparison between the edge data of experiment 12 and the bar data of experiment 13, because the edge and bar images had not been matched for detectability, and the experimental design did not include stimulus jitter. The absence of jitter means that an edge could be detected with 'bar detectors' that are offset from image centre, and a bar could be detected with offset 'edge detectors'. Thus a future experiment could jitter the features between trials in order to prevent the observers from using these cues. If the z-scores for discrimination between a bar and an edge were $\sqrt{2}$ higher than their detection z-scores, then it could be assumed that the edge and bar channels are orthogonal (independent). However, this approach would require that the edge and bar contrast levels had previously been made equivalent (i.e. that each pairing of edge contrast and bar contrast produced the same detection z-score). Therefore a preliminary experiment would be required that examined the detectability of edges and bars, and that produced a bar-edge contrast scaling factor.

The existence of channels sensitive to a restricted range of contrasts has received some physiological (Albrecht & Hamilton, 1982) and psychophysical (Yu et al., 2004) support. Contrast uncertainty could potentially have influenced all of the experiments in this thesis, due to the use of interleaved contrast levels. If trials were blocked by contrast, then this potential source of uncertainty would be removed and performance might become even more linear. A future experiment could employ a polarity identification task to explore this blocking effect. If the blocked contrast data were the same as interleaved contrast data, then the absence of a blocking effect could be concluded. The model offered in chapter 6 would indeed predict no such effect. But if a blocking effect were observed, then the question would arise as to why the use of interleaved contrast levels caused impaired performance. Two hypotheses are presented below, followed by a method of distinguishing between them.

The first hypothesis concerns fluctuations in the position of the observer's internal criterion. To take the example of a single interval task, the criterion's optimal position lies at the midpoint between the peaks of the 'light target' and 'dark target' distributions in decision space. If trials were blocked by contrast, then within each block, the criterion should remain at this location. However, if contrast levels were interleaved, then the criterion would be expected to drift somewhat, depending on the

contrast levels and responses in the previous few trials. This drift in criterion position over the course of an experiment has been well-documented and modelled by Treisman & Williams (1984). The lack of stability in the criterion's position effectively adds a source of noise, which would therefore impair performance to a level below that observed in blocked contrast trials.

The second hypothesis involves an increase in uncertainty, and relies on the existence of 'contrast channels'. When trials are blocked by contrast, then the observer would soon learn to attend to only the relevant contrast channel(s) and ignore the others. But when interleaved trial are used, the observer would be uncertain about which channels to monitor, and so would monitor several. This would lead to poorer performance, as demonstrated by the above Monte Carlo simulation when the number of irrelevant channels is high. Thus again, interleaved trials would lead to poorer performance than trials blocked by contrast. However, if only one channel served all contrast levels in the tested range, then performance should be very similar in both conditions.

These two hypotheses suggest two slightly different patterns of results. If the first hypothesis (criterion variance) were true then the effect of interleaved trials would apply across all contrast levels to the same degree. In other words, a plot of performance z-scores *vs* contrast would have a lower slope in the interleaved condition than the blocked condition. But the degree of non-linearity would be the same in both conditions. In other words, the amount of curvature at low z-scores would be very similar in both conditions. However, if the 2nd hypothesis (uncertainty reduction) were true, then a switch from blocked to interleaved trials would increase uncertainty and therefore the degree of non-linearity in the data. In other words, the amount of curvature at low z-scores would be greater in the interleaved condition than in the blocked condition. Future work could address these issues.

One feature of the uncertainty model presented in this thesis, and versions presented elsewhere, is that the source of the uncertainty is not prescribed. Pelli (1985) supposed that the uncertainty was about the spatial and temporal limits of the target stimulus, but did not speculate on a physiological locus. In order to rule out or support a peripheral locus, a future experiment could attempt to determine whether stimulus certainty (i.e. near-linear performance) can transfer across orientation, spatial frequency or eyes. In the context of an uncertainty model, this approach would examine the properties of the observers ability to attend to a small number of irrelevant channels, arising from extensive practice.

7.5 Thesis summary

The experiments in this thesis have demonstrated the success of a derivative-based multi-scale approach to feature detection in human vision. The 3rd derivative of the luminance profile was shown to have a critical role in the perception of an exciting novel phenomenon 'Mach edges'. This discovery clearly demonstrated the incompleteness of the well-established use of lower derivatives in edge detection.

Although previous research has suggested a role for 2nd derivatives in the perception of Mach bands, the work in this thesis demonstrated the inadequacy of a single-scale approach, and the importance of a multi-scale 2nd derivative computation, where information is combined across scale in scale-space.

Possibly the most exciting novel discovery, presented in this thesis, was the near-linear performance for detection and polarity identification at low-contrast levels, which accompanied extensive practice and low extrinsic uncertainty. This degree of linearity has never been observed previously in human vision, and it redefines the upper limits of recorded performance.

Bibliography

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.

Adini, Y., Sagi, D. & Tsodyks, M. (1997). Excitatory-inhibitory network in the visual cortex: Psychophysical evidence. *Proceedings of the National Academy of Sciences of the United States of America*, 94, 10426–10431.

Ahumada, A. & Lovell, J. (1971). Stimulus features in signal detection. *Journal of the Acoustical Society of America*, 49, 1751–1756.

Akutsu, H. & Legge, G. E. (1995). Discrimination of compound gratings: spatial-frequency channels or local features? *Vision Research*, 35, 2685–2695.

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

Allik, J., Dzhafarov, E. & Rauk, M. (1982). Position discrimination may be better than detection. *Vision Research*, 22, 1079 – 1081.

Anzai, A., Ohzawa, I. & Freeman, R. D. (1999). Neural mechanisms for processing binocular information ii. complex cells. *Journal of Neurophysiology*, 82, 909–924.

Argyle, E. (1971). Techniques for edge detection. *Proceedings of the Institute of Electrical and Electronics Engineers*, 59, 285–286.

Argyle, E. (1972). Techniques for edge detection - reply. *Proceedings of the Institute of Electrical and Electronics Engineers*, 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. (1984a). How do we discriminate relative spatial phase. *Vision Research*, 24, 1847–1857.

Badcock, D. R. (1984b). Spatial phase or luminance profile discrimination? *Vision Research*, 24, 613–623.

Barbieri, G. S. A. (2000). *The role of spatial derivatives in feature detection.* Ph.D. thesis, Aston University.

Barlow, H. B. (1958). Temporal and spatial summation in human vision at different background intensities. *Journal of Physiology*, 141, 337–350.

Bergen, J. & Landy, M. (1991). Computational models of visual texture segmentation. In Landy, M. & Movshon, J. (eds.), *Computational models of visual processing*. MIT Press, Cambridge MA.

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

Bergen, J. R., Wilson, H. R. & Cowan, J. D. (1979). Further evidence for four mechanisms mediating vision at threshold - sensitivities to complex gratings and aperiodic stimuli. *Journal of the Optical Society of America*, 69, 1580–1587. Pdf hard copy Nice explanation of subthreshold summation on page 1584.

Bergholm, F. (1987). Edge focusing. *IEEE Transactions On Pattern Analysis and Machine Intelligence*, 9, 726–741.

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.

Blakemore, C. & Nachmias, J. (1971). The orientation specificity of two visual after-effects. *The Journal of Physiology*, 213, 157–174.

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

Bovik, A. C., Clark, M. & Geisler, W. S. (1990). Multichannel texture analysis using localized spatial filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12, 55–73.

Bowen, R. W. (1995). Isolation and interaction of on and off pathways in human vision: Pattern-polarity effects on contrast discrimination. *Vision Research*, 35, 2479–2490.

Brainard, D., Pelli, D. & Robson, T. (2002). Display characterisation. In Hornak, J. (ed.), *Encyclopedia of imaging science and technology*, (172–188). Wiley.

Burr, D. & Morrone, C. (1994). The role of features in structuring visual images. In Bock, G. & Goode, J. (eds.), *Higher-order processing in the visual system*, (129–146). John Wiley and Sons Ltd, Chichester.

Burr, D. C. (1980). Sensitivity to spatial phase. *Vision Research*, 20, 391–396.

Burr, D. C. & Morrone, M. C. (1993). Impulse-response functions for chromatic and achromatic stimuli. *Journal of the Optical Society of America A*, 10, 1706–1713.

Burr, D. C., Morrone, M. C. & Spinelli, D. (1989). Evidence for edge and bar detectors in human vision. *Vision Research*, 29, 419–431.

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

Campbell, F. W. & Green, D. G. (1965). Optical and retinal factors affecting visual resolution. *Journal of Physiology*, 181, 576–593.

Campbell, F. W. & Gubisch, R. W. (1966). Optical quality of the human eye. *Journal of Physiology*, 186, 558–578.

Campbell, F. W. & Kulikowski, J. J. (1966). Orientational selectivity of the human visual system. *The Journal of Physiology*, 187, 437–445.

Campbell, F. W. & Robson, J. G. (1968). Application of fourier analysis to the visibility of gratings. *Journal of Physiology*, 197, 551–556.

Cannon, M. W. & Fullemkamp, S. C. (1991). Spatial interactions in apparent contrast - inhibitory effects among grating patterns of different spatial-frequencies, spatial positions and orientations. *Vision Research*, 31, 1985–1998.

Canny, J. (1986). A computational approach to edge-detection. *IEEE Transactions On Pattern Analysis and Machine Intelligence*, 8, 679–698.

Charman, W. N. & Watrasiewicz, B. M. (1964). Mach effect associated with microscope images. *Journal of the Optical Society of America*, 54, 791–794.

Chen, J. S. & Medioni, G. (1989). Detection, localization, and estimation of edges. *IEEE Transactions On Pattern Analysis and Machine Intelligence*, 11, 191–198.

Chichilnisky, E. J. & Kalmar, R. S. (2002). Functional asymmetries in on and off ganglion cells of primate retina. *Journal of Neuroscience*, 22, 2737–2747.

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*, 86, 9631–9635.

Clark, J. (1989). Authenticating edges produced by zero-crossing algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11, 43–57.

Cohn, T. E. & Lasley, D. J. (1975). Spatial summation of foveal increments and decrements. *Vision Research*, 15, 389–399.

Cohn, T. E. & Lasley, D. J. (1985). Site of the accelerating nonlinearity underlying luminance-change detection. *Journal of the Optical Society of America A*, 2, 202–205.

Collins, J., Imhoff, T. & Greg, P. (1996). Noise-enhanced tactile sensation. *Nature*, 383, 770.

Cornsweet, T. (1970). *Visual perception.* Academic Press, New York.

Creelman, C. & Macmillan, N. (1979). Auditory phase and frequency discrimination: a comparison of nine procedures. *Journal of Experimental Psychology: Human Perception and Performance*, 5, 146–156.

Croner, L. J. & Kaplan, E. (1995). Receptive fields of p-ganglion and m-ganglion cells across the primate retina. *Vision Research*, 35, 7–24.

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

De Monasterio, F. M., Gouras, P. & Tolhurst, D. J. (1975). Trichromatic colour opponency in ganglion cells of the rhesus monkey retina. *Journal of Neurophysiology*, 251, 197–216.

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

De Valois, R., Thorell, L. & Albrecht, D. (1985). Periodicity of striate-cortex-cell receptive fields. *Journal of the Optical Society of America A*, 2, 1115–1122.

De Valois, R. L. & Cottaris, N. P. (1998). Inputs to directionally selective simple cells in macaque striate cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 95, 14488–14493.

Derrington, A. M. & Lennie, P. (1984). Spatial and temporal contrast sensitivities of neurons in lateral geniculate nucleus of macaque. *Journal of Physiology-London*, 357, 219–240.

Dreher, B. & Sanderson, K. (1973). Receptive field analysis - responses to moving visual contours by single lateral geniculate neurons in cat. *Journal of Physiology-London*, 234, 95–118.

Eckstein, M. P. & Ahumada, Albert J., J. (2002). Classification images: A tool to analyze visual strategies. *Journal of Vision*, 2, i–i.

Efron, B. (1979). 1977 rietz lecture - bootstrap methods - another look at the jackknife. *Annals of Statistics*, 7, 1–26.

Efron, B. (1981). Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods. *Biometrika*, 68, 589–599.

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. (1984). Functional characteristics and diversity of cat retinal ganglion cells: basic characteristics and quantitative description. *Investigative Ophthalmology Visual Science*, 25, 250–267.

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

Enroth-Cugell, C., Robson, J. G., Schweitzer-Tong, D. E. & Watson, A. B. (1983). Spatio-temporal interactions in cat retinal ganglion cells showing linear spatial summation. *The Journal of Physiology*, 341, 279–307.

Enroth-Cugell, C. & Shapley, R. M. (1973a). Adaptation and dynamics of cat retinal ganglion cells. *The Journal of Physiology*, 233, 271–309.

Enroth-Cugell, C. & Shapley, R. M. (1973b). Flux, not retinal illumination, is what cat retinal ganglion cells really care about. *The Journal of Physiology*, 233, 311–326.

Fechner, G. (1860). *Elemente der psychophysik*. Breitkopf und Härtel.

Ferster, D. & Miller, K. (2000). Neural mechanisms of orientation selectivity in the visual cortex. *Annual Review of Neuroscience*, 23, 441–471.

Field, D. J. & Nachmias, J. (1984). Phase reversal discrimination. *Vision Research*, 24, 333–340.

Finney, D. (1971). *Probit Analysis*. Cambridge University Press, London, 3rd edn.

Fiorentini, A. (1972). Mach band phenomena. In Jameson, D. & Hurvich, L. (eds.), *Handbook of Sensory Physiology*, vol. 7, chap. 8, (188–201). Springer, Berlin.

Fiorentini, A., Baumgartner, G., Magnussen, S., Schiller, P. & Thomas, J. (1990). The perception of brightness and darkness: Relations to neuronal receptive fields. In Spillmann, L. & Werner, J. (eds.), *Visual perception: The neurophysiological foundations*, chap. 7, (129–161). Academic Press, San Diego.

Fleet, D. J., Wagner, H. & Heeger, D. J. (1996). Neural encoding of binocular disparity: Energy models, position shifts and phase shifts. *Vision Research*, 36, 1839–1857.

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

Foley, J. M. & Legge, G. E. (1981). Contrast detection and near-threshold discrimination in human vision. *Vision Research*, 21, 1041–1053.

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*, C 24, 616–628.

Freeman, R. D. & Ohzawa, I. (1990). On the neurophysiological organization of binocular vision. *Vision Research*, 30, 1661–1676.

Geisler, W. S., Perry, J. S., Super, B. J. & Gallogly, D. P. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research*, 41, 711–724.

Georgeson, M. & Wallis, S. (2008). Seeing light vs dark lines: Psychophysical performance is based on separate channels, limited by noise and uncertainty. *Journal of Vision*, 8, 821.

Georgeson, M. A. (1984). Eye-movements, afterimages and monocular rivalry. *Vision Research*, 24, 1311–1319.

Georgeson, M. A. (1992). Human vision combines oriented filters to compute edges. *Proceedings of the Royal Society of London Series B-Biological Sciences*, 249, 235–245.

Georgeson, M. A. (1994). From filters to features: location, orientation, contrast and blur. In Bock, G. & Goode, J. (eds.), *Higher-order processing in the visual system*, (147–169). John Wiley and Sons Ltd, Chichester.

Georgeson, M. A. (1996). Perceived structure of plaids implies variable combination of oriented filters in edge-finding. In Rogowitz, B. & Allebach, J. (eds.), *Human Vision and Electronic Imaging*, vol. 2657, (175–189). Proceedings SPIE.

Georgeson, M. A. (2006). Bars and edges: A multi-scale gaussian derivative model for feature coding in human vision. *Journal of Vision*, 6, 191.

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. & Georgeson, J. M. (1987). Facilitation and masking of briefly presented gratings: time-course and contrast dependence. *Vision Research*, 27, 369–379.

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

Georgeson, M. A., May, K. A., Freeman, T. C. A. & Hesse, G. S. (2007). From filters to features: Scale-space analysis of edge and blur coding in human vision. *Journal of Vision*, 7, 1–21.

Georgeson, M. A. & Turner, R. S. E. (1985). Afterimages of sinusoidal, square-wave and compound gratings. *Vision Research*, 25, 1709–1720.

Gerrits, H. J. M. & Vendrik, A. J. H. (1970). Simultaneous contrast, filling-in process and information processing in man's visual system. *Experimental Brain Research*, 11, 411–430.

Gescheider, G. (1997). *Psychophysics: the fundamentals.* Lawrence Erlbaum Associates, New Jersey, 3rd edn.

Goris, R. L. T., Zaenen, P. & Wagemans, J. (2008). Some observations on contrast detection in noise. *Journal of Vision*, 8, 1–15.

Gourevitch, V. & Galanter, E. (1967). A significance test for one-parameter isosensitivity functions. *Psychometrika*, 32, 25–33.

Graham, N. (1977). Visual detection of aperiodic spatial stimuli by probability summation among narrowband channels. *Vision Research*, 17, 637–652.

Graham, N. (1980). Spatial frequency channels in human vision: detecting edges without edge detectors. In Harris, C. (ed.), *Visual Coding and Adaptability*, (215–262). Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Graham, N. & Nachmias, J. (1971). Detection of grating patterns containing two spatial frequencies: A comparison of single-channel and multiple-channels models. *Vision Research*, 11, 251 – 259.

Graham, N., Robson, J. & Nachmias, J. (1978). Grating summation in fovea and periphery. *Vision Research*, 18, 815 – 825.

Green, D. & Swets, J. (1966). *Signal detection theory and psychophysics.* John Wiley and Sons, New York.

Gur, M. & Syrkin, G. (1993). Color enhances mach bands detection threshold and perceived brightness. *Vision Research*, 33, 2313–2319.

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

Heggelund, P., Krekling, S. & Skottun, B. C. (1984). Spatial summation in subregions of simple-cell receptive fields in cat striate cortex as a function of slit length. *The Journal of Physiology*, 352, 327–337.

Henning, G. B., Millar, R. W. & Hill, N. J. (2000). Detection of incremental and decremental bars at different locations across mach bands and related stimuli. *Journal of the Optical Society of America*, 17, 1147–1159.

Henning, G. B. & Wichmann, F. A. (2007). Some observations on the pedestal effect. *Journal of Vision*, 7, 1–15.

Hering, E. (1964). *Outlines of a theory of the light sense.* (trans. by Hurvich, L.M. and Jameson, D) Harvard University Press, Massachusetts.

Hess, R. F. & Holliday, I. E. (1992). The coding of spatial position by the human visual system: Effects of spatial scale and contrast. *Vision Research*, 32, 1085–1097.

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., Schmid, K. L., Dumoulin, S. O., Field, D. J. & Brinkworth, D. R. (2006). What image properties regulate eye growth? *Current Biology*, 16, 687–691.

Hesse, G. S. & Georgeson, M. A. (2005). Edges and bars: where do people see features in 1-d images? *Vision Research*, 45, 507–525.

Hibbard, P. B. (2008). Binocular energy responses to natural images. *Vision Research*, 48, 1427–1439.

Hirsch, J. A., Alonso, J.-M., Reid, R. C. & Martinez, L. M. (1998). Synaptic integration in striate cortical simple cells. *Journal of Neuroscience*, 18, 9517–9528.

Huang, P., Kingdom, F. & Hess, R. (2006). Only two phase mechanisms, ±cosine, in human vision. *Vision Research*, 46, 2069–2081.

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

Hubel, D. & Wiesel, T. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of physiology*, 160, 106–154.

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

Hubel, D. H. & Wiesel, T. N. (1977). Functional architecture of macaque monkey visual cortex. *Proceedings of the Royal Society of London Series B-Biological Sciences*, 198, 1–59.

Hung, C. P., Ramsden, B. M. & Roe, A. W. (2007). A functional circuitry for edge-induced brightness perception. *Nature Neuroscience*, 10, 1185–1190.

Jacobson, L. D., Gaska, J. P., Hai-Wen, C. & Pollen, D. A. (1993). Structural testing of multi-input linear-nonlinear cascade models for cells in macaque striate cortex. *Vision Research*, 33, 609–626.

Jesteadt, W. & Bilger, R. C. (1974). Intensity and frequency discrimination in one- and two-interval paradigms. *The Journal of the Acoustical Society of America*, 55, 1266–1276.

Jesteadt, W. & Sims, S. L. (1975). Decision processes in frequency discrimination. *The Journal of the Acoustical Society of America*, 57, 1161–1168.

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.

Kagan, I., Gur, M. & Snodderly, D. M. (2002). Spatial organization of receptive fields of v1 neurons of alert monkeys: Comparison with responses to gratings. *Journal of Neurophysiology*, 88, 2557–2574.

Kelly, D. H. & Martinez-Uriegas, E. (1993). Measurements of chromatic and achromatic afterimages. *Journal of the Optical Society of America A*, 10, 29–37.

Kingdom, F. & Moulden, B. (1992). A multichannel approach to brightness coding. *Vision Research*, 32, 1565–1582.

Kingdom, F. A. & Whittle, P. (1996). Contrast discrimination at high contrasts reveals the influence of local light adaptation on contrast processing. *Vision Research*, 36, 817 – 829.

Klein, S. A. (1985). Double-judgment psychophysics: problems and solutions. *Jounal of the Optical Society of America A*, 2, 1560–1585.

Knau, H. (2000). Thresholds for detecting slowly changing ganzfeld luminances. *Journal of the Optical Society of America A*, 17, 1382–1387.

Koenderink, J. J. (1984). The structure of images. *Biological Cybernetics*, 50, 363–370.

Kontsevich, L. L., Chen, C. C. & Tyler, C. W. (2002). Separating the effects of response nonlinearity and internal noise psychophysically. *Vision Research*, 42, 1771–1784.

Kontsevich, L. L. & Tyler, C. W. (1999a). Distraction of attention and the slope of the psychometric function. *Journal of the Optical Society of America A*, 16, 217–222.

Kontsevich, L. L. & Tyler, C. W. (1999b). Nonlinearities of near-threshold contrast transduction. *Vision Research*, 39, 1869 – 1880.

Korn, A. F. (1988). Toward a symbolic representation of intensity changes in images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10, 610–625.

Kovesi, P. (2000). Phase congruency: a low-level image invariant. *Psychological Research*, 64, 136–148.

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

Kulikowski, J., Abadi, R. & King-Smith, P. (1973). Orientational selectivity of grating and line detectors in human vision. *Vision Research*, 13, 1479–1486.

Kulikowski, J. J. & King-Smith, P. E. (1973). Spatial arrangement of line, edge and grating detectors revealed by subthreshold summation. *Vision Research*, 13, 1455–1478.

Lasley, D. J. & Cohn, T. E. (1981). Why luminance discrimination may be better than detection. *Vision Research*, 21, 273 – 278.

Lawden, M. C. (1983). An investigation of the ability of the human visual system to encode spatial phase relationships. *Vision Research*, 23, 1451–1463.

Legge, G. (1979). Spatial frequency masking in human vision: binocular interactions. *Journal of the Optical Society of America*, 69, 838–847.

Legge, G. E. (1978a). Space domain properties of a spatial frequency channel in human vision. *Vision Research*, 18, 959 – 969.

Legge, G. E. (1978b). Sustained and transient mechanisms in human-vision - temporal and spatial properties. *Vision Research*, 18, 69–81.

Legge, G. E. (1984). Binocular contrast summation 1. detection and discrimination. *Vision Research*, 24, 373–383.

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

Legge, G. E. & Kersten, D. (1983). Light and dark bars - contrast discrimination. *Vision Research*, 23, 473–483.

Legge, G. E., Kersten, D. & Burgess, A. E. (1987). Contrast discrimination in noise. *Journal of the Optical Society of America A*, 4, 391–404.

Leshowitz, B. (1969). Comparison of roc curves from one- and two-interval rating-scale procedures. *The Journal of the Acoustical Society of America*, 46, 399–402.

Leshowitz, B., Taub, H. & Raab, D. (1968). Visual detection of signals in the presence of continuous and pulsed backgrounds. *Perception and Psychophysics*, 4, 207–213.

Levi, D. M. & Klein, S. A. (1990a). Equivalent intrinsic blur in amblyopia. *Vision Research*, 30, 1995–2022.

Levi, D. M. & Klein, S. A. (1990b). Equivalent intrinsic blur in spatial vision. *Vision Research*, 30, 1971–1993.

Lindeberg, T. (1990). Scale-space for discrete signals. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12, 234–254.

Lindeberg, T. (1993). On scale selection for differential operators. *Proceedings of the 8th Scandinavian Conference on Image Analysis*, May, 857–866.

Lindeberg, T. (1994). *Scale-space theory in computer vision.* Kluwer, Dordrecht.

Lindeberg, T. (1996). Edge detection and ridge detection with automatic scale selection. *Proceedings CVPR-96*, (465–471).

Lindeberg, T. (1998). Edge detection and ridge detection with automatic scale selection. *International Journal of Computer Vision*, 30, 117–154.

Livingstone, M. S. (1998). Mechanisms of direction selectivity in macaque v1. *Neuron*, 20, 509 – 526.

Lowry, E. M. & DePalma, J. J. (1961). Sine-wave response of the visual system 1. the mach phenomenon. *Journal of the Optical Society of America*, 51, 740–746.

Lu, Z. L. & Dosher, B. A. (2008). Characterizing observers using external noise and observer models: Assessing internal representations with external noise. *Psychological Review*, 115, 44–82.

Lukaszewski, J. S. & Elliot, D. N. (1962). Auditory threshold as a function of forced-choice technique, feedback, and motivation. *Journal of the Acoustical Society of America*, 34, 223–263.

Macleod, I. (1970). On finding structure in pictures. In Kaneff, K. (ed.), *Picture Language Machines*, (231–256). Academic Press, London.

Macleod, I. D. G. (1972). Techniques for edge detection. *Proceedings of the Institute of Electrical and Electronics Engineers*, 60, 344.

Macmillan, N. & Creelman, C. (2005). *Detection theory: a user's guide.* Lawrence Erlbaum Associates, New Jersey.

Maffei, L. (1985). Complex cells control simple cells. In Rose, D. & Dobson, V. (eds.), *Models of the visual cortex*, (334–340). Wiley, New York.

Maffei, L. & Fiorentini, A. (1973). Visual cortex as a spatial frequency analyzer. *Vision Research*, 13, 1255–1267.

Maffei, L., Morrone, C., Pirchio, M. & Sandini, G. (1979). Responses of visual cortical cells to periodic and non-periodic stimuli. *The Journal of Physiology*, 296, 27–47.

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

Mallat, S. & Zhong, S. (1992). Characterization of signals from multiscale edges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14, 710–732.

Markowitz, J. & Swets, J. (1967). Factors affecting the slope of empirical roc curves: Comparison of binary and rating responses. *Perception and Psychophysics*, 2, 91–100.

Marr, D. (1976). Early processing of visual information. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, 275, 483–519.

Marr, D. (1982). *Vision*. W.H.Freeman and Co., New York.

Marr, D. & Hildreth, E. (1980). Theory of edge-detection. *Proceedings of the Royal Society of London Series B-Biological Sciences*, 207, 187–217.

Marr, D., Ullman, S. & Poggio, T. (1979). Bandpass channels, zero-crossings, and early visual information-processing. *Journal of the Optical Society of America*, 69, 914–916.

Martinez, L. M. & Alonso, J.-M. (2001). Construction of complex receptive fields in cat primary visual cortex. *Neuron*, 32, 515–525.

Martinez, L. M., Wang, Q. B., Reid, R. C., Pillai, C., Alonso, J. M., Sommer, F. T. & Hirsch, J. A. (2005). Receptive field structure varies with layer in the primary visual cortex. *Nature Neuroscience*, 8, 372–379.

Mata, M. L. & Ringach, D. L. (2005). Spatial overlap of on and off subregions and its relation to response modulation ratio in macaque primary visual cortex. *Journal of Neurophysiology*, 93, 919–928.

Mather, G. & Morgan, M. (1986). Irradiation - implications for theories of edge localization. *Vision Research*, 26, 1007–1015.

May, K. & McIlhagga, W. (2009). Probing edge-blur perception with reverse correlation. *Perception*, 38, 621.

May, K. A. (2003). *Edge coding in human vision: a psychophysical and computational investigation*. Ph.D. thesis, Aston University.

May, K. A. & Georgeson, M. A. (2007a). Added luminance ramp alters perceived edge blur and contrast: A critical test for derivative-based models of edge coding. *Vision Research*, 47, 1721–1731.

May, K. A. & Georgeson, M. A. (2007b). Blurred edges look faint, and faint edges look sharp: The effect of a gradient threshold in a multi-scale edge coding model. *Vision Research*, 47, 1705–1720.

Mayer, M. J. & Tyler, C. W. (1986). Invariance of the slope of the psychometric function with spatial summation. *Journal of the Optical Society of America A*, 3, 1166–1172.

McGraw, P. V., Levi, D. M. & Whitaker, D. (1999). Spatial characteristics of the second-order visual pathway revealed by positional adaptation. *Nature Neuroscience*, 2, 479–484.

Meese, T. S. (1993). *Feature coding in human pattern vision.* Ph.D. thesis, Bristol University.

Meese, T. S. (1995). Phase-reversal discrimination in one and 2 dimensions - performance is limited by spatial repetition, not spatial-frequency content. *Vision Research*, 35, 2157–2167.

Meese, T. S. (2004). Area summation and masking. *Journal of Vision*, 4, 930–943.

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. (1991). Edge computation in human vision - what controls the combining of oriented filters. *Perception*, 20, 81–81.

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

Meese, T. S. & Summers, R. J. (2009). Neuronal convergence in early contrast vision: Binocular summation is followed by response nonlinearity and area summation. *Journal of Vision*, 9, 1–16.

Meese, T. S., Summers, R. J., Holmes, D. J. & Wallis, S. A. (2007). Contextual modulation involves suppression and facilitation from the center and the surround. *Journal of Vision*, 7, 1–21.

Meese, T. S. & Williams, C. B. (2000). Probability summation for multiple patches of luminance modulation. *Vision Research*, 40, 2101–2113.

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 Series B-Biological Sciences*, 235, 221–245.

Morrone, M. C., Burr, D. C. & Ross, J. (1994). Illusory brightness step in the chevreul illusion. *Vision Research*, 34, 1567–1574.

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.

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

Mullikin, W. H., Jones, J. P. & Palmer, L. A. (1984). Periodic simple cells in cat area 17. *Journal of Neurophysiology*, 52, 372–387.

Mumford, D., Kosslyn, S. M., Hillger, L. A. & Herrnstein, R. J. (1987). Discriminating figure from ground - the role of edge-detection and region growing. *Proceedings of the National Academy of Sciences of the United States of America*, 84, 7354–7358.

Nachmias, J. (1967). Effect of exposure duration on visual contrast sensitivity with square-wave gratings. *Journal of the Optical Society of America*, 57, 421–427.

Nachmias, J. (1981). On the psychometric function for contrast detection. *Vision Research*, 21, 215 – 223.

Nachmias, J. & Kocher, E. C. (1970). Visual detection and discrimination of luminance increments. *Journal of the Optical Society of America*, 60, 382–389.

Nachmias, J. & Sansbury, R. V. (1974). Grating contrast - discrimination may be better than detection. *Vision Research*, 14, 1039–1042.

Nachmias, J. & Weber, A. (1975). Discrimination of simple and complex gratings. *Vision Research*, 15, 217 – 223.

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

Ohzawa, I., DeAngelis, G. C. & Freeman, R. D. (1990). Stereoscopic depth discrimination in the visual-cortex - neurons ideally suited as disparity detectors. *Science*, 249, 1037–1041.

Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America A*, 2, 1508–1532.

Pessoa, L. (1996). Mach bands: how many models are possible? recent experimental findings and modeling attempts. *Vision Research*, 36, 3205–3227.

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

Poirier, F. & Gurnsey, R. (2005). Non-monotonic changes in performance with eccentricity modeled by multiple eccentricity-dependent limitations. *Vision Research*, 45, 2436–2448.

Pollen, D. & Ronner, S. (1981). Phase-relationships between adjacent simple cells in the visual cortex. *Science*, 212, 1409–1411.

Pynn, C. T., Braida, L. D. & Durlach, N. I. (1972). Intensity perception. iii. resolution in small-range identification. *The Journal of the Acoustical Society of America*, 51, 559–566.

Ratliff, F. (1965). *Mach bands: quantitative studies on neural networks in the retina.* Holden-Day, San Francisco.

Ratliff, F. (1984). Why mach bands are not seen at the edges of a step. *Vision Research*, 24, 163–165.

Ratliff, F. & Hartline, H. K. (1959). The responses of limulus optic nerve fibers to patterns of illumination on the receptor mosaic. *Journal of General Physiology*, 42, 1241–1255.

Ratliff, F., Milkman, N. & Rennert, N. (1983). Attenuation of mach bands by adjacent stimuli. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences*, 80, 4554–4558.

Richter, J. & Ullman, S. (1986). Nonlinearities in cortical simple cells and the possible detection of zero crossings. *Biological Cybernetics*, 53, 195–202.

Ringach, D. L. (2002). Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *Journal of Neurophysiology*, 88, 455–463.

Roberts, L. (1965). Machine perception of three dimensional solids. In Tippett, J., Berkowitz, D., Clapp, L., Koester, C. & Vanderburgh, A. (eds.), *Optical and electro-optical information processing*, (159–197). Massachusetts Institue of Technology Press, Massachusetts.

Robson, J. (1975). Receptive fields: neural representation of the spatial and intensive attributes of the visual image. In Carterette, E. & Friedman, M. (eds.), *Handbook of perception*, vol. 5, (81–116). Academic Press, New York.

Robson, J. (1983). Frequency domain visual processing. In Braddick, O. & Sleigh, A. (eds.), *Physical and biological processing of images*, vol. 11, (73–87). Springer-Verlag, Berlin.

Robson, J. & 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. & Stone, J. (1965). Analysis of receptive field of cat retinal ganglion cells. *Journal of Physiology*, 28, 833–849.

Rosenfeld, A. (1970). A nonlinear edge detection technique. *Proceedings of the Institute of Electrical and Electronics Engineers*, 58, 814–816.

Ross, J., Holt, J. J. & J Johnstone, J. R. (1981). High-frequency limitations on mach bands. *Vision Research*, 21, 1165–1167.

Ross, J., Morrone, M. C. & Burr, D. C. (1989). The conditions under which mach bands are visible. *Vision Research*, 29, 699–715.

Sachs, M., Nachmias, J. & Robson, J. (1971). Spatial frequency channels in human vision. *Journal of the Optical Society of America*, 61, 1176–1186.

Saito, H.-A. & Fukada, Y. (1986). Gain control mechanisms in x- and y-type retinal ganglion cells of the cat. *Vision Research*, 26, 391 – 408.

Schiller, P., Sandell, J. & Maunsell, J. (1986). Functions of the on and off channels of the visual-system. *Nature*, 322, 824–825.

Schiller, P. H., Finlay, B. L. & Volman, S. F. (1976). Quantitative studies of single-cell properties in monkey striate cortex. i. spatiotemporal organization of receptive fields. *Journal of Neurophysiology*, 39, 1288–1319.

Schulman, A. J. & Mitchell, R. R. (1966). Operating characteristics from yes-no and forced-choice procedures. *The Journal of the Acoustical Society of America*, 40, 473–477.

Schumer, R. A. & Movshon, J. A. (1984). Length summation in simple cells of cat striate cortex. *Vision Research*, 24, 565 – 571.

Scott-Samuel, N. E. & Georgeson, M. A. (1999). Does early non-linearity account for second-order motion? *Vision Research*, 39, 2853–2865.

Shapley, R. (1986). The importance of contrast for the activity of single neurons, the vep and perception. *Vision Research*, 26, 45–61.

Shapley, R. & Perry, V. H. (1986). Cat and monkey retinal ganglion-cells and their visual functional roles. *Trends in Neurosciences*, 9, 229–235.

Shapley, R. M. & Tolhurst, D. J. (1973). Edge detectors in human vision. *Journal of Physiology-London*, 229, 165–183.

Skottun, B. C., De Valois, R. L., Grosof, D. H., Movshon, J. A., Albrecht, D. G. & Bonds, A. B. (1991). Classifying simple and complex cells on the basis of response modulation. *Vision Research*, 31, 1078–1086.

Solomon, J. A. (2007). Intrinsic uncertainty explains second responses. *Spatial Vision*, 20, 45–60.

Solomon, J. A., Lavie, N. & Morgan, M. J. (1997). Contrast discrimination function: Spatial cuing effects. *Journal of the Optical Society of America A*, 14, 2443–2448.

Stromeyer, C. & Klein, S. (1974). Spatial frequency channels in human vision as asymmetric (edge) mechanisms. *Vision Research*, 14, 1409–1420.

Stromeyer, C. F. & Julesz, B. (1972). Spatial frequency masking in vision: critical bands and spread of masking. *Journal of the Optical Society of America*, 62, 1221–1232.

Sullivan, G., Georgeson, M. & Oatley, K. (1972). Channels for spatial frequency selection and detection of single bars by human visual system. *Vision Research*, 12, 383–394.

Summers, R. J. & Meese, T. S. (2009). The influence of fixation points on contrast detection and discrimination of patches of grating: Masking and facilitation. *Vision Research*, 49, 1894–1900.

Swets, J., Tanner, W. & Birdsall, T. G. (1961). Decision processes in perception. *Psychological Review*, 68, 301–340.

Swets, J. A., Shipley, E. F., McKey, M. J. & Green, D. M. (1959). Multiple observations of signals in noise. *The Journal of the Acoustical Society of America*, 31, 514–521.

Szulborski, R. G. & Palmer, L. A. (1990). The two-dimensional spatial structure of nonlinear subunits in the receptive fields of complex cells. *Vision Research*, 30, 249 – 254.

Tanner, W. & Swets, J. (1954). A decision-making theory of visual detection. *Psychological Review*, 61, 401–409.

Tanner, W. P. (1961). Physiological implications of psychophysical data. *Annals of the New York Academy of Sciences*, 89, 752–765.

ter Haar Romeny, B. (2003). *Front-end vision and multi-scale image analysis.* Kluwer, Dordrecht.

Thomas, J. P. (1966). Brightness variations in stimuli with ramp-like contours. *Journal of the Optical Society of America*, 56, 238–242.

Thurstone, L. (1927). A law of comparative judgement. *Psychological Review*, 34, 278–286.

Tolhurst, D. & Barfield, L. (1978). Interactions between spatial-frequency channels. *Vision Research*, 18, 951–958.

Tolhurst, D. & Dealy, R. (1975). The detection and identification of lines and edges. *Vision Research*, 15, 1367–1372.

Tolhurst, D. J. (1972a). Adaptation to square-wave gratings: inhibition between spatial frequency channels in the human visual system. *The Journal of Physiology London*, 226, 231–248.

Tolhurst, D. J. (1972b). On the possible existence of edge detector neurones in the human visual system. *Vision Research*, 12, 797–804.

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.

Torre, V. & Poggio, T. (1986). On edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8, 147–163.

Treisman, M. & Williams, T. C. (1984). A theory of criterion setting with an application to sequential dependencies. *Psychological Review*, 91, 68–111.

Troy, J. B. (1983). Spatio-temporal interaction in neurones of the cat's dorsal lateral geniculate nucleus. *The Journal of Physiology*, 344, 419–432.

Tyler, C. W. & Gorea, A. (1986). Different encoding mechanisms for phase and contrast. *Vision*, 26, 1073–1082.

Viemeister, N. (1970). Intensity discrimination: performance of three paradigms. *Perception and Psychophysics*, 8, 417–419.

von Bekesy, G. (1967). Mach band type lateral inhibition in different sense organs. *The Journal of General Physiology*, 50, 519–532.

von Bekesy, G. (1968). Brightness distribution across the mach bands measured with flicker photometry, and the linearity of sensory nervous interaction. *Journal of the Optical Society of America*, 58, 1–8.

von der Heydt, R., Peterhans, E. & Dursteler, M. (1992). Periodic pattern-selective cells in monkey visual cortex. *Journal of Neuroscience*, 12, 1416–1434.

von Helmholtz, H. (2000). *Helmholtz's treatise on physiological optics*, vol. 2. (J. Southall, Trans) Thoemmes Press, Bristol. (1911).

Wallis, S. & Georgeson, M. (2007a). Mach edges: a key role for 3rd derivative filters in spatial vision. *Perception*, 36, 314–315.

Wallis, S. & Georgeson, M. (2007b). Seeing light vs dark lines: Psychophysical performance is based on separate channels, limited by noise and uncertainty. *Perception*, 37, 315.

Wallis, S. & Georgeson, M. (2007c). Third-derivative filters predict edge locations in spatial vision. *Perception*, 36, 43.

Wallis, S. A. & Georgeson, M. A. (2009). Mach edges: Local features predicted by 3rd derivative spatial filtering. *Vision Research*, 49, 1886–1893.

Watrasiewicz, B. M. (1966). Some factors affecting the appearance of the mach bands. *Journal of the Optical Society of America*, 56, 499–503.

Watson, A. B. & Robson, J. G. (1981). Discrimination at threshold - labeled detectors in human vision. *Vision Research*, 21, 1115–1122.

Watson, C., Kellogg, S., Kawanishi, D. & Lucas, P. (1973). The uncertain response in detection-oriented psychophysics. *Journal of Experimental Psychology*, 99, 180–185.

Watt, R. (1988). *Visual processing, computational, psychophysical and cognitive research.* Lawrence Erlbaum Associates, Hove.

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.

Westhiemer, G. (2007). Irradiation, border location, and the shifted-chessboard pattern. *Perception*, 36, 483–493.

Whiteside, J. A. & Davidson, M. L. (1971). Symmetrical appearance of bright and dark mach bands from an exponential illumination gradient. *Journal of the Optical Society of America*, 61, 958–961.

Wichmann, F. A. & Hill, N. J. (2001a). The psychometric function: I. fitting, sampling, and goodness of fit. *Perception and Psychophysics*, 63, 1293–1313.

Wichmann, F. A. & Hill, N. J. (2001b). The psychometric function: Ii. bootstrap-based confidence intervals and sampling. *Perception and Psychophysics*, 63, 1314–1329.

Wiesel, T. N. (1960). Receptive fields of ganglion cells in the cat's retina. *The Journal of Physiology*, 153, 583–594.

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.

Williams, D. W. & Wilson, H. R. (1983). Spatial-frequency adaptation affects spatial-probability summation. *Journal of the Optical Society of America*, 73, 1367–1371.

Williams, P. E. & Shapley, R. M. (2007). A dynamic nonlinearity and spatial phase specificity in macaque v1 neurons. *Journal of Neuroscience*, 27, 5706–5718.

Wilson, H. R. & Bergen, J. R. (1979). 4-mechanism model for threshold spatial vision. *Vision Research*, 19, 19–32.

Wilson, H. R. & Gelb, D. J. (1984). Modified line-element theory for spatial-frequency and width discrimination. *Journal of the Optical Society of America A*, 1, 124–131.

Wilson, H. R. & Giese, S. C. (1977). Threshold visibility of frequency gradient patterns. *Vision Research*, 17, 1177–1190.

Wilson, H. R., Mcfarlane, D. K. & Phillips, G. C. (1983). Spatial frequency tuning of orientation selective units estimated by oblique masking. *Vision Research*, 23, 873–882.

Witkin, A. (1983). Scale-space filtering. In Bundy, A. (ed.), *Proceedings of the 8th international joint conference on artificial intelligence*, vol. 2, (1019–1022). William Kauffman, Los Altos.

Young, R. A. (1985). The gaussian derivative theory of spatial vision: analysis of cortical cell receptive field line-weighting profiles. Tech. rep., General Motors Research Laboratories, Warren, Mich.

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.

Yu, C., Klein, S. A. & Levi, D. M. (2002). Facilitation of contrast detection by cross-oriented surround stimuli and its psychophysical mechanisms. *Journal of Vision*, 2, 243–255.

Yu, C., Klein, S. A. & Levi, D. M. (2004). Perceptual learning in contrast discrimination and the (minimal) role of context. *Journal of Vision*, 4, 169–182.

Yuille, A. L. & Poggio, T. A. (1986). Scaling theorems for zero crossings. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8, 15–25.

Zanker, J. & Meso, A. (2008). On the role of fourier spectra in the separation of transparent motion components. *Perception*, 37, 959.

Zeng, F., Fu, Q. & Morse, R. (2000). Human hearing enhanced by noise. *Brain Research*, 869, 251–255.

Zhang, W. & Bergholm, F. (1997). Multi-scale blur estimation and edge type classification for scene analysis. *International Journal of Computer Vision*, 24, 219–250.