

ON THE SUMMATION OF VISUAL NOISE

by

Christopher Patrick Taylor

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Department of Psychology, Neuroscience and Behaviour
McMaster University

Copyright © 2011 by Christopher Patrick Taylor

Abstract

On the summation of visual noise

Christopher Patrick Taylor

Doctor of Philosophy

Department of Psychology, Neuroscience and Behaviour

McMaster University

2011

What information is used by the visual system to detect patterns? A standard model hypothesizes that both spatial frequency and orientation information are processed by independent channels, meaning there is no summation among channels. Despite consensus among researchers on how the visual system sums spatial frequency and orientation information there are data in the literature (Kersten, 1987) that ostensibly contradict the standard model. To resolve this conflict, we measured the efficiency of spatial frequency and orientation of filtered noise. To learn what information the visual system uses when detecting filtered noise, we applied a technique that can determine the information used to detect and discriminate filtered visual noise. In Chapter 2 the detection of spatial frequency filtered noise is not only efficient but remains so with stimulus uncertainty and extremely brief (10ms) stimulus duration. When the spatial frequency channel used was measured, we found a fixed bandwidth channel as the spatial frequency of the pattern was increased. To test the standard model, we implemented simulations of the standard model and contrary to the interpretation, the standard model could predict detection of spatial frequency filtered noise. Chapter 3 used spatial frequency filtered noise to relate the detection and discrimination of filtered visual noise. A simple rule relates what information observers use to detect and discriminate spatial frequency filtered noise. Chapter 4 extends the work of Chapter 2 to orientation information and found that orientation filtered noise is detected efficiently. We again measured what information observers used and found that, unlike SF filtered noise, observers use orientation in a flexible or adjustable manner.

Descriptive Note

Doctor of Philosophy (2011)

Psychology, Neuroscience and Behaviour

McMaster University

Hamilton, Ontario

TITLE: On the summation of visual noise

AUTHOR: Christopher Patrick Taylor, Hon. B.Sc. (University College, University of Toronto)

SUPERVISORS: Professor Patrick J. Bennett & Professor Allison B. Sekuler

NUMBER OF PAGES: ix, 123

Acknowledgements

The work in this thesis has spanned many years through out the process there have been two constants, Patrick Bennett and Allison Sekuler. Who are fantastic mentors and great people. They have allowed me to have an experience in their lab that many graduate student peers of mine have envied. It is hard to put the many years into words but their supervision has allowed me to learn as if I were independent while always feeling as if I am part of a team. This does not seem to go far enough. I've become a better person for the experiences I have had and through their efforts, which have on many occasions gone far above and beyond what could reasonably be required of graduate supervisors. The role they've played in my life can not be over-stated.

Donna Waxman's influence as lab research assistant – she made even the most difficult data collection a breeze. As all science comes down to the data that are collected Donna's influence can be seen in almost each and every plot and table in this thesis.

This penultimate paragraph is a mere list of people, animals and things that all merit a tip of my hat for their help: Meghan McConnell, Caffeine, Ike O'Bean, Cat Chup, Dr. David McLeod, Clonazepam, Lamotrogine, Richard Murray, Jason Gold, Paul Fauvre, Kathy Murphy, Sam Hannah, Baseball, Kevin Quain and the Mad Bastards, Zahra Hussain, Carl Gaspar, Guillaume Rousselet, Jesse Husk, Stan Govenlock, Lisa Betts, Andrzej Pietrewicz, Julia Pietrus, Adélaïde de Heering, and John W. Duncan IV.

Love and thanks to my family: my wonderful, bright-eyed little sister Katherine (Charlie) Taylor; my Aunt Christina Taylor whose love and support saw me through both good times and bad; my father Miles Taylor; and to my mother Patricia Taylor whose love for me has no bounds.

Preface

Chapter 2 of this dissertation has been published in a peer-reviewed journal, The Journal of the Optical Society of America A. Chapter 4 has been submitted to Vision Research. Chapter 3 will be prepared for submission to a journal yet to be determined.

My advisors Patrick Bennett and Allison Sekuler are co-authors on the publication that resulted from Chapter 2. My role in conducting the research presented in all data chapters (2–4) consisted of collaborating with my advisors on formulating and refining the hypotheses, the experimental design, and preparation of the manuscripts. The tasks of implementing the experiments and simulations, data analysis and managing the observers each were predominantly left to the student. My advisors were available for meetings to address any problems, issues or concerns that stymied the student.

Contents

1	Introduction	1
	References	9
2	Spatial frequency summation in visual noise	14
2.1	Experiment 1	16
2.1.1	Methods	16
2.1.1.1	Observers	16
2.1.1.2	Apparatus	16
2.1.1.3	Stimuli	16
2.1.1.4	Procedures	17
2.1.2	Results	18
2.1.3	Discussion	20
2.2	Experiment 2	23
2.2.1	Methods	23
2.2.1.1	Observers	23
2.2.1.2	Apparatus	23
2.2.1.3	Stimuli	23

2.2.1.4	Procedures	24
2.2.2	Results	24
2.2.3	Discussion	24
2.3	Experiment 3	28
2.3.1	Methods	28
2.3.1.1	Observers	28
2.3.1.2	Apparatus	28
2.3.1.3	Stimuli	28
2.3.1.4	Procedure	29
2.3.2	Results	29
2.3.3	Discussion	29
2.4	Experiment 4	32
2.4.1	Methods	32
2.4.1.1	Observers	32
2.4.1.2	Apparatus	32
2.4.1.3	Stimuli	32
2.4.1.4	Procedures	33
2.4.1.5	Analysis	33
2.4.2	Results	33
2.4.3	Non-linear contributions to the classification image?	37
2.4.4	Discussion	38
2.5	Conclusion	41
	References	44

3	Detection and discrimination of visual noise	47
3.1	Experiment 3.1	51
3.1.1	Methods	51
3.1.1.1	Observers	51
3.1.1.2	Apparatus	51
3.1.1.3	Stimuli	52
3.1.1.4	Procedures	52
3.1.2	Results	53
3.1.3	Discussion	56
3.1.3.1	The Wilson-Gelb Model	56
3.2	Experiment 3.2	64
3.2.1	Methods	64
3.2.1.1	Observers	64
3.2.1.2	Apparatus	64
3.2.1.3	Stimuli	65
3.2.1.4	Procedures	65
3.2.2	Results	68
3.2.3	Discussion	69
3.2.3.1	Do the classification images predict the d' of human ob- servers?	71
3.2.4	The Wilson-Gelb Model and Classification Images	72
3.3	Summary	73
3.4	Conclusion	74
	References	74

4	Orientation summation in visual noise	77
4.1	Experiment 4.1	79
4.1.1	Methods	80
4.1.1.1	Observers	80
4.1.1.2	Apparatus	80
4.1.1.3	Stimuli	80
4.1.1.4	Procedure	81
4.1.2	White Noise Thresholds	81
4.1.3	Results	82
4.1.4	Discussion	88
4.2	Experiment 4.2	88
4.2.1	Methods	89
4.2.1.1	Observers	89
4.2.1.2	Apparatus	89
4.2.1.3	Stimuli	89
4.2.1.4	Procedures	89
4.2.2	Results	89
4.2.3	Analysis	93
4.2.4	Discussion	98
4.2.5	The Cigars and Donuts Model	98
4.2.6	Inhibition/Response Suppression in the Classification Image? . . .	101
4.3	Conclusion	102
	References	102

5 Conclusion	106
References	118

Chapter 1

Introduction

Background

Beginning in the 1950s and 1960s vision researchers turned to techniques adapted from engineering and communication theory [4, 22] to investigate the visual system [23, 24, 39, 41, 45]. The first attempts at characterizing the visual system tended to use models that characterized the visual system as a single channel.¹ Barlow [5] provides an excellent summary of the transformative effect that information/communication theory has had on psychology, both behavioral and physiological.

Schade [41] measured contrast sensitivity functions (CSFs) by asking observers to detect gratings that varied in contrast. Schade interpreted his results as support for the idea that the visual system is well-described by a single channel that passes visual information to later stages of the visual system. Measurements of the CSF however, are not definitive evidence that the operation of the visual system is best described by a single-channel model [48], but merely that the hypothesis is reasonable.

An alternative to the single-channel model hypothesis is that visual stimuli are encoded by an array of multiple independent channels [9, 20]. A key development in the

¹The channel concept originated in information/communication theory [43]. When a sender and receiver attempt to communicate there must be a medium to transmit the signal. A channel may be (but is not limited to being) a band of radio frequencies, a beam of light, the synapse, or a set neurons wired together. Attneave [2] introduced several important theoretical concepts, including channels, from information/communication theory into Psychology.

investigation into the multiple channel hypothesis was the application of Fourier Analysis to the ability of human observers to detect and discriminate sinusoidal gratings [9]. Debate surrounded the development of the details of multiple-channel models [29]. However, the standard model is one in which the model channels are tuned to relatively narrow ranges of spatial frequency and orientation [14, 17, 48, 50, 51, 52]. The important distinction between single-channel and multiple-channel models is not that the two models contain a different number of channels, but that the channels in the multiple-channel model are, at least in some circumstances, approximately independent. In a multiple-channel model with strict channel independence the activation of one channel does not influence the activation in any of the other channels in the ensemble. When viewed through the lens of channel independence, the difference between single- and multiple-channel models becomes a continuum rather than a categorical distinction. A strict single-channel model is equivalent to a multiple-channel model where the channels completely lack independence; the channels act together and are indistinguishable.

Evidence supporting the idea of channel independence has been obtained from adaptation [7, 35, 42], masking [8, 30, 38, 53], detection/discrimination double-judgement [28, 49], and summation experiments [18, 19, 20, 40]. Of the four lines of evidence, the experiments on spatial frequency and orientation summation form an important background and experimental motivation for the work described in this thesis. Therefore, a brief description of two types of summation experiments will be provided in the remainder of this section.

In a typical spatial frequency summation experiment, contrast detection thresholds are measured for two component sinusoidal gratings of frequency F and nF ($n \geq 0$) and a compound waveform consisting of the sum of the two sine wave gratings ($F + nF$). Summation is said to occur if the detection threshold for the compound is lower than threshold for either component. One form of summation, probability summation, has been defined by [20] as the expected improvement in detection threshold that results from the system having two independent detectors that each have an opportunity to detect the compound. When the spatial frequency of two gratings exceeds a difference of 1-2 octaves, as those shown in the example stimuli of Figure 1.1, probability summation is found. Summation exceeding what would be expected on the basis of probability summation has been found only when the spatial frequencies contained in the compound grating do not differ by more than 1-2 octaves [18, 19, 20, 40]. Consequently, the results of summation experiments are consistent with the idea that spatial frequencies that differ by 1-2 octaves are detected by independent mechanisms.

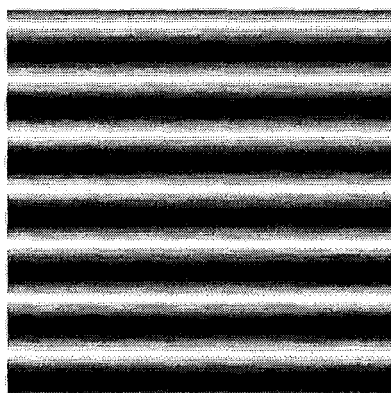
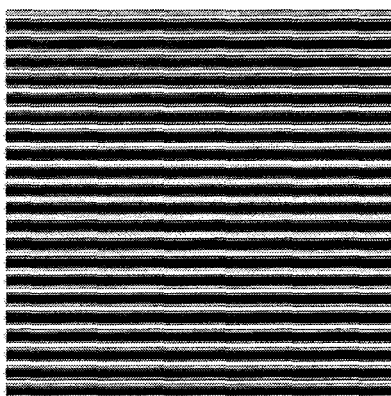
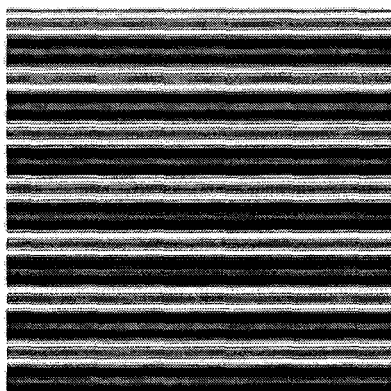
 f  $3f$  $f+3f$

Figure 1.1: Three example stimuli for a spatial frequency summation experiment. Two stimuli are component gratings, one of frequency f and its third harmonic. The bottom stimulus the compound of the two components $f + 3f$.

Analogous experiments have been conducted to support the idea that the visual system contains multiple, independent orientation channels [30, 46]. Sinusoidal gratings, such as those shown in Figure 1.1, have been used in orientation summation experiments [10]. In an orientation summation experiment, the frequency of the stimulus is kept constant and the orientation of the gratings is manipulated. The logic of the experiment is the same: contrast detection thresholds are measured for two component gratings of orientations θ and $\theta \pm \delta$. The variable δ represents the difference in orientation between the two gratings. If $\delta = 0$ the gratings have the same orientation and if $\delta = 90$ the gratings are orthogonal. A compound grating that is the sum of the two stimulus components is created and the threshold for the compound grating is measured. Typically, significant summation has been found only when the orientations of the two components are within $15\text{--}20^\circ$ [20], suggesting that orientations separated by more than $15\text{--}20^\circ$ are detected by independent mechanisms.

Context within Vision Research

The standard multiple-channel model of detection and discrimination does not exist in a vacuum within vision research. Multiple-channel models provide the foundation for experiments and models in areas of vision outside the sub-field of spatial vision. Issues in areas as diverse as letter recognition/reading [31, 36, 44], face recognition [12, 15, 16] and attention [3, 11, 13, 32, 47] have been addressed via standard multiple-channel models. The many applications of the standard model of spatial vision highlights the importance of ensuring the details of the standard model are correct. For example, if the application of the standard model to an attention task fails, researchers want to be certain that the model fails because it fails to capture an aspect of attentional processing and not because a detail of the standard model is incorrect.

Challenges to the Standard Model of Summation

Multiple channel models have provided satisfactory explanations of a wide range of psychophysical [20, 29, 48] and physiological [14] data. Nevertheless, several challenges have been posed to the standard multiple independent channel model. In particular, the assumption that channels are independent has been challenged by results from double-judgement [33, 34], masking [37], and adaptation experiments that used grating mixtures [26, 27]. These experiments (unlike summation experiments) all used supra-threshold patterns, and therefore they do not rule out the possibility that the independence as-

sumption is valid in detection experiments that use stimuli that (by definition) are barely visible. However, there is one detection experiment reported by Kersten [25] that also poses a challenge to the standard model. Prior to the work contained in this thesis, the challenge posed by Kersten’s results had not been addressed in the spatial vision literature. Kersten’s experiment provides a key motivation for the experiments conducted in this work, therefore, a detailed explanation of how the results from Kersten’s experiment conflict with the data and interpretation of summation experiments is provided in the following section.

Absolute Efficiency of Noise Detection

Kersten [25] measured detection thresholds for a stimulus that was presented in a background of white noise. The stimulus was itself visual noise that was filtered in a manner in which the number of spatial frequency components contained within the noise could be varied and examples of this stimulus are shown in Figure 1.3. This method, unlike previous summation experiments, allowed the number of spatial frequency components in the stimulus to be altered via a single parameter, spatial frequency bandwidth, over a large range.

Kersten then measured human observers’ contrast detection thresholds as a function of the spatial frequency bandwidth of the stimulus. Figure 1.2 shows the threshold versus bandwidth (TvB) function for one of Kersten’s observers. Detection threshold, expressed as root mean squared² (RMS) contrast, increased with increasing stimulus bandwidth.

The dashed grey line in Figure 1.2 is Equation 1.1:

$$c_{RMS} \propto (BW)^{\frac{1}{4}} \quad (1.1)$$

which says that RMS detection threshold is proportional to the quarter-root of the spatial frequency bandwidth. Equation 1.1, which provides a good fit to the TvB function at all but the narrowest bandwidth, corresponds to the pattern of data that would be produced by an ideal observer in this task [21, 25].

Equation 1.2 and Figure 1.4 show mathematically and schematically how the ideal observer for a noise detection task operates. The ideal detector sums the power at each

²The use of RMS contrast is common in visual psychophysics, particularly when expressing the contrast of complex stimuli, and has been found to be the best index of detectability of complex patterns such as natural images [6].

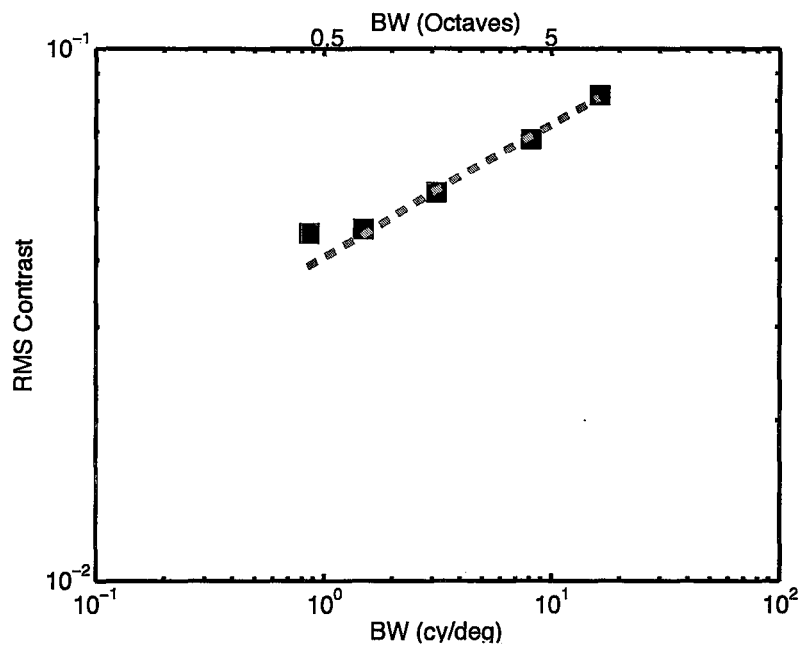


Figure 1.2: Data re-plotted from [25]. The squares are a human observer's contrast detection thresholds. The dashed grey line is a power function with a quarter-root slope that was fit to the data.

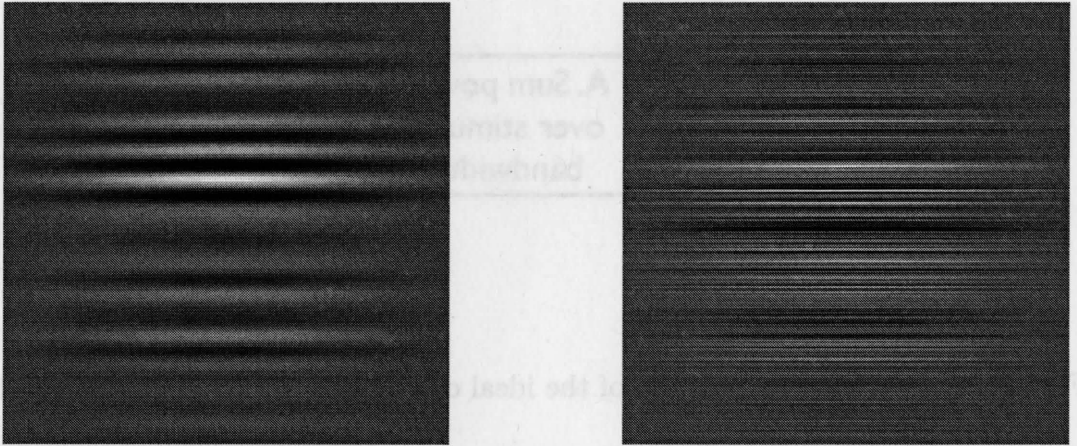


Figure 1.3: Filter noise stimuli similar to those used by [25].

frequency (f^2) in the signal bandwidth (BW_{low} to BW_{high}) resulting in the decision variable, P_S . The ideal observer is illustrated schematically in Figure 1.4. The ideal observer represents the visual stimulus by summing the squared-amplitude, or power, at each spatial frequency within the range of frequencies in the target stimulus:

$$P_S = \sum_{BW_{low}}^{BW_{high}} f^2 \quad (1.2)$$

The frequency summation portion of the ideal observer, depicted in box A in Figure 1.4, can be described as a single-channel that has a bandwidth that is adjusted to match the signal bandwidth. In a two-interval forced choice task, the ideal observer applies the summation operation to the stimuli in both intervals, and then selects the interval with the greater P_S or signal power (Figure 1.4, box B). The match between the slope of the ideal TvB function and the human TvB data leads to a striking interpretation; human observers sum information optimally across spatial frequencies.

Thus, the ideal observer becomes crucial for interpreting the data in the context of previous summation experiments. The difference between the multiple-channel models used to explain the data from summation experiments and the ideal detector used to explain Kersten's [25] results could not be greater. The multiple-channel model uses several fixed bandwidth independent channels to explain summation experiments, whereas the ideal observer is at the other end of the continuum of summation as it uses a single, adjustable channel to sum spatial frequency.

One final aspect of the TvB functions of human observers lends support to the inter-

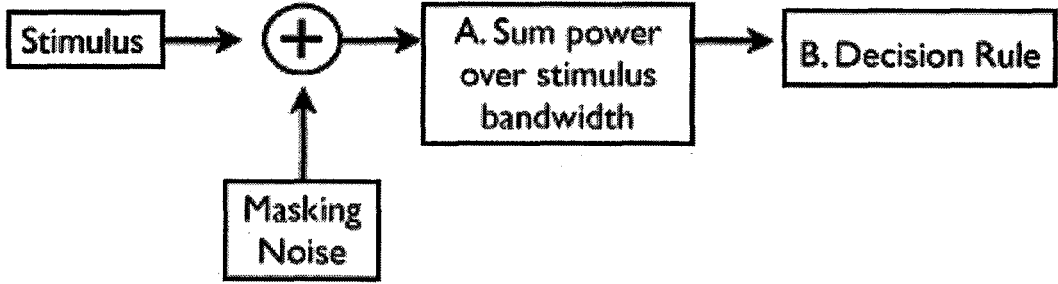


Figure 1.4: A schematic diagram of the ideal observer for a noise detection task.

pretation that human observers may be summing spatial frequency in a manner similar to the ideal detector: absolute efficiency for noise detection is constant for bandwidths greater than one octave and surprisingly high compared to many other visual tasks (e.g., letter and face recognition [16]). Absolute efficiency is defined as:

$$\eta = \left(\frac{c_{ideal}}{c_{observer}} \right)^2 \quad (1.3)$$

where c_{ideal} and $c_{observer}$ are the thresholds for ideal and real observers in RMS contrast. In Kersten's task, the slopes of the TvB curves estimated for real and ideal observers were equal. Therefore, absolute efficiency was constant across a wide range of stimulus bandwidths. Furthermore, Kersten found that absolute efficiency was quite high (i.e., nearly 50% in some conditions). In other words, Kersten's results suggest that human observers integrate information across a wide range of spatial frequencies in a manner that loses very little information.

The high efficiency found in this task causes the interpretation of the match between ideal TvB function slope and human TvB data to become even more of a problem for the standard multiple-channel model. Recall that the standard model assumes channel independence when spatial frequencies are widely separated. Stating that two spatial frequencies are completely independent is equivalent to stating that efficiency ought to be zero; information in the two channels is not summed at all. A strict single channel model would predict high efficiency as information is summed across spatial frequency.

Chapter 2 of this work addresses the challenge just described directly by: replicating and extending Kersten's results, varying stimulus uncertainty/duration and using the classification image technique to measure the frequency information used by observers

when detecting visual noise [1]. Chapter 3 addresses a curious result of Chapter 2 and the channel independence of visual noise centered at 5 and 15 cy/deg using double-judgement procedures [28, 49] and measuring classification images for both detection and identification. Chapter 4 extends noise detection methods to orientation summation by measuring TvB functions and classification images for orientation filtered noise.

References

- [1] A. Ahumada and J. Lovell. Stimulus features in signal detection. Journal of the Acoustical Society of America, 49:1751–1756, Jan 1971.
- [2] F. Attneave. Some informational aspects of visual perception. Psychological Review, 61(3):183–193, 1954.
- [3] S. Baldassi and D.C. Burr. Feature-based integration of orientation signals in visual search. Vision Research, 40(10):1293–1300, 2000.
- [4] H.B. Barlow. The coding of sensory messages. In W.H. Thorpe and O.L. Zangwill, editors, Current Problems in Animal Behaviour, pages 331–360, 1961.
- [5] H.B. Barlow. The exploitation of regularities in the environment by the brain. Behavioral and Brain Sciences, 24(04):602–607, 2002.
- [6] P.J. Bex and W. Makous. Spatial frequency, phase, and the contrast of natural images. Journal of the Optical Society of America A, 19(6):1096–1106, 2002.
- [7] C. Blakemore and J. Nachmias. The orientation specificity of two visual after-effects. The Journal of Physiology, 213(1):157, 1971.
- [8] F.W. Campbell and J.J. Kulikowski. Orientational selectivity of the human visual system. The Journal of physiology, 187(2):437, 1966.
- [9] F.W. Campbell and J.G. Robson. Application of Fourier analysis to the visibility of gratings. The Journal of Physiology, 197(3):551, 1968.
- [10] C.R. Carlson, R.W. Cohen, and I. Gorog. Visual processing of simple two-dimensional sine-wave luminance gratings. Vision Research, 17(3):351, 1977.
- [11] M. Carrasco, C. Penpeci-Talgar, and M. Eckstein. Spatial covert attention increases contrast sensitivity across the CSF: Support for signal enhancement. Vision Research, 40(10-12):1203–1215, 2000.

- [12] S.C. Dakin and R.J. Watt. Biological bar codes in human faces. Journal of Vision, 9(4):2, 2009.
- [13] E.T. Davis and J. Palmer. Visual search and attention: An overview. Spatial Vision, 17(4):249–256, 2004.
- [14] R.L. DeValois and K.K. DeValois. Spatial Vision. Oxford University Press, New York, 1988.
- [15] C. Gaspar, A.B. Sekuler, and P.J. Bennett. Spatial frequency tuning of upright and inverted face identification. Vision Research, 48(28):2817–2826, 2008.
- [16] J. Gold, P.J. Bennett, and A.B. Sekuler. Identification of band-pass filtered letters and faces by human and ideal observers. Vision Research, 39(21):3537–3560, 1999.
- [17] N. Graham. Detection and identification of near-threshold visual patterns. Journal of the Optical Society of America A, 2(9):1468–1482, 1985.
- [18] N. Graham and J. Nachmias. Detection of grating patterns containing two spatial frequencies: a comparison of single-channel and multiple-channel models. Vision Research, 11(3):251–259, Jan 1971.
- [19] N. Graham, J.G. Robson, and J. Nachmias. Grating summation in fovea and periphery. Vision Research, 18(7):815–25, Jan 1978.
- [20] N.V.S. Graham. Visual Pattern Analyzers. Oxford University Press, New York, 1989.
- [21] D.M. Green. Auditory detection of a noise signal. In J.A. Swets, editor, Signal detection and recognition by human observers, pages 523–547, New York, 1960. Wiley.
- [22] R.L. Gregory. The brain as an engineering problem. In W.H. Thorpe and O.L. Zangwill, editors, Current Problems in Animal Behaviour, pages 331–360, 1961.
- [23] D.H. Kelly. Visual response to time-dependent stimuli. i. amplitude sensitivity measurements. Journal of the Optical Society of America, 51:422–9, Apr 1961.
- [24] D.H. Kelly and C.A. Burbeck. Critical problems in spatial vision. Critical Reviews in Biomedical Engineering, 10(2):125, 1984.

- [25] D. Kersten. Statistical efficiency for the detection of visual noise. Vision Research, 27(6):1029–1040, Jan 1987.
- [26] S. Klein and C.F. Stromeyer. Adaptation to square-wave gratings: In search of the elusive third harmonic. Vision Research, 13:1335–1342, 1973.
- [27] S. Klein and C.F. Stromeyer. On inhibition between spatial frequency channels: adaptation to complex gratings. Vision Research, 20(5):459–466, 1980.
- [28] S.A. Klein. Double-judgment psychophysics: problems and solutions. Journal of the Optical Society of America A, 2(9):1560–85, Sep 1985.
- [29] S.A. Klein. Channels: bandwidth, channel independence. detection vs. discrimination. In B Blum, editor, Channels in the visual nervous system: neurophysiology, psychophysics and models, pages 11–27, London, 1992. Freund Publishing House.
- [30] G.E. Legge and J.M. Foley. Contrast masking in human vision. Journal of the Optical Society of America, 70(12):1458–1471, 1980.
- [31] G.E. Legge, G.S. Rubin, and A. Luebker. Psychophysics of reading. V. The role of contrast in normal vision. Vision Research, 27(7):1165–1177, 1987.
- [32] Z.L. Lu and B.A. Doshier. External noise distinguishes attention mechanisms. Vision Research, 38(9):1183–1198, 1998.
- [33] L.A. Olzak and J.P. Thomas. Neural recoding in human pattern vision: model and mechanisms. Vision Research, 39(2):231–256, 1999.
- [34] L.A. Olzak and T.D. Wickens. Paradigm shifts: New techniques to answer new questions. Perception, 28(12):1509–1531, Jan 1999.
- [35] A. Pantle and R. Sekuler. Size-detecting mechanisms in human vision. Science, 162(858):1146, 1968.
- [36] D.G. Pelli, C.W. Burns, B. Farell, and D.C. Moore. Identifying letters. Vision Research, 46(28):4646–4674, 2006.
- [37] M.E. Perkins and M.S. Landy. Nonadditivity of masking by narrow-band noises. Vision Research, 31:1053–1065, 1991.
- [38] G.C. Phillips and H.R. Wilson. Orientation bandwidths of spatial mechanisms measured by masking. Journal of the Optical Society of America A, 1(2):226–232, 1984.

- [39] J.G. Robson. Spatial and temporal contrast-sensitivity functions of the visual system. Journal of the Optical Society of America, 56(8):1141–1142, 1966.
- [40] M.B. Sachs, J. Nachmias, and J.G. Robson. Spatial-frequency channels in human vision. Journal of the Optical Society of America, 61(9):1176–1186, 1971.
- [41] O.H. Schade. Optical and photoelectric analog of the eye. Journal of the Optical Society of America, 46(9):721–39, Sep 1956.
- [42] R. Sekuler and L. Ganz. A new aftereffect of seen movement with a stabilized retinal image. Science, 139:419–420, 1963.
- [43] C.E. Shannon and W. Weaver. A mathematical theory of communication. Bell Syst. Tech. J., 27(379):623, 1948.
- [44] J.A. Solomon and D.G. Pelli. The visual filter mediating letter identification. Nature, 369(6479):395–397, 1994.
- [45] J.P. Thomas. Spatial vision then and now. Vision Research, 26(9):1523, 1986.
- [46] J.P. Thomas and J. Gille. Bandwidths of orientation channels in human vision. Journal of the Optical Society of America, 69(5):652–660, 1979.
- [47] C.W. Tyler and C.C. Chen. Signal detection theory in the 2AFC paradigm: Attention, channel uncertainty and probability summation. Vision Research, 40(22):3121–3144, 2000.
- [48] B.A. Wandell. Foundations of Vision. Wiley, New York, 1995.
- [49] A.B. Watson and J.G. Robson. Discrimination at threshold: labelled line detectors in human vision. Vision Research, 21:1115–1122, 1981.
- [50] H.R. Wilson and J.R. Bergen. A four mechanism model for threshold spatial vision. Vision Research, 19(1):19–32, Jan 1979.
- [51] H.R. Wilson and D.J. Gelb. Modified line-element theory for spatial-frequency and width discrimination. Journal of the Optical Society of America A, 1(1):124–131, 1984.
- [52] H.R. Wilson and F. Wilkinson. Evolving concepts of spatial channels in vision: From independence to nonlinear interactions. Perception, 26:939–960, Jan 1997.

- [53] H.R. Wilson, D.K. McFarlane, and G.C. Phillips. Spatial frequency tuning of orientation selective units estimated by oblique masking. Vision Research, 23(9):873–82, 1983.

Chapter 2

Spatial frequency summation in visual noise

Abstract

Kersten (*Vision Res*, 1987, 27(6), 1029-40) reported that absolute efficiency for the detection of static, one-dimensional band-pass noise was high and approximately constant for stimulus bandwidths ranging from 1 to 6 octaves. This result implies that human observers integrated information efficiently across a wide range of spatial frequency. One interpretation of this result – and similar results obtained with auditory stimuli (Green, DM, *J Acoust Soc Am*, 1960, 32(1), 121-131) – is that human observers, like ideal observers, can detect stimuli using an internal filter that has an adjustable bandwidth. The current experiments replicate Kersten’s findings; extend them to the case where observers are uncertain about stimulus bandwidth; and use the classification image technique to estimate the filter used to detect noise stimuli that differ in bandwidth. Our results suggest that observers do not adjust channel bandwidth to match the stimulus, and that detection thresholds are consistent with the predictions of a multiple-channel model.

Introduction

Models of the initial stages of visual processing typically include a stage in which stimuli are encoded by an array of multiple independent filters, or channels, that are

tuned to relatively narrow ranges of spatial frequency and orientation [9, 25, 27, 28, 29]. Support for multiple-channel models comes (in part) from spatial frequency summation studies. In a typical summation experiment, contrast detection thresholds are measured for two sinusoidal gratings of frequency F and nF , where n is an integer, and a compound waveform consisting of the sum of the two sine wave gratings ($F + nF$). Summation is said to occur if the detection threshold for the compound is lower than threshold for either component. Typically, summation exceeding what would be expected on the basis of probability summation has been found only when the spatial frequencies contained in the compound grating do not differ by more than 1-2 octaves [13, 14, 15].

Another approach to investigating spatial frequency summation was described by Kersten [20], who measured detection thresholds for static, one-dimensional bandpass noise stimuli presented in a background of white noise. The bandwidth of the noise was varied across conditions. For bandwidths ranging from approximately 0.5 to 6 octaves, Kersten found that detection thresholds, when expressed as root-mean squared (RMS) contrast, were proportional to the quarter-root of the stimulus bandwidth. Kersten also calculated the performance of an ideal detector that summed contrast power optimally across all spatial frequency components in the stimulus. Surprisingly, the ideal detector's threshold (expressed as RMS contrast) was only 0.5 units lower than human thresholds and was proportional to the quarter-root of stimulus bandwidth. Hence, absolute efficiency, defined as

$$\eta = \left(\frac{c_{ideal}}{c_{observer}} \right)^2 \quad (2.1)$$

where c_{ideal} and $c_{observer}$ are the ideal and human observer's RMS contrast thresholds, was high ($\approx 50\%$) and nearly constant across stimulus bandwidths ranging from 0.5 to 6 octaves. To explain this result, Kersten [20] speculated that human observers used an internal channel, or filter, whose bandwidth could be adjusted to match the bandwidth of the noise signal. (A similar model was proposed by Green [17, 18] to account of similar findings in experiments with filtered auditory noise.) An adjustable channel can be thought of as a template in the spatial frequency domain which could be constructed by combining the outputs of narrow-band channels centered on different spatial frequencies. The fact that efficiency was nearly constant across a wide range of bandwidths suggests that the process of combining responses across channels is efficient, a conclusion that appears to be inconsistent with the results of spatial frequency summation studies using sine wave grating stimuli.

The goal of this paper is to address the apparent inconsistency between the results of

summation studies using grating and noise stimuli. Four experiments are described. The first replicates and extends the findings of Kersten [20]. The second examines spatial frequency summation for very briefly flashed stimuli in conditions where observers are uncertain about stimulus bandwidth. The third experiment demonstrates that varying the center frequency of visual noise does not affect optimal summation. Finally, the fourth experiment uses the response classification technique [3, 10] to estimate the tuning characteristics of the internal filters used in this noise detection task.

2.1 Experiment 1

2.1.1 Methods

2.1.1.1 Observers

The observers were students at McMaster University (21-26 years old). All observers were naïve with respect to the experimental hypotheses, had normal or corrected-to-normal Snellen acuity, and had previous experience in visual psychophysical tasks. Observers were paid for participating in this experiment.

2.1.1.2 Apparatus

A Macintosh G4 running MATLAB and the Psychophysics toolbox [7, 22] was used to generate and present the stimuli and collect responses. The stimuli were presented with a nVidia GeForce2 MX card on a Sony GDM-F520 monitor set to a resolution of 1920 by 1440 pixels. The entire display subtended a visual angle of 10.8 by 8.3 degrees at the viewing distance of 2m. The frame rate of the display was 85 Hz and the mean luminance 45 cd/m^2 . The luminance of the display was calibrated using a PhotoResearch PR-650 photometer. A bit stealing method [24] was used to increase the contrast resolution. Head position was stabilized with a chin-forehead rest.

2.1.1.3 Stimuli

The stimuli were filtered, one-dimensional Gaussian white noise that varied in spatial frequency bandwidth. The center-frequency of the patterns was fixed at 5 cy/deg on a logarithmic scale. In this experiment, the seven spatial frequency bandwidths were 0.5, 1, 1.25, 1.5, 2, 3, 4 octaves, which corresponded to a two-sided bandwidth of approximately 2.5, 10, 11.25, 15, 20, 30 and 40 cy/deg. Stimulus contrast was modulated with a

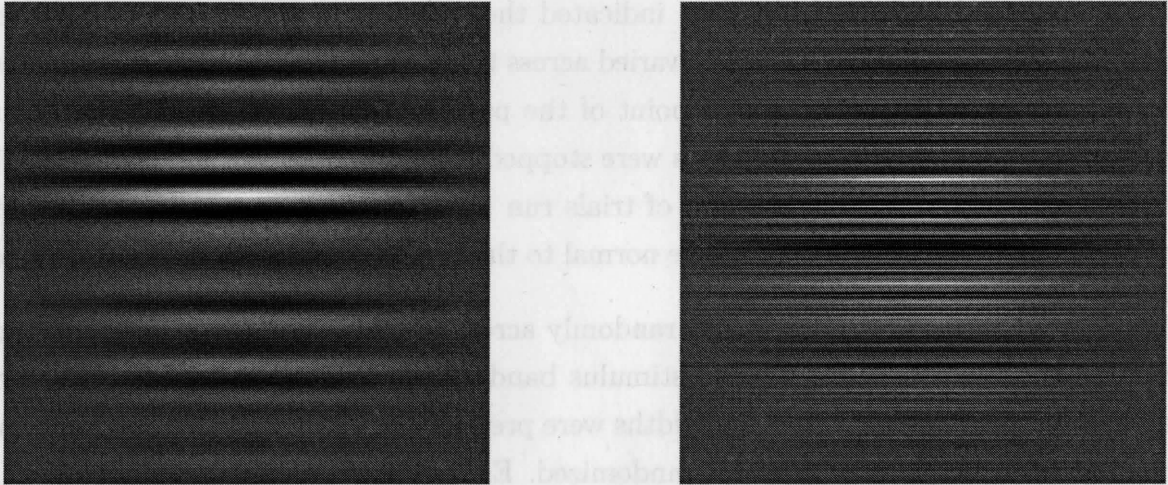


Figure 2.1: The left pattern is an example of a narrow (0.5 octaves) bandwidth stimulus and the right pattern is an example of a wide (4 octaves) bandwidth stimulus.

circularly-symmetric Gaussian envelope with a standard deviation of 1.08 degrees of visual angle. Fig 2.1 shows two high contrast examples of the stimuli. Stimuli were presented for 200 ms. Stimuli were presented in a background of white masking noise that had a variance of 0.04, 0.08, 0.16 or 0.32. The white masking noise was static and presented only during the stimulus intervals and not during the inter-stimulus interval.

2.1.1.4 Procedures

Observers viewed the stimuli binocularly through natural pupils and responses were recorded with a computer keyboard. A 3×3 degree frame of maximum contrast surrounded the stimuli to reduce spatial uncertainty. To reduce adaptation, the frame had a 50% probability of being black or white on each trial. The frame had a width of two pixels and was on the screen for the entire duration of each trial, from the presentation of the fixation point until the observer's response. A two-interval forced-choice (2-IFC) procedure was used. The observer was instructed to fixate a high-contrast dot located in the center of the display. The observer initiated each trial by pressing the spacebar on the keyboard. After a delay of 50 ms, the fixation point was removed, then after another 50 ms delay the first stimulus interval appeared. The first stimulus interval was 200 ms in duration and was followed by a 300 ms blank inter-stimulus interval. After the inter-stimulus interval a second 200 ms stimulus interval appeared. The stimulus intervals were marked by a clearly-audible tones. Observers were asked to select, via a button press, which of the two intervals contained the target. Auditory feedback, in the

form of low- and high-pitched tones, indicated the accuracy of the response after each trial. Stimulus contrast variance was varied across trials using four interleaved staircases, two converging on the 71% correct point of the psychometric function and two on the 84% correct point [26]. The staircases were stopped when the observer had completed 75 trials in each staircase. The number of trials run in each session was 2100. Thresholds were estimated by fitting a cumulative normal to the combined data of all four staircases.

The level of masking noise varied randomly across test session but was held constant within a test session. The order of stimulus bandwidth was block randomized in each test session, meaning that the bandwidths were presented in separate blocks of trials and the order of blocks/bandwidths was randomized. Each observer completed all bandwidth conditions during a single session. The masking noise variance was selected randomly for each session. Four sessions were run at each of the four levels of masking noise.

2.1.2 Results

Figure 2.2 shows threshold versus bandwidth (TvB) functions for one typical observer (AMC). Each symbol type (circles, inverted and upright triangles) represents the thresholds measured with stimuli embedded in different levels of external noise. Not surprisingly, detection thresholds increased with increasing levels of external noise. The dashed lines in the figure have a slope of 0.25 and have been shifted vertically to fit the data. It can be seen that they provide good fits to the data when the stimulus bandwidth was greater than ≈ 1 octave, but that thresholds at narrower bandwidths tended to be nearly constant. We therefore computed best-fitting power functions for thresholds obtained with stimulus bandwidths ≥ 1 octave: the slopes for the estimated TvB functions are shown in Table 2.1. Except for observer AMC in the condition using the highest level of external noise, all of the estimated slopes were very close to 0.25.

Table 2.1: Threshold vs. bandwidth slopes for three observers in Experiment 2.1.

Noise σ^2	AMC	AM	AG
0.04	0.26	0.27	0.26
0.08	0.25	0.25	0.26
0.16	0.27	0.23	0.25
0.32	0.20	0.25	0.25

Figure 2.3 plots absolute efficiency (Eq 2.1) as a function of stimulus bandwidth at

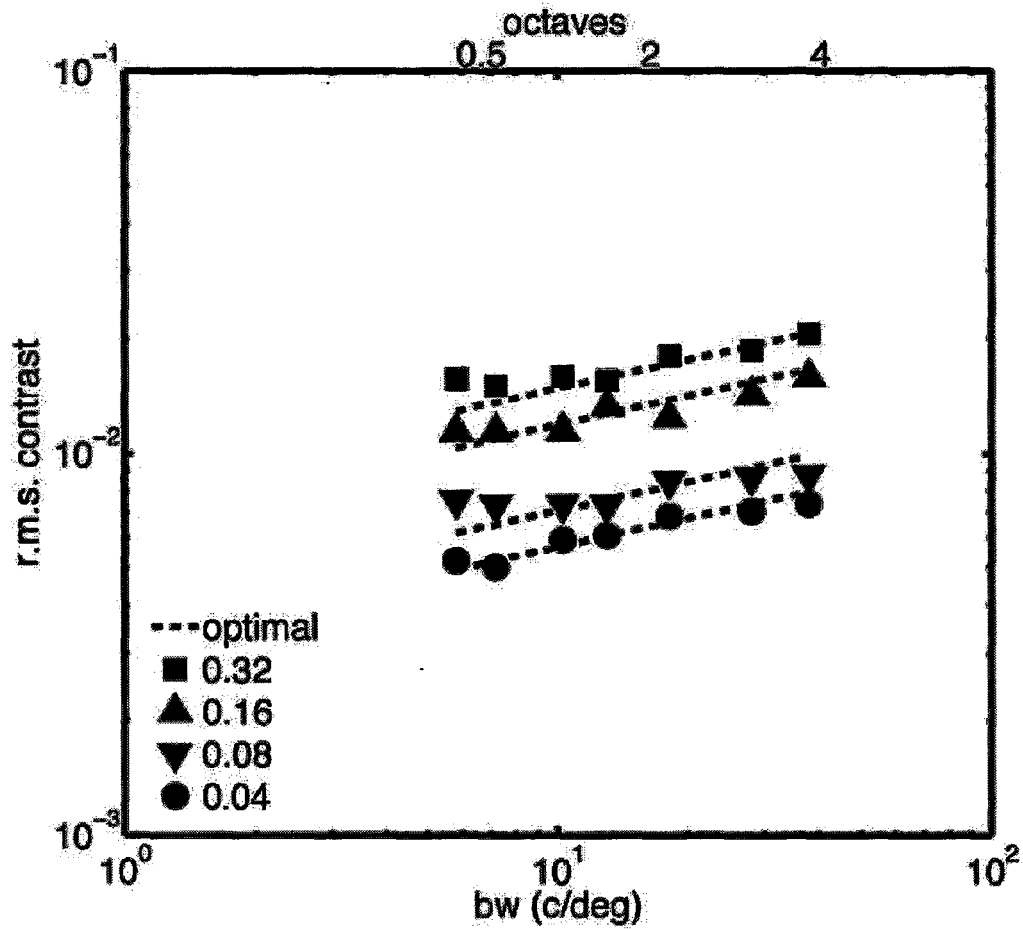


Figure 2.2: Noise detection threshold versus stimulus bandwidth for observer AMC in Experiment 2.1. The different symbols represent thresholds obtained with different levels of masking noise. The variance of the masking noise is indicated in the legend. The dotted lines have a slope of 0.25 and have been shifted vertically to fit the data.

three levels of external noise. Efficiency was high (i.e., greater than 0.4) at all levels of external noise and approximately constant across spatial frequency bandwidth. Similar results were obtained from two other observers (see Figures 2.4 and 2.5).

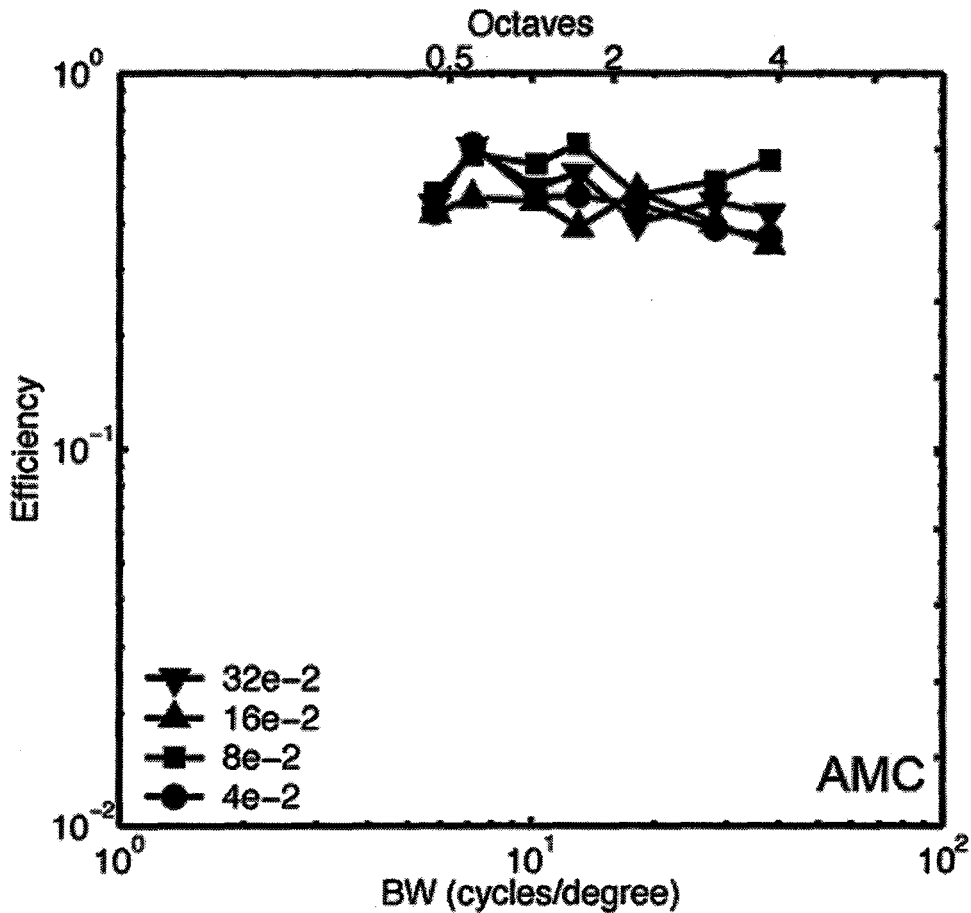


Figure 2.3: A plot of absolute efficiency (Eq 1.) versus stimulus bandwidth plot for observer AMC from Experiment 2.1. The different symbols represent efficiency obtained with different levels of masking noise. The variance of the masking noise is indicated in the legend.

2.1.3 Discussion

The results of Experiment 2.1 replicate the key findings reported by Kersten[20] and extend them to a greater range of external noise levels. Experiment 2.1 found that the TvB functions measured in three observers had slopes (in log-log coordinates) of approximately 0.25, which is the same slope exhibited by an ideal observer [20]. Moreover, absolute efficiency was high, averaging 40% across all observers and conditions. Finally, the results did not vary systematically with the level of masking noise. These findings are

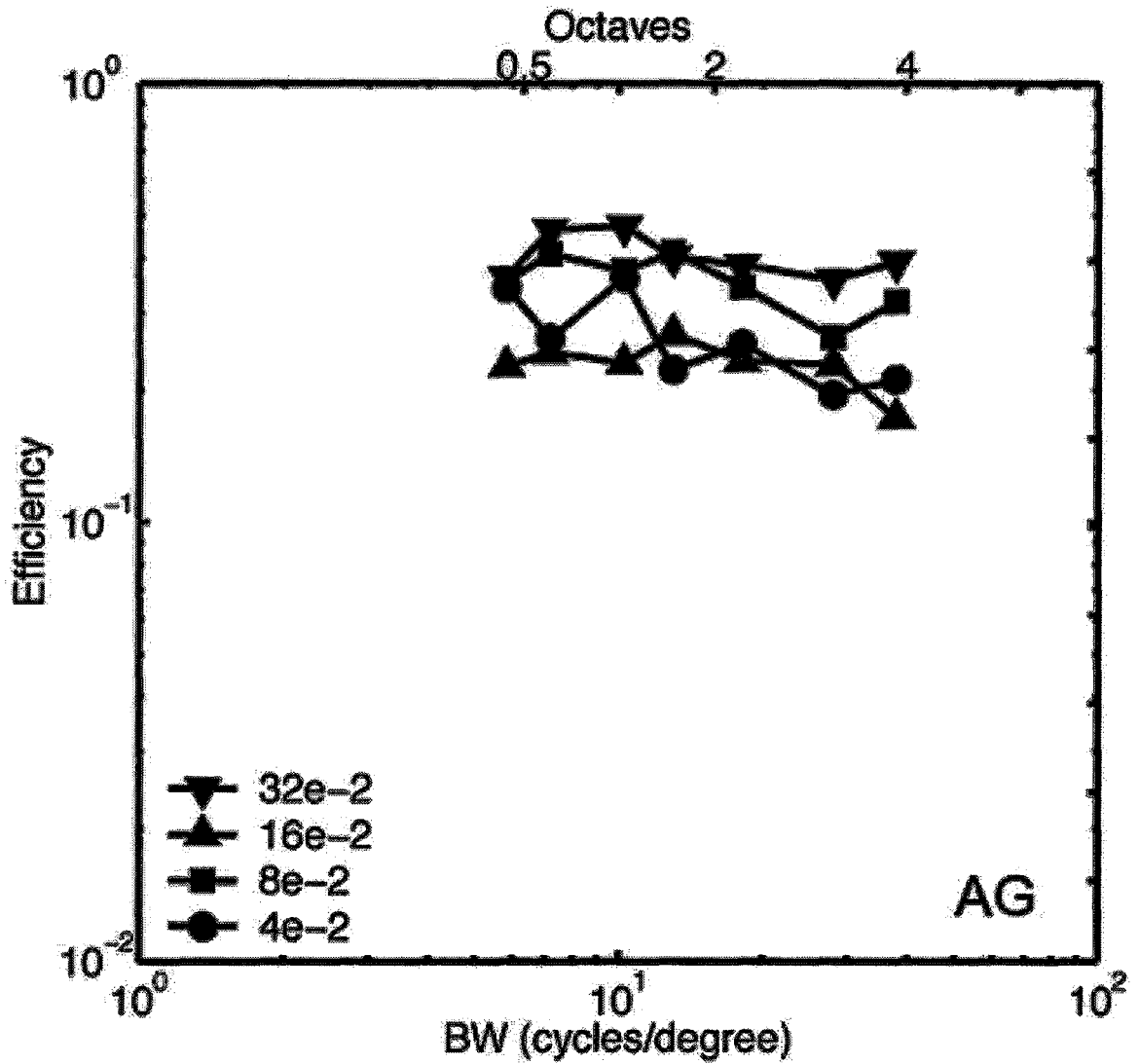


Figure 2.5: A plot of absolute efficiency versus bandwidth plot for observer AG from Experiment 2.1. The different symbols represent efficiency obtained with different levels of masking noise. The variance of the masking noise is indicated in the legend.

consistent with the idea that noise patterns are detected by an internal filter that has an adjustable frequency bandwidth [16, 20] with a minimum bandwidth of approximately one octave. The next three experiments explore the idea that an adjustable bandwidth channel is involved in the detection of visual noise.

2.2 Experiment 2

In Experiment 2.1 and previous experiments [16, 20], stimulus bandwidth was varied across blocks of trials but held constant within blocks. Hence, subjects could have used knowledge about stimulus bandwidth to adjust the bandwidth of an internal channel used to detect the target. If the bandwidth adjustment depends on prior knowledge of the stimulus bandwidth, then randomizing stimulus bandwidth should disrupt this mechanism and result in less efficient frequency summation (i.e., increase the slope of the TvB function). Experiment 2.2 tested this idea by randomizing stimulus bandwidth across trials.

2.2.1 Methods

2.2.1.1 Observers

The two observers who participated in this experiment were both 28 years old, were students at McMaster University, and were paid for their participation. Both observers were naïve with respect to the experimental hypotheses, had either normal or corrected-to-normal Snellen acuity, and had extensive practice in psychophysical tasks

2.2.1.2 Apparatus

The apparatus used to run this experiment was identical to that used in Experiment 2.1.

2.2.1.3 Stimuli

As in Experiment 2.1, all stimuli were one-dimensional filtered Gaussian white noise that varied in spatial frequency bandwidth. In this experiment, the stimuli were presented at one of two temporal durations, 200 ms or 12 ms. Unlike Experiment 2.1, the stimuli were not presented in masking noise.

2.2.1.4 Procedures

The procedures were identical to those used in Experiment 2.1 save that stimulus bandwidth was blocked or randomized within each testing session. In the blocked condition, stimulus bandwidth was fixed within a block of 150 trials, and the order of the seven stimulus bandwidths was randomized across blocks. In the randomized condition, stimulus bandwidth was randomly selected on each trial, with the constraint that each bandwidth was presented on 150 trials. Observers alternated between blocked and randomized presentation across days; one observer started the experiment with the randomized presentation and the other observer started with the blocked presentation.

Thresholds at each bandwidth for each session were estimated by fitting a cumulative normal to the data. Three sessions were run for each of the presentation types and temporal durations. The mean threshold from the three sessions in each condition was calculated and plotted in a threshold versus bandwidth (TvB) function.

2.2.2 Results

Figures 2.6 and 2.7 show TvB functions for each observer in each of the blocked and randomized bandwidth conditions at both the short and long temporal durations. Shortening the temporal duration of the stimulus from 200 ms to 12 ms increased thresholds by approximately 0.2 log units in all conditions, but did not alter the slope of the TvB function. Randomizing stimulus bandwidth across trials had virtually no effect on thresholds and also did not alter the slopes of the TvB function. The dotted lines in each figure have a slope of 0.25 – the slope of the TvB curve for an ideal detector – and provide a good fit to the data obtained with stimulus bandwidths greater than or equal to one octave. To confirm this observation, the method of least squares was used to compute the best-fitting line for log-transformed thresholds measured with stimulus bandwidths ≥ 1 octave. The slopes of best-fitting lines are shown in Table 2.2, along with 95% confidence intervals estimated using a bootstrap procedure [11]. The slopes of the lines did not differ significantly from 0.25, and were similar in the blocked and randomized bandwidth presentations.

2.2.3 Discussion

The slopes of the TvB curves measured in the current experiment, which did not embed the targets in external noise, did not differ from those measured in Experiment 2.1,

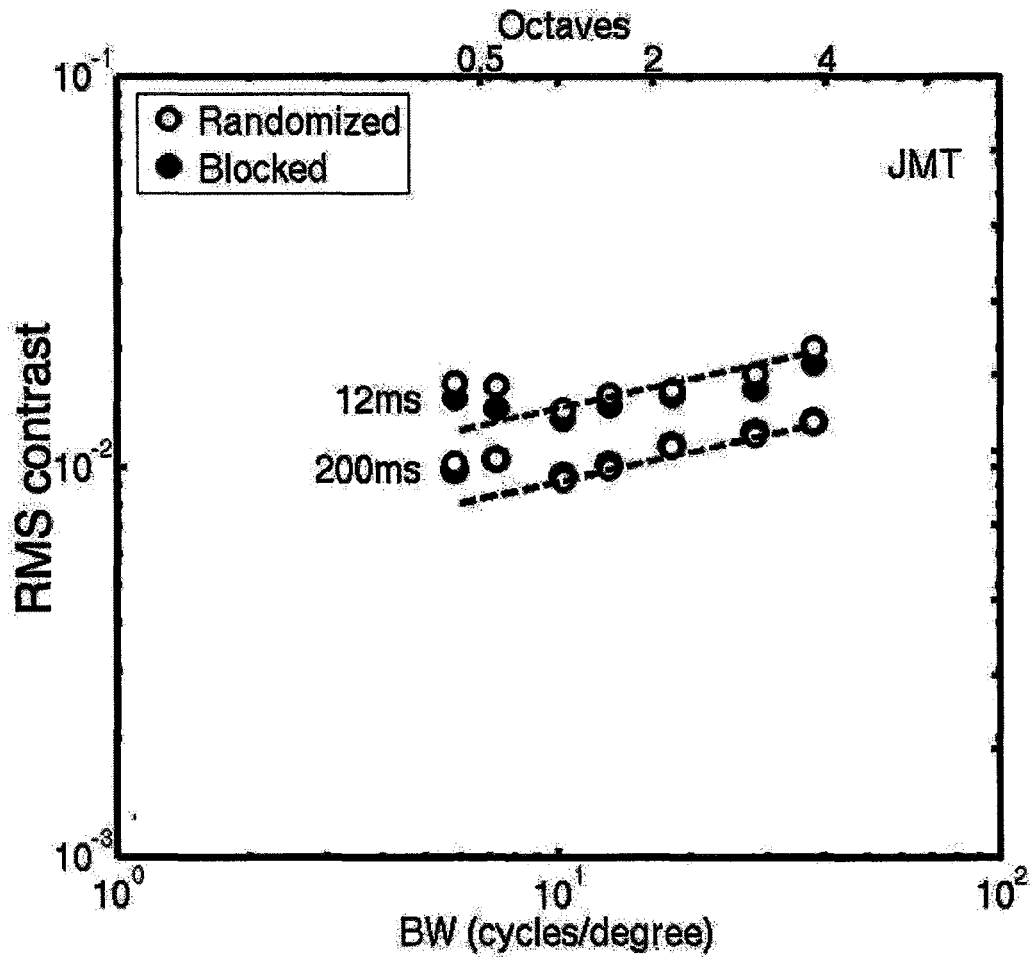


Figure 2.6: Detection thresholds for observer JMT in Experiment 2.2 in the randomized-bandwidth (unfilled symbols) and blocked-bandwidth (filled symbols) conditions. Error bars ($\pm$ one standard error) are plotted but are smaller than the symbols. The dashed lines have a slope of 0.25 and have been shifted vertically to fit the thresholds obtained in the randomized condition.

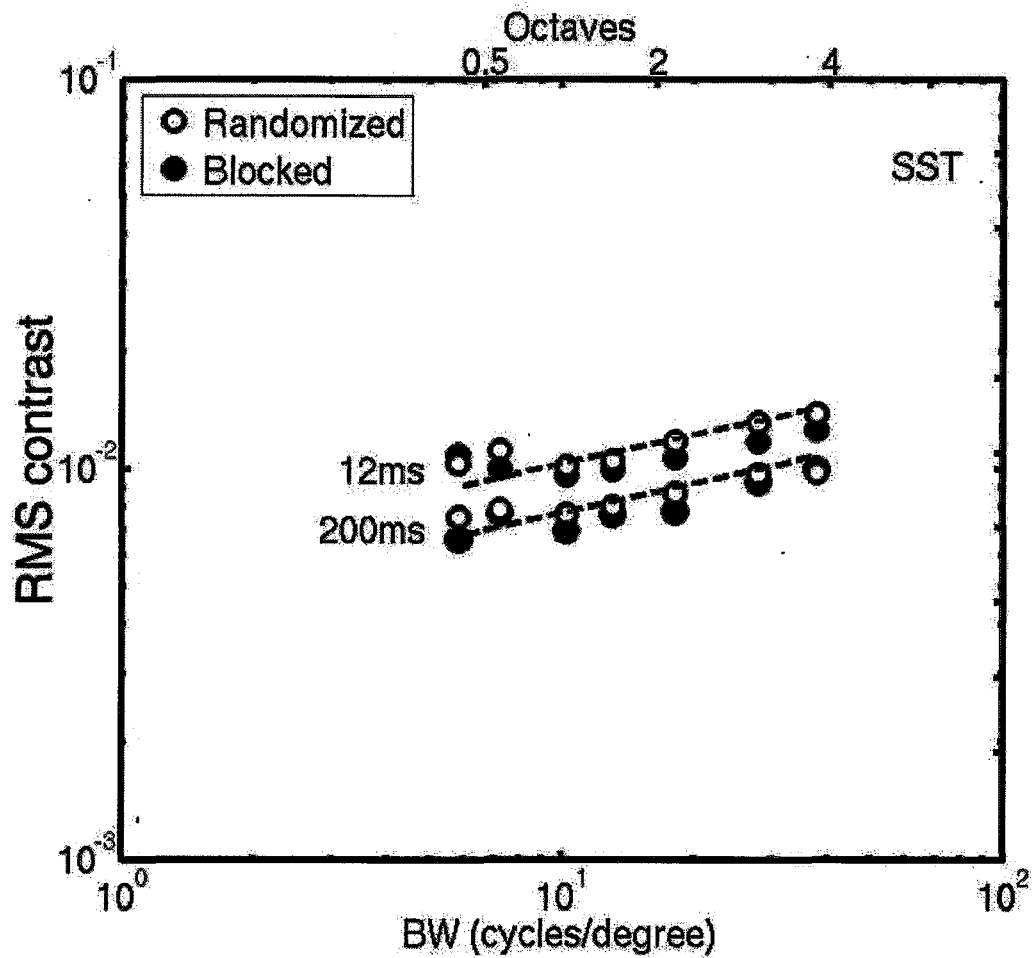


Figure 2.7: Detection thresholds for observer SST in Experiment 2.2. Plotting conventions are the same as in Figure 2.6.

which did use external noise. Hence, the current findings support the idea that the slope of the TvB curve is not sensitive to the level of external noise. The current experiment also found that the slope of the TvB curve was not affected by significant changes in stimulus duration or by the introduction of uncertainty about stimulus bandwidth. Previous experiments using sinusoidal gratings have reported that uncertainty about a target's spatial frequency increases detection thresholds [8, 19]. The explanation for this effect is that uncertainty increases the number of noisy frequency-selective channels that must be monitored to detect the target, and therefore lowers the signal-to-noise ratio. Within this multiple channels framework, the failure to find an effect of uncertainty in the current experiment suggests that increasing the uncertainty about stimulus bandwidth did not alter the number of channels observers monitored to detect the noise target.

One explanation for the optimal frequency summation found in Experiment 2.1 (and by Kersten [20]) is that observers use prior knowledge of stimulus bandwidth to adjust the bandwidth of an internal channel. The finding that randomizing stimulus bandwidth had no effect on the slope of the TvB function is inconsistent with this hypothesis. If the adjustable bandwidth hypothesis is to be maintained, then it must be assumed that some process extracts information about stimulus bandwidth on a trial-by-trial basis using information from the stimulus itself. The data from the two temporal duration conditions place limits on this process. If one considers only the data from the 200 ms temporal presentation condition, one could argue that observers adjusted their channel bandwidths during the presentation of the stimulus via feedback connections from higher cortical areas to V1 [21]. The data from the 12 ms condition, however, show that optimal summation occurs with very brief stimulus presentations. Therefore, it seems that feedback – if it is operating in this task – must adjust channel bandwidth based on information contained in a very brief stimulus.

Table 2.2: Slopes of threshold vs. bandwidth functions from Experiment 2.2.

Bandwidth	duration	SST	JMT
Blocked	12 ms	0.22 ± 0.15	0.22 ± 0.12
Randomized	12 ms	0.25 ± 0.11	0.25 ± 0.10
Blocked	200 ms	0.27 ± 0.078	0.23 ± 0.21
Randomized	200 ms	0.21 ± 0.08	0.25 ± 0.09

2.3 Experiment 3

Kersten[20] used stimuli that had a center frequency of 2 cy/deg. In our first two experiments, we used a slightly higher center-frequency of 5 cy/deg. In all cases the center frequencies were close to the peak of the contrast sensitivity function [9]. In Experiment 2.3 we examined whether evidence of optimal summation could be found with high spatial frequency stimuli. If a high center spatial frequency is used, one would expect that contrast thresholds for the detection of the noise signal would increase because of the reduced sensitivity to high spatial frequencies. However, the finding of optimal summation does not depend on the absolute efficiency of noise detection, but rather the slope of the TvB function. Because optimal summation depends not on absolute efficiency but how efficiency changes with bandwidth, it is possible to find evidence of optimal summation even if the overall absolute efficiency is lower than the efficiency we found in Experiment 2.1. In this experiment we examine whether noise signals can still be detected optimally if the stimulus center frequency is increased to 15 cy/deg.

2.3.1 Methods

2.3.1.1 Observers

The observers in this experiment were the author (CPT) and two individuals recruited from the McMaster community who were paid for their participation. The observers were 25-29 years of age and had normal or corrected-to-normal Snellen acuity.

2.3.1.2 Apparatus

The apparatus was the same as in Experiment 2.1, except for one change: contrast resolution was increased by using a Bits++ device (Cambridge Research Systems) in Mono++ mode. The Bits++ setup enabled us to achieve 14-bit resolution of contrast.

2.3.1.3 Stimuli

The stimuli were one-dimensional filtered Gaussian white noise that varied in spatial frequency bandwidth. The center-frequency of the patterns was fixed at 15 cy/deg on a logarithmic scale. The seven spatial frequency bandwidths used in this experiment were the same, when expressed in octaves, as Experiment 2.1 (0.5, 1, 1.25, 1.5, 2, 3, 4 octaves wide). A Gaussian envelope with a standard deviation of 1.5 degrees of visual angle to reduce edge artifacts. Stimuli were presented for 200 ms. Stimuli were presented in a

Gaussian white noise mask that had a contrast variance of 0.08.

2.3.1.4 Procedure

This experiment used a procedure that was identical to that of Experiment 2.1.

2.3.2 Results

Figure 2.8 shows the contrast thresholds for three observers in the noise detection task. As was done for Experiment 2.1, linear fits to the data were computed for all the points greater than one octave. For all observers, the 95% confidence interval (computed via a bootstrap simulation) of the slope of the best fitting line contained the value 0.25, which is again consistent with the predictions of optimal summation.

As was done for Experiment 2.1, the thresholds for an ideal observer were estimated and then used to calculate absolute efficiency (see Figure 2.9). Absolute efficiency was approximately 0.2 in observers MB and CPT and slightly lower (on average) in observer AL. These efficiencies are lower than the values (i.e., 0.3-0.5) obtained in Experiment 2.1 with 5 cy/deg stimuli. As discussed above, this decrease was expected because optical factors reduce retinal contrast, and therefore efficiency, much more at higher spatial frequencies [4]. As was found in Experiment 2.1, efficiency was constant as a function of stimulus bandwidth for two observers (MB and CPT). In observer AL, however, efficiency fell by approximately 50% as stimulus bandwidth increased from 0.5 to 4 octaves. Interestingly, for this observer efficiency was nearly constant (≈ 0.2) when stimulus bandwidth was ≤ 1.5 octaves, and then fell to approximately 0.1 for bandwidths greater than 1.5 octaves.

2.3.3 Discussion

The results of Experiment 2.3 extend the finding of optimal summation to patterns with a high center spatial frequency. This result demonstrates that the efficient combination of spatial frequency information is not limited to frequencies at or near the peak of the contrast sensitivity function.

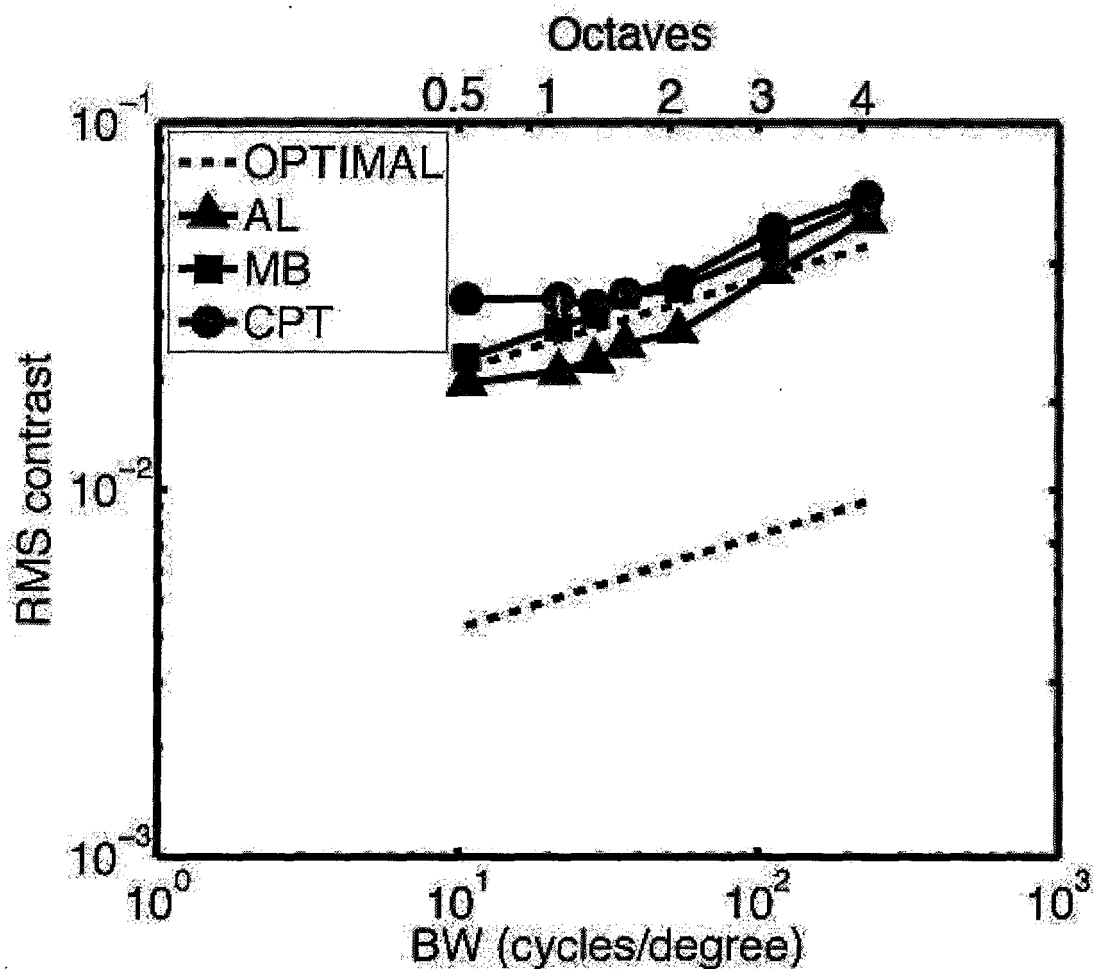


Figure 2.8: Threshold versus bandwidth functions for the three observers in Experiment 2.3. The dashed lines are provided as a reference and illustrate the expected slope (0.25) if the summation of spatial frequency information was optimal.

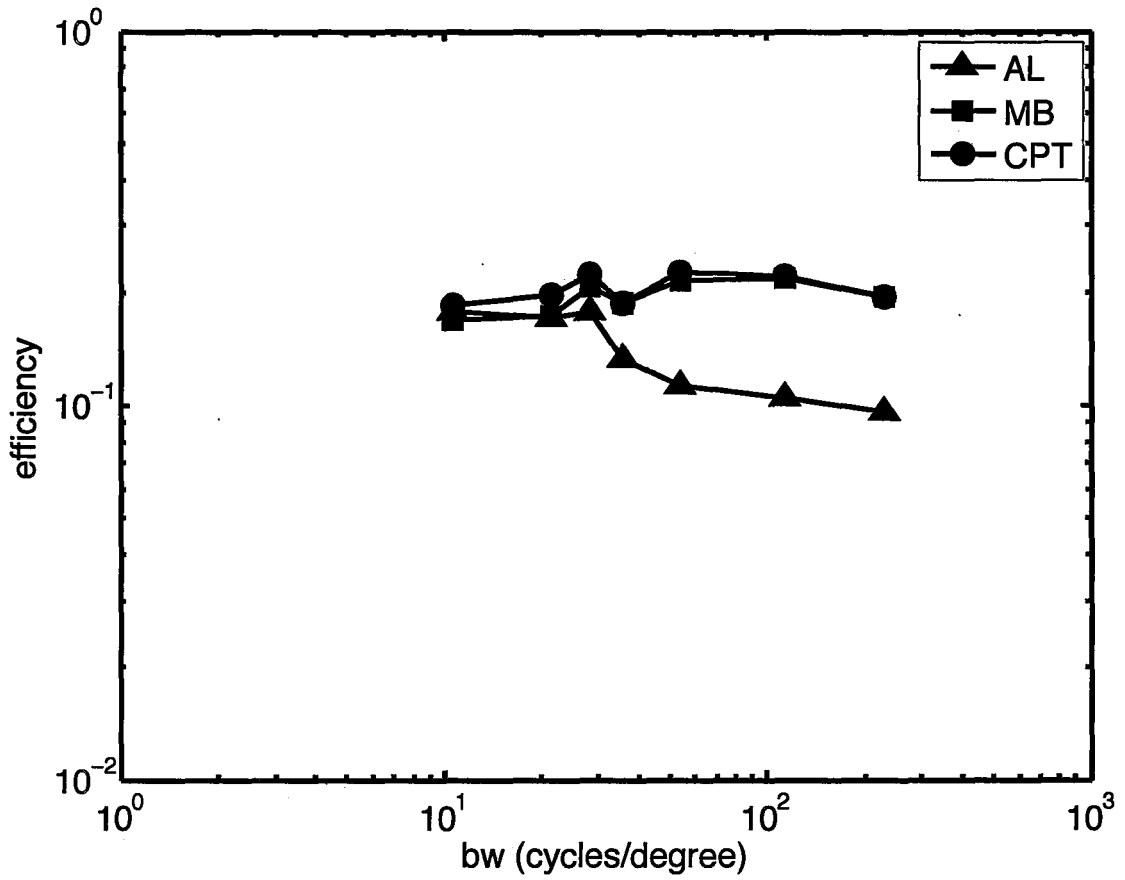


Figure 2.9: Absolute efficiency versus pattern bandwidth for three observers detecting 15 cy/deg center-frequency patterns.

2.4 Experiment 4

The results of the first three experiments show that spatial frequency information can be summed optimally across a wide range of spatial frequency, and that optimal summation does not depend critically on prior knowledge of stimulus bandwidth, the level of external noise, stimulus duration, or center frequency. However, it is still an open question as to what stimulus information observers use to detect visual noise. The current experiment uses the classification image technique to examine this issue.

Recently the classification image technique [3] has been applied to wide variety of visual tasks such as vernier acuity [6], grouping [12], face recognition [23] and attention [10]. This wide range of results has shown that the classification image technique can reveal what aspects of the stimulus are used by observers to perform perceptual tasks. Here we apply the classification technique to examine the template used when a noise detection task is performed.

2.4.1 Methods

2.4.1.1 Observers

One observer from Experiment 2.1 and two additional observers participated in this experiment. All observers were naïve with respect to the experimental hypotheses and had practice in this and other psychophysical tasks. Observers were 25-29 years of age and all had normal Snellen acuity.

2.4.1.2 Apparatus

The apparatus was identical to that used in Experiment 2.1.

2.4.1.3 Stimuli

As in Experiment 2.1, all stimuli were one-dimensional filtered Gaussian white noise that varied in spatial frequency bandwidth. The center-frequency of the patterns was fixed during a session at either 5 or 15 cy/deg. Spatial frequency bandwidth was 1, 2, 4 or 6 octaves. Stimulus contrast was windowed with a two-dimensional Gaussian envelope with a standard deviation of 1.5 degrees of visual angle. Stimuli were presented for 200 ms in a Gaussian white noise mask that had a contrast variance of 0.32.

2.4.1.4 Procedures

This experiment used a 2-IFC procedure that was identical to the procedure used in Experiment 2.1. Stimulus bandwidth was blocked within each testing session, but the order was randomized across testing sessions. Each session began with 50 trials during which stimulus rms contrast was adjusted with a 2-down/1-up staircase procedure [26] to find the observer's 71% correct detection threshold. Stimulus contrast was then fixed for the remaining 400 trials of the block. Each observer completed 3200 trials in each of eight conditions (i.e., two center-frequencies and four bandwidths).

2.4.1.5 Analysis

We used the 2-IFC variant of the response classification image task which has been described in detail previously [1, 2]. The major difference between the classification images measured in previous work and the classification images presented here is that our classification images were calculated using the *power spectra* of the noise masks, rather than the noise masks themselves. On each trial, the Fourier transform of the noise mask in each interval was computed. Next, the difference between the pair of power spectra was calculated at each spatial frequency, and the difference spectra was placed into one of four bins based on which interval contained the signal (1 or 2) and the observer's response (correct or incorrect). The power spectra within each bin were averaged: the two average spectra computed from correct responses were summed, as were the two average spectra computed from incorrect trials. Finally, the difference between the correct and incorrect averaged spectra was computed and the resulting classification image was normalized to have a peak value of 1. Classification images calculated using this procedure are proportional to the linear template applied to the power spectra [1, 2].

2.4.2 Results

The dotted lines in Figures 2.10 and 2.11 show the classification images obtained from one representative observer with stimulus center frequencies of 5 and 15 cy/deg, respectively. In both figures, the y-axis can be thought of as representing the relative strength of the linear association between the power at a given frequency and the observer's response.

The plots in Figure 2.10 show that this observer was influenced most by power at 5 cy/deg, which corresponds to the center frequency of the stimulus. Surprisingly, the plots in Figure 2.11 show that behaviour was most influenced by power at 5 cy/deg even

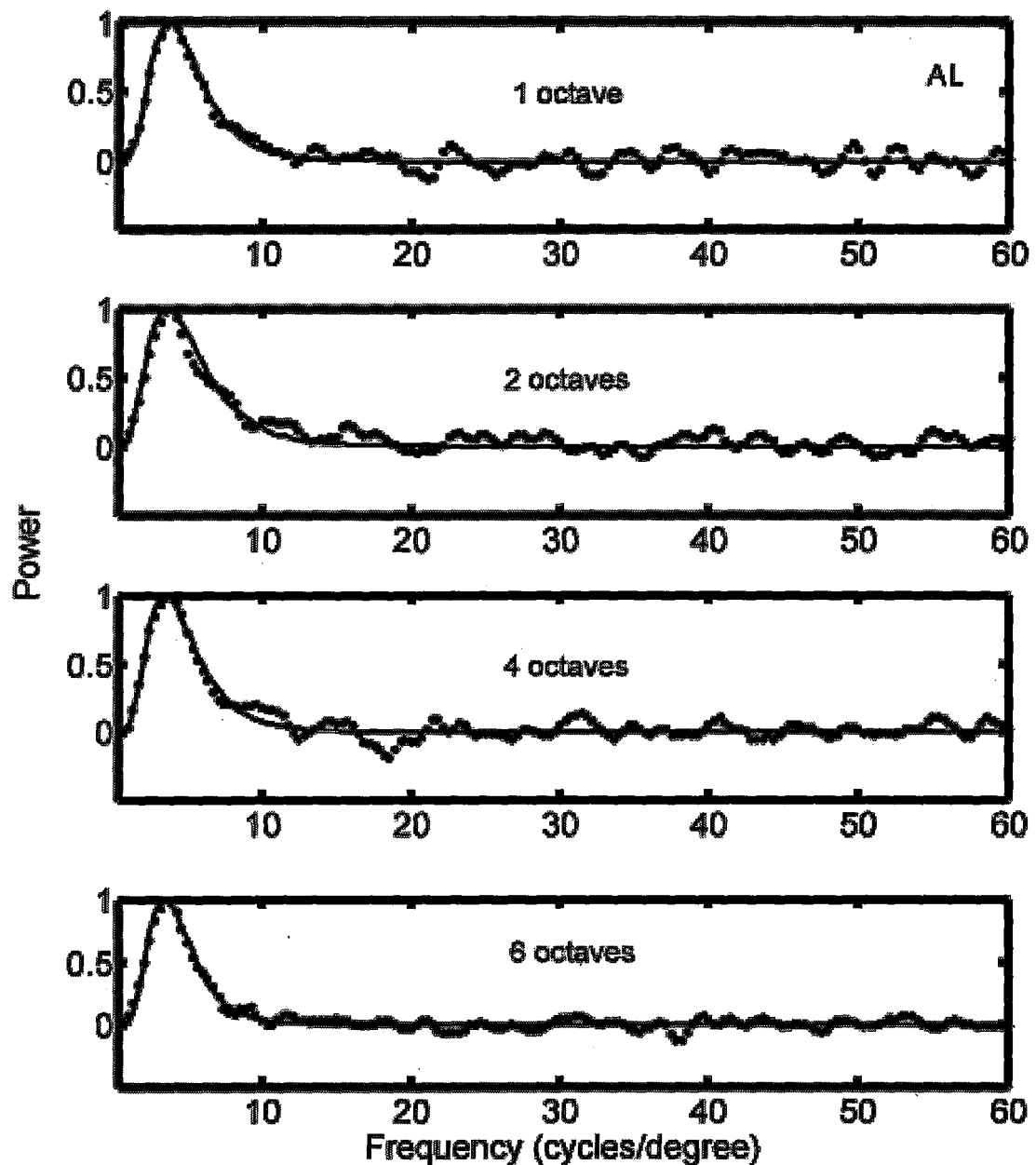


Figure 2.10: Classification images for one observer (AL) in the 5 cy/deg center-frequency conditions. Each panel shows the classification image obtained with a different stimulus bandwidth.

when the center frequency of the target was 15 cy/deg. A comparison of Figures 2.10 and 2.11 shows that the bandwidths of the classification images obtained with 5 and 15 cy/deg stimuli also were very similar, although it appears that the bandwidth was slightly greater with the high spatial frequency stimulus. These qualitative impressions were investigated quantitatively by fitting the classification images with a log-normal equation:

$$Power(x|\mu, \sigma) = \frac{1}{\log(\sigma)\sqrt{2\pi}} e^{\frac{-(\log(x)-\mu)^2}{2\log(\sigma)^2}} \quad (2.2)$$

Prior to fitting the log-normal function to the data, the classification images were smoothed by averaging values at adjacent frequencies and then re-normalizing the smoothed images to have a peak of one. The best-fitting log-normal functions are indicated by the smoothed lines in Figures 2.10 and 2.11. Goodness of fit was evaluated using three statistics: the Mean Absolute Deviation (*MAD*), Hotelling T^2 statistic, and R^2 . In all cases the *MAD* for the residuals of our fitted normalized templates was less 0.04 units of normalized power for all observers in all conditions. Hotelling's T^2 statistic was used to evaluate the deviations between the classification images and the best-fitting log-normal functions [1]: For each observer in each condition, the Hotelling T^2 statistic was very small ($T^2 < 1$) which indicated that there were no significant differences between the fitted functions and our calculated templates. R^2 was also calculated for each observer in each condition for the fitting function and the classification images and it always exceeded a value of 0.92. Each of the three statistics computed indicates that the log-normal channels fitted to the data provided a very good description of the shape of the template for noise detection.

Confidence intervals for the estimates of the center-frequency and bandwidth of the classification images were estimated using a parametric bootstrap procedure on the residuals of the best-fitting log-normal function for each observer in each condition [11]. On each run of the simulation, a zero mean Gaussian random variable with a variance equal to the observed residual variance was added to the fitted log-normal function to create synthetic data. A log-normal was then fit to the synthetic data, and the resulting estimates of center-frequency and bandwidth were recorded. This procedure repeated 10,000 times, and the bootstrapped distributions of center-frequency and bandwidth were used to estimate the 95% confidence intervals that are shown in Tables 2.3 and 2.4 .

For all observers, the center-frequency of the template was between 3-5 cy/deg, which is near the peak of the contrast sensitivity function [9], and did not differ systematically

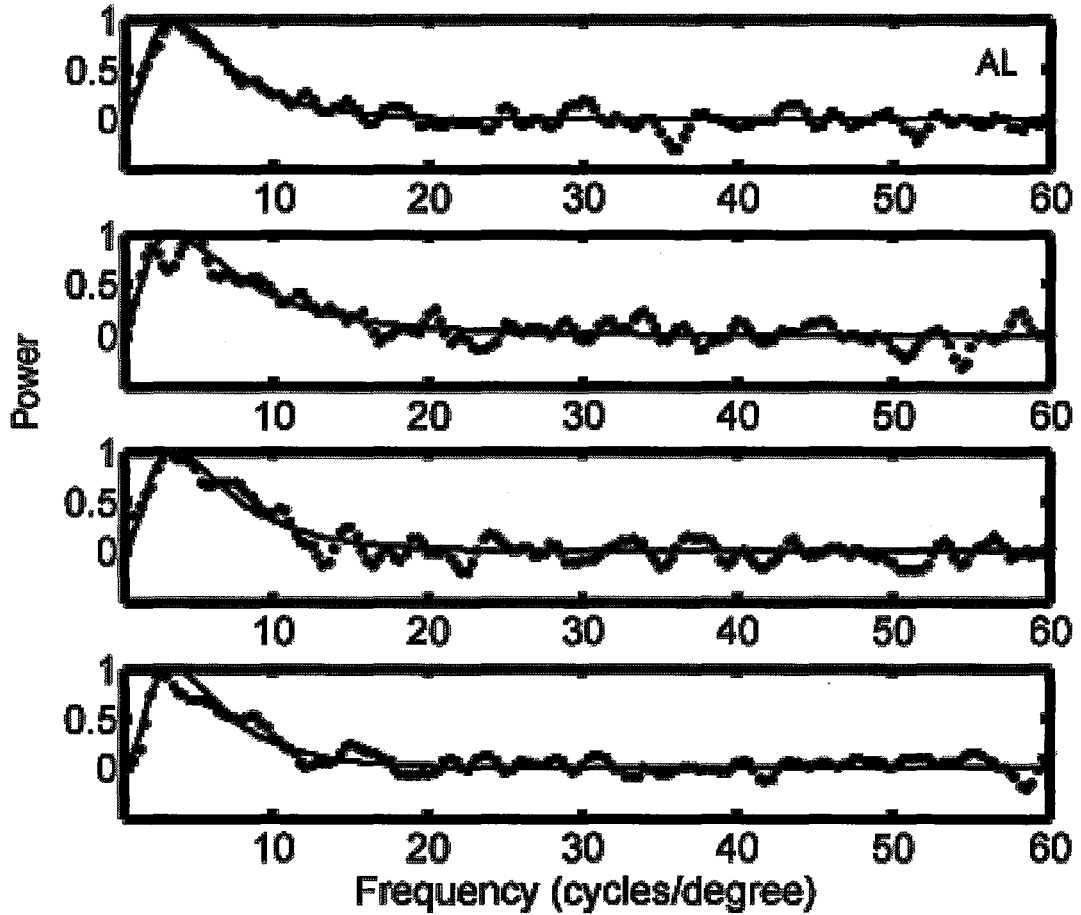


Figure 2.11: Classification images for one observer (AL) in the 15 cy/deg center-frequency condition, shown in each of the four noise bandwidth conditions. The order of the panels is the same as in the previous figure, from 1 octave to 4 octaves.

Table 2.3: Center-frequencies and 95% confidence intervals for three observers.

sf	octaves	AL	MB	VA
5	1	3.71 ± 0.11	3.96 ± 0.21	5.65 ± 0.61
	2	3.62 ± 0.11	3.58 ± 0.34	4.92 ± 0.81
	3	3.51 ± 0.22	3.55 ± 0.66	5.82 ± 0.57
	4	3.49 ± 0.89	3.41 ± 1.08	5.32 ± 0.57
15	1	3.45 ± 0.24	4.41 ± 0.28	4.56 ± 0.81
	2	3.45 ± 0.25	4.71 ± 0.32	5.35 ± 0.71
	3	3.60 ± 0.77	5.93 ± 0.45	5.37 ± 0.50
	4	3.54 ± 0.38	3.94 ± 0.80	5.34 ± 0.22

across conditions. The bandwidths of the classification images did not vary systematically with stimulus bandwidths, but was on average lower in the 5 cy/deg condition (mean = 1.1 octaves) than in the 15 cy/deg condition (mean = 1.4 octaves).

Table 2.4: Octave bandwidths and 95% confidence intervals for three observers.

sf	octaves	AL	MB	VA
5	1	1.07 ± 0.04	1.02 ± 0.09	1.30 ± 0.15
	2	1.18 ± 0.04	1.01 ± 0.09	1.24 ± 0.24
	3	1.09 ± 0.03	1.07 ± 0.06	1.24 ± 0.13
	4	1.03 ± 0.02	0.99 ± 0.08	1.39 ± 0.15
15	1	1.37 ± 0.05	1.58 ± 0.08	1.45 ± 0.11
	2	1.48 ± 0.03	1.48 ± 0.08	1.34 ± 0.12
	3	1.36 ± 0.04	1.46 ± 0.03	1.41 ± 0.16
	4	1.33 ± 0.02	1.58 ± 0.07	1.49 ± 0.13

2.4.3 Non-linear contributions to the classification image?

The classification images shown in Figures 2.10 and 2.11 are estimates of the weight of the observer's linear template applied to the contrast power at each frequency. What these templates do not show, however, are the non-linear operations on the power spectrum that could have influenced observers' responses. To examine the possible influence of such mechanisms on observers' responses, we computed separate classification images from the signal-absent and signal-present noise fields. The classification images derived from the signal-absent noise fields are shown for one observer with the 5 cy/deg center frequency stimuli in Figure 2.12. Templates from other observers in other conditions were very similar.

Under a linear observer model, the shape of the templates derived from signal-present and signal-absent noise fields should be the same except for a change in sign [1]. If, however, non-linear operations significantly influenced behaviour in our task, then the templates generally will differ. A quantitative statistical test based on Hotelling T^2 [1] failed to find a significant difference between the shapes of the classification images obtained from signal-absent and signal-present noise fields. Hence, we did not find evidence that non-linearities contributed significantly to the decisions made by observers in our task.

It is important to be cautious when interpreting the results of the test for non-

linearities described above. If one observes similar templates in both the signal-present and signal-absent conditions, the result does not support the linear model but instead should be interpreted as providing no support that response non-linearities that can be revealed by this test are present. This test has been shown to reveal both spatial and phase uncertainty [5, 6] when classification images are computed in the spatial domain. However, here we are computing classification images in the power spectrum and not the spatial domain. In the power spectrum, the non-linearity that is analogous to spatial uncertainty would be spatial-frequency uncertainty. The current analysis is sensitive to spatial-frequency uncertainty, [1, 2] and therefore the failure to find an effect of non-linearities suggests that spatial-frequency uncertainty is not influencing performance in our task. However, we cannot rule out the possibility that other kinds of non-linearities are involved in the detection of visual noise.

2.4.4 Discussion

Experiment 2.4 yielded two unexpected findings. First, the classification images suggest that observers were most influenced by power at 3-5 cy/deg when detecting noise stimuli centered at 5 and 15 cy/deg. Second, although the bandwidth of the classification image was slightly higher in the high spatial frequency condition, it was independent of stimulus bandwidth. In other words, the classification images suggest that observers did *not* adjust the bandwidth of an internal filter to match the stimulus bandwidth. Overall, the results of Experiment 2.4 suggest that the spatial frequencies that were linked to observers' behaviour were remarkably constant across conditions.

The results of Experiment 2.4 present us with a conundrum. On the one hand, the slopes of the TvB functions measured in previous experiments suggest that observers sum information optimally across a wide range of spatial frequency, perhaps by adjusting the bandwidth of an internal channel. On the other hand, the classification images provide no evidence for an adjustable channel. This failure to find evidence of an adjustable channel raises the question of whether bandwidth adjustment is necessary to produce TvB functions with slopes of 0.25. To address this question, we used a multiple-channel model described by Wilson and Gelb [28], which has been used to account for a wide variety of experimental results.

The Wilson-Gelb model includes six fixed spatial frequency channels. A fixed-variance is added to the output of each channel, and the responses are then combined using non-linear (Minkowski) summation. To apply this model to our noise detection data, we first

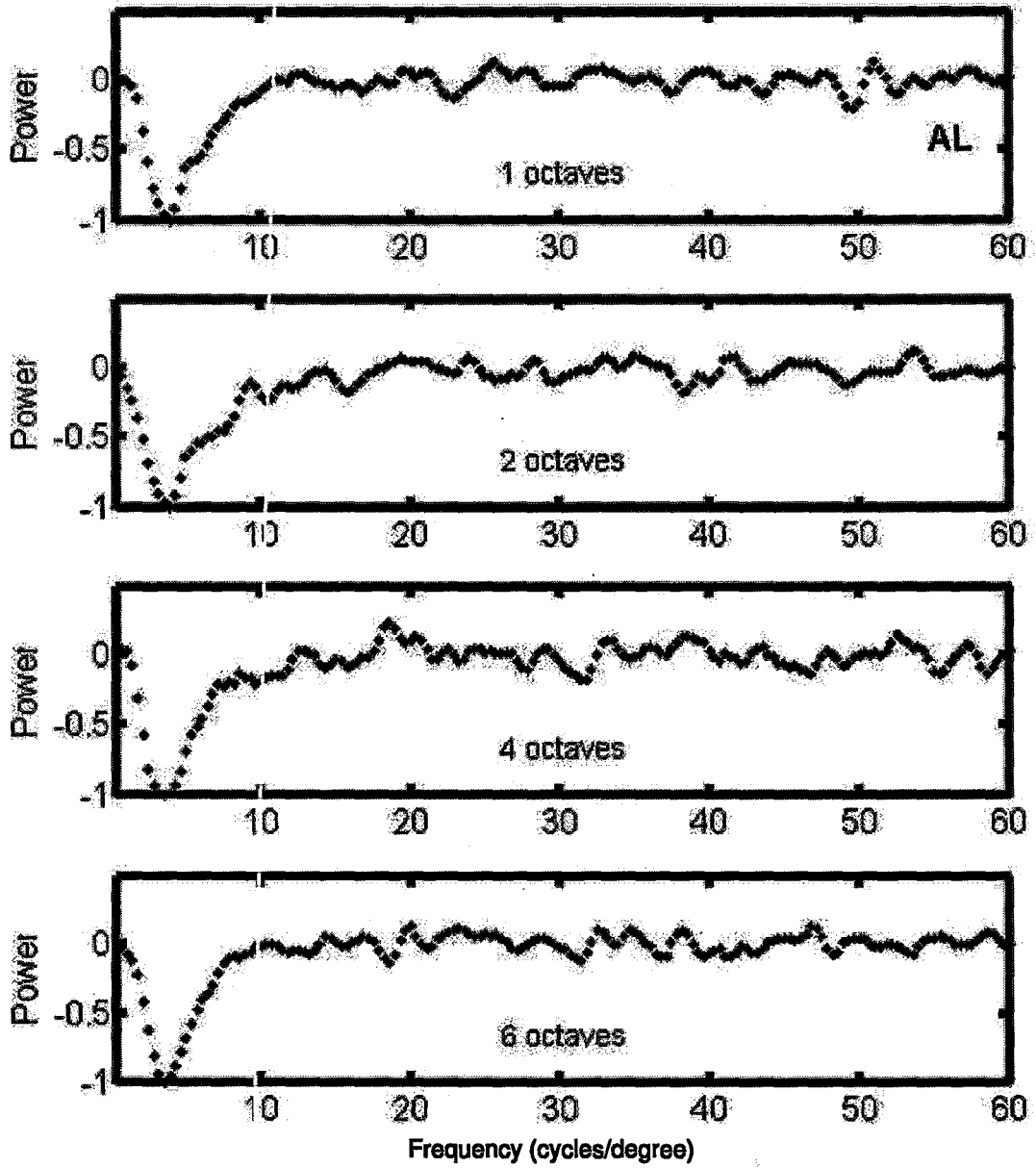


Figure 2.12: Classification images for one observer (AL) generated from the signal-absent noise fields.

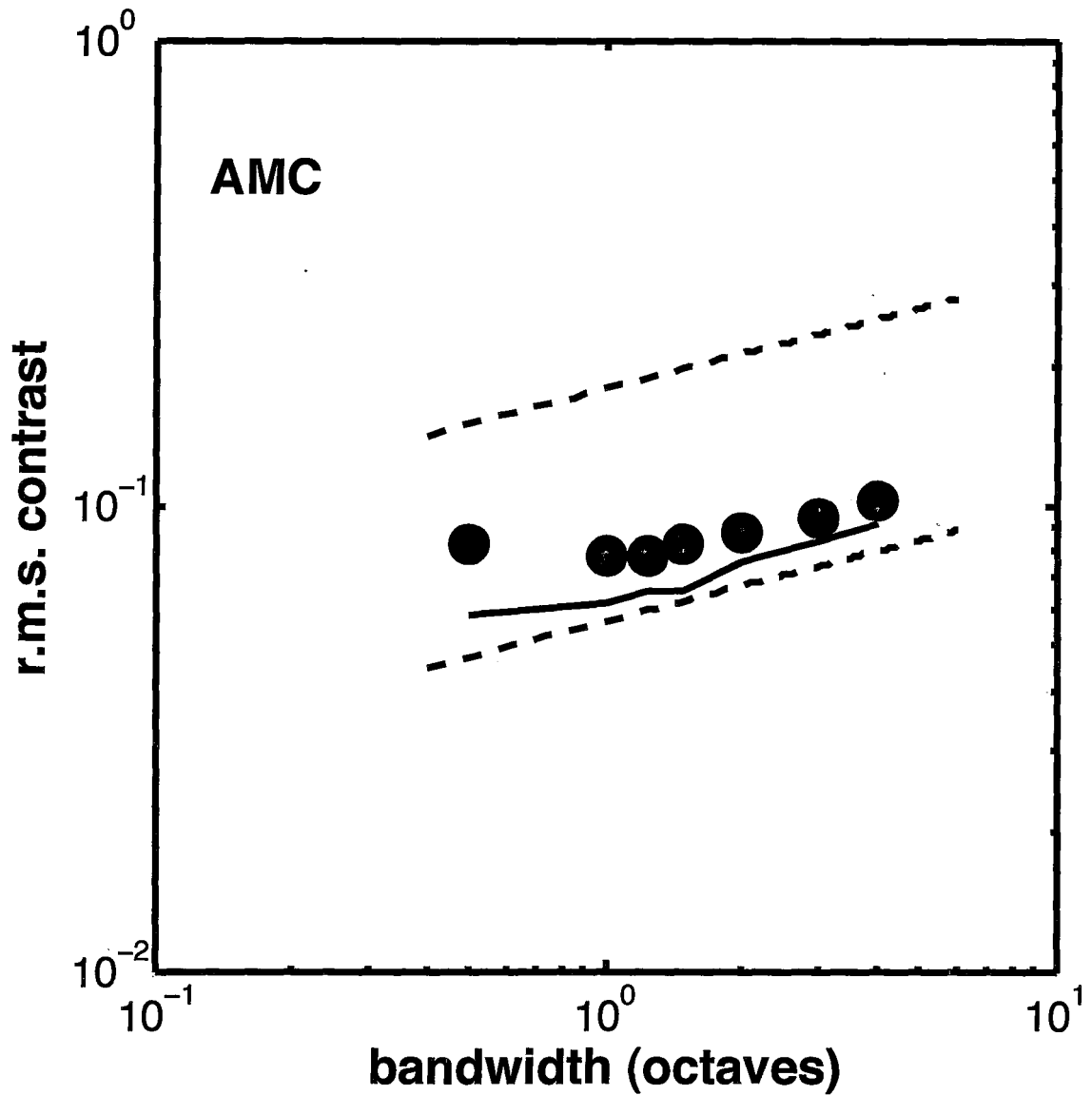


Figure 2.13: The filled circles represent the TvB function for one observer. The solid line is the TvB produced by the Wilson-Gelb model which has no free parameters and is not shifted to fit the data. The dashed lines represent a slope of 0.25.

measured the CSF for two observers using narrow-band (0.5 octaves) noise at a range of center-frequencies. The CSF was used to set the sensitivities of the six spatial-frequency channels in the model: this adjustment allowed the model to produce a CSF that had the same shape as the human observer. Next, we adjusted the variance of the noise so that peak sensitivity of the model's CSF matched that of the human observer. Finally, we tested the effects of varying the Minkowski summation exponent across a range of values: similar results were obtained with exponents in the range of 2-5, but we settled on a value of 2 because it had been used with success previously to capture a variety of other results [28]. The parameters of the model were then fixed, and simulations were performed in MATLAB to estimate the model's thresholds for our noise stimuli.

For one observer, the model slightly under-predicted thresholds in all conditions, although the error was quite small for the broadest bandwidth conditions (Figure 2.13). For the other observer, the model accurately accounted for thresholds obtained with narrow bandwidth stimuli, but slightly under-predicted thresholds when stimulus bandwidth was greater than one octave (Figure 2.14). However, in both cases, for stimulus bandwidths greater than one octave, the slopes of the TvB functions generated by the model and obtained from the human observer were very similar to each other and to the ideal value of 0.25. Figure 2.15 shows classification images for the model: as was found with human observers (Figure 2.10), the classification images are remarkably constant across conditions both in terms of peak power ($\approx 5\text{cy/deg}$) and bandwidth.

2.5 Conclusion

Experiment 2.1 replicated the findings of Kersten [20]: absolute efficiency for noise detection was high and nearly constant across a wide range of stimulus bandwidths. Experiments 2 and 3 obtained similar results when subjects were uncertain about stimulus bandwidth, when stimulus duration was reduced from 200 ms to 12 ms, and when the center frequency of the stimulus was increased from 5 to 15 cy/deg. These results are surprising because they indicate that human observers can sum spatial frequency information across bandwidths that exceed the bandwidths of individual channels typically used in multiple-channel models of pattern detection [15].

One explanation for these results is that noise is detected with an internal filter that has a bandwidth that is adjusted to match the stimulus [18, 20]. Based on the results of Experiment 2.2, adjustment of the internal filter could be based on a bottom-up process that uses information available in a brief presentation of the stimulus. The results of

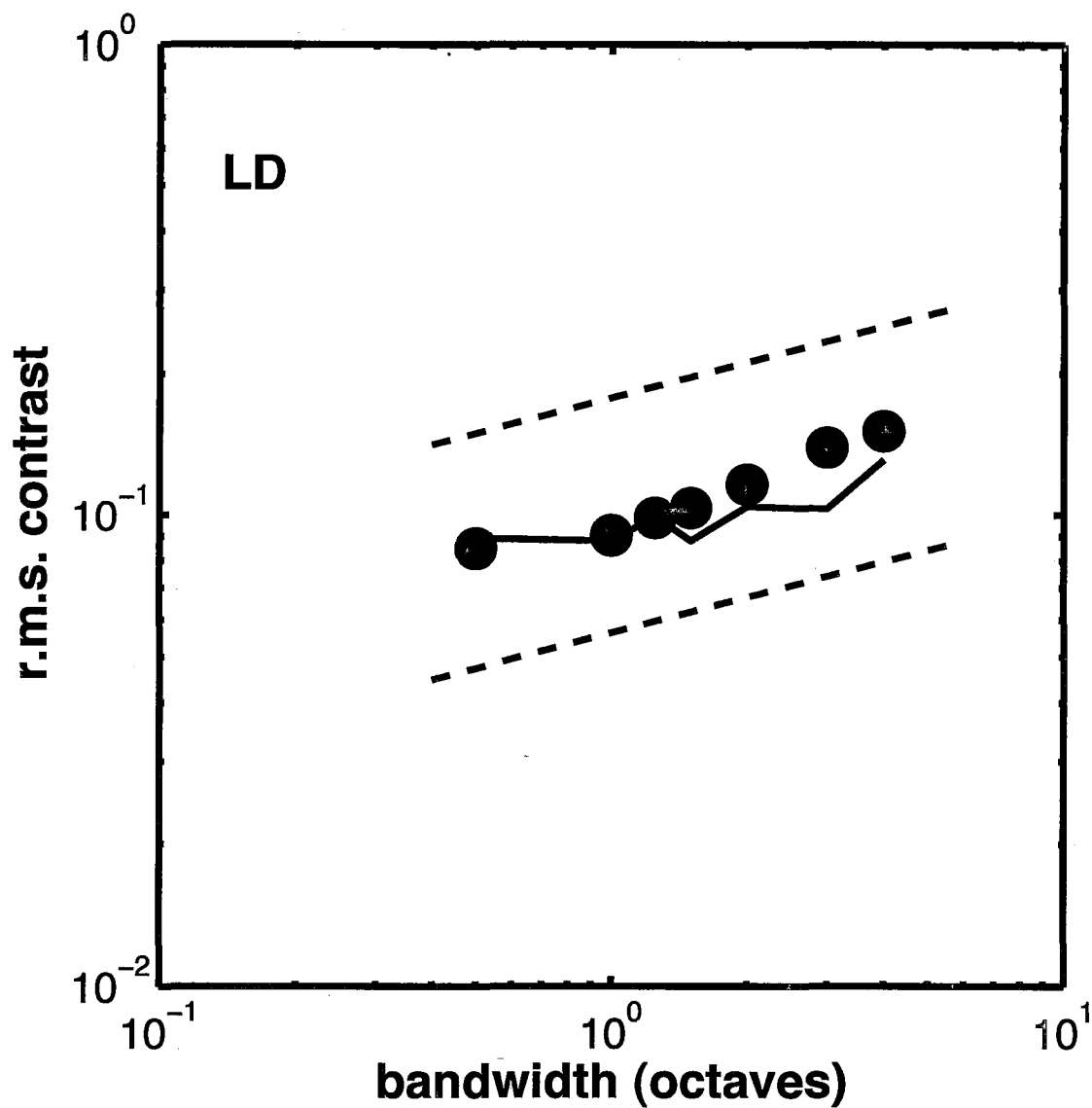


Figure 2.14: The filled circles represent the TVB function for one observer. The solid line is the TvB produced by the model which has no free parameters and is not shifted to fit the data. The dashed lines represent a slope of 0.25.

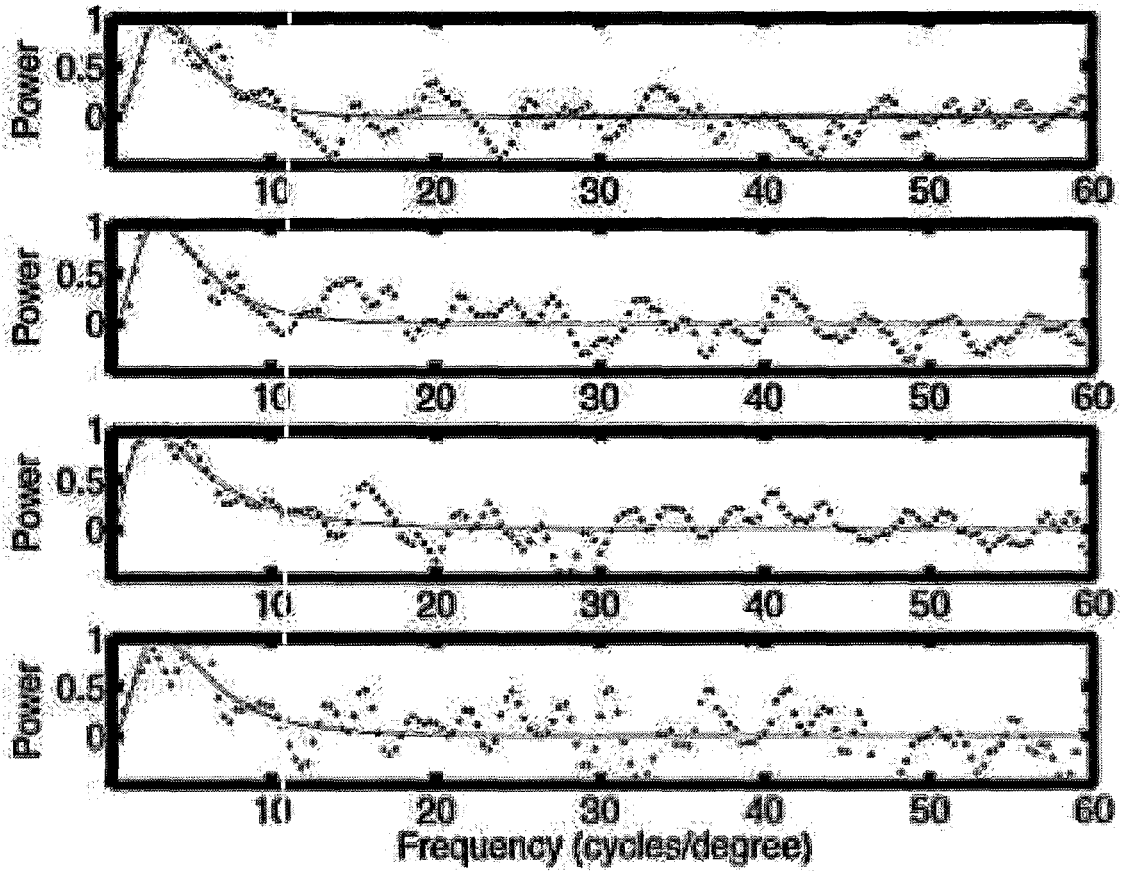


Figure 2.15: Classification images for the Wilson-Gelb model applied to the noise detection task.

Experiment 2.4, however, are inconsistent with a adjustable-bandwidth model: the classification images measured in that experiment did not vary significantly or systematically with stimulus bandwidth. Indeed, the results of Experiment 2.4 are consistent with the hypothesis that used a single, fixed channel – or a single, fixed algorithm for combining the responses of multiple channels – to detect noise stimuli of various bandwidths. Simulations based on the Wilson-Gelb multiple-channel model [28] were consistent with this idea. Once the model parameters were adjusted to account for the contrast sensitivity functions of our observers, the Wilson-Gelb model was able to account for the main findings of our experiments. Specifically, the model had a high absolute efficiency for noise detection that was nearly constant for stimulus bandwidths greater than or equal to one octave, and the peak frequencies and bandwidths of the classification images derived from the model responses were approximately constant across a four-fold range of stimulus bandwidths.

The Wilson-Gelb model does not have variable-bandwidth channels, nor does it combine the outputs of different channels in an optimal matter. Nevertheless, that model produces TvB curves that have nearly the same slope as the curve produced by an ideal detector. We conclude, therefore, that standard multiple-channel detection models can account for ideal frequency summation for visual (and, presumably, auditory) noise [18, 20].

References

- [1] C.K. Abbey and M.P. Eckstein. Classification image analysis: estimation and statistical inference for two-alternative forced-choice experiments. Journal of Vision, 2 (1):66–78, Jan 2002.
- [2] C.K. Abbey, M.P. Eckstein, and F.O. Bochud. Estimation of human-observer templates in two-alternative forced-choice experiments. Proceedings of SPIE, 3663: 284–295, Jan 1999.
- [3] A. Ahumada and J. Lovell. Stimulus features in signal detection. Journal of the Acoustical Society of America, 49:1751–1756, Jan 1971.
- [4] M.S. Banks, W.S. Geisler, and P.J. Bennett. The physical limits of grating visibility. Vision Research, 27(11):1915–24, 1987.
- [5] E. Barth, B.L. Beard, and A.J. Ahumada. Nonlinear features in vernier acuity. Proceedings of SPIE, 3644:88–96, May 1999.

- [6] B.L. Beard and A.J. Ahumada. Detection in fixed and random noise in foveal and parafoveal vision explained by template learning. Journal of the Optical Society of America A, 16(3):66–78, Jan 1999.
- [7] D. Brainard. The psychophysics toolbox. Spatial Vision, 10:443–446, Jan 1997.
- [8] E.T. Davis and N. Graham. Spatial frequency uncertainty effects in the detection of sinusoidal gratings. Vision Research, 21(5):705–712, 1981.
- [9] R.L. DeValois and K.K. DeValois. Spatial Vision. Oxford University Press, New York, 1988.
- [10] M.P. Eckstein, S.S. Shimozaki, and C.K. Abbey. The footprints of visual attention in the Posner cueing paradigm revealed by classification images. Journal of Vision, 2(1):25–45, 2002.
- [11] B. Efron and B.J. Tibshirani. Introduction to the Bootstrap. Chapman & Hall, 1994.
- [12] J.M. Gold, R.F. Murray, P.J. Bennett, and A.B. Sekuler. Deriving behavioural receptive fields for visually completed contours. Current Biology, 10(11):663–6, Jun 2000.
- [13] N. Graham and J. Nachmias. Detection of grating patterns containing two spatial frequencies: a comparison of single-channel and multiple-channel models. Vision Research, 11(3):251–259, Jan 1971.
- [14] N. Graham, J.G. Robson, and J. Nachmias. Grating summation in fovea and periphery. Vision Research, 18(7):815–25, Jan 1978.
- [15] N.V.S. Graham. Visual Pattern Analyzers. Oxford University Press, New York, 1989.
- [16] D.M. Green. Auditory detection of a noise signal. In J.A. Swets, editor, Signal detection and recognition by human observers, pages 523–547, New York, 1960. Wiley.
- [17] D.M. Green. Auditory detection of a noise signal. Journal of the Acoustical Society of America, 32(1):121–131, Jan 1960.

- [18] D.M. Green. Auditory detection of a noise signal. In J.A. Swets, editor, Signal detection and recognition by human observers, pages 523–547, New York, 1960. Wiley.
- [19] R. Hübner. The efficiency of different cue types for reducing spatial-frequency uncertainty. Vision Research, 36(3):401–408, 1996.
- [20] D. Kersten. Statistical efficiency for the detection of visual noise. Vision Research, 27(6):1029–1040, Jan 1987.
- [21] V.A. Lamme and P.R. Roelfsema. The distinct modes of vision offered by feedforward and recurrent processing. Trends in Neurosciences, 23(11):571–9, Nov 2000.
- [22] D.G. Pelli. The videotoolbox software for visual psychophysics: transforming numbers into movies. Spatial Vision, 10(4):437–442, Jan 1997.
- [23] A.B. Sekuler, C.M. Gaspar, J.M. Gold, and P.J. Bennett. Inversion leads to quantitative, not qualitative, changes in face processing. Current Biology, 14(5):391–6, Mar 2004.
- [24] C.W. Tyler, H. Chan, L. Liu, B. McBride, and L.L. Kontsevich. Bit stealing: how to get 1786 or more gray levels from an 8-bit color monitor. In B.E. Rogowitz, editor, Human Vision, Visual Processing, and Digital Display III, volume 1666, pages 351–364. SPIE, 1992.
- [25] B.A. Wandell. Foundations of Vision. Wiley, New York, 1995.
- [26] G.H. Wetherill and H. Levitt. Sequential estimation of points on a psychometric function. British Journal of Mathematical & Statistical Psychology, 18:1–10, Jan 1965.
- [27] H.R. Wilson and J.R. Bergen. A four mechanism model for threshold spatial vision. Vision Research, 19(1):19–32, Jan 1979.
- [28] H.R. Wilson and D.J. Gelb. Modified line-element theory for spatial-frequency and width discrimination. Journal of the Optical Society of America A, 1(1):124–131, 1984.
- [29] H.R. Wilson and F. Wilkinson. Evolving concepts of spatial channels in vision: From independence to nonlinear interactions. Perception, 26:939–960, Jan 1997.

Chapter 3

Detection and discrimination of visual noise

Abstract

In Chapter 2, our classification image experiments found that the peak-sensitivity of the channel used for noise detection did not vary with the center frequency (CF) of the stimulus. For both 5 and 15 cy/deg stimuli, the classification image peaked at approximately 5 cy/deg. This result leads to the counter-intuitive hypothesis that observers ought to be unable to discriminate filtered noise with CFs of 5 and 15 cy/deg when the contrast of the patterns are at detection threshold. This hypothesis was investigated in this chapter. In Experiment 3.1, we used the methods of double-judgment psychophysics (Klein, SA, *J Opt Soc Am A*, 1985, 2(9), 1560-1585) to examine the relation between detection and discrimination thresholds for stimuli with CFs of 5 and 15 cy/deg. The results of Experiment 3.1 showed that the discriminability of the two patterns can be predicted from the detection data (Watson, AB, *Vision Res*, 1981, 21(7) 1115-1122). In Experiment 3.2, we used the classification image technique to measure the channel used for both detection and discrimination of visual noise stimuli that differed in CF. We found that the difference in bandwidth between the two templates for stimulus detection was a good match to the channel estimated for stimulus discrimination.

Introduction

In Chapter 2 we found the puzzling result: classification images for the detection of 5 and 15 cy/deg patterns did not differ in their center frequency and, differed only slightly in their bandwidth. The goal of this chapter is to investigate this result using the methods of double-judgement psychophysics and the classification image technique. To put the experiments in context, what follows is a brief description of the methods that lead to our surprising results.

In our previous experiments, observers performed a one-dimensional noise detection task [12] in a background of Gaussian white noise. We used the classification image technique [2] to measure which spatial frequencies were used to do the detection task. The results of our classification images analysis revealed that observers used a template that had a center frequency of 3–5 cy/deg when the signal had a center-frequency of 5 or 15 cy/deg. As mentioned previously, the two classification images did not differ in center frequency and differed only slightly in their bandwidth. This difference was less pronounced than expected, because an ideal observer would use channels that centered on the stimulus center frequency that matched the bandwidth of the stimulus exactly. As we found only small differences in the bandwidth of the channels in our classification images for the detection of visual noise, we wondered whether the classification images could capture visual discrimination performance. The relationship between detection and discrimination has been well described by signal detection theory [11, 14] and if our observers are using the channels we measured for detection, then both observers' thresholds and classification images ought to predict discrimination thresholds.

The methods of double-judgement psychophysics relate detection and discrimination. The results of Chapter 2 lead to the hypothesis that observers ought to be unable to discriminate filtered noise with CFs of 5 and 15 cy/deg when the contrast of the patterns are at detection threshold, which one would not predict from previous work [7, 10]. Relating detection and discrimination addresses this hypothesis because if, as the classification images suggest, there is little difference between the channel used for detecting filtered noise with CFs of 5 and 15 cy/deg then at detection threshold, discrimination ought to be difficult for human observers. The next section explains the double-judgment psychophysics in more detail.

Relating detection and discrimination

The work that has addressed the relationship between detection and discrimination has been called double-judgement psychophysics [13]. In double-judgement psychophysics, detection and discrimination are measured together to relate the two measurements. Often, the double-judgement procedure uses a variant of the two interval forced-choice (2-IFC) procedure. In a double-judgement task, observers provide two responses, a detection response that indicates the interval that the target appeared. Then a discrimination response that indicates which of two target stimuli were presented. A classic example of this technique is presented in a paper by Nachmias and Weber [16] in which a double-judgment procedure was used to show that 3 and 9 cy/deg sinusoidal gratings are discriminated as soon as they are detected.

To make inferences about the relation between detection and discrimination, a model of visual processing must be assumed. The assumptions that go into the model, however, are common and are incorporated into the standard, back-pocket, multiple channel model of pattern vision [6]. The key aspects of the back-pocket model that are relevant to double-judgement task are that visual stimuli are encoded by multiple, labeled, noisy, spatial frequency channels. The output of the channels in the model are then passed through a decision stage that often uses non-linear summation to transform the responses from multiple channels into a single decision variable. The back-pocket model performs a detection task by using a criterion to determine whether the signal is reliably present or absent. When the observer makes a discrimination response, the observer uses a decision rule that determines which of the two stimuli that was presented. Models of this sort have been used to account for a great variety of detection and discrimination data [24].

Figure 3.1 represents how responses of a multiple channel can be used to detect and discriminate two stimuli, S1 and S2. The figure illustrates a decision space that is formed by two dimensions, A and B. These dimensions are not totally abstract — they represent the responses of two channels, A and B. S1 and S2 represent the *mean* activity produced on trials that contained either stimulus 1 or stimulus 2, whereas N represents the mean activity produced on noise, or no-stimulus trials. The decision space model is general: the two stimuli could vary on any number of stimulus dimensions (e.g., motion, color, spatial frequency, orientation, etc.). Likewise, the selectivity of the two channels represented by A and B could vary on a variety of dimensions. The left panel of Figure 3.1 illustrates the case in which the responses of the two channels are orthogonal and independent, whereas the right panel illustrates the case where the responses are not orthogonal.

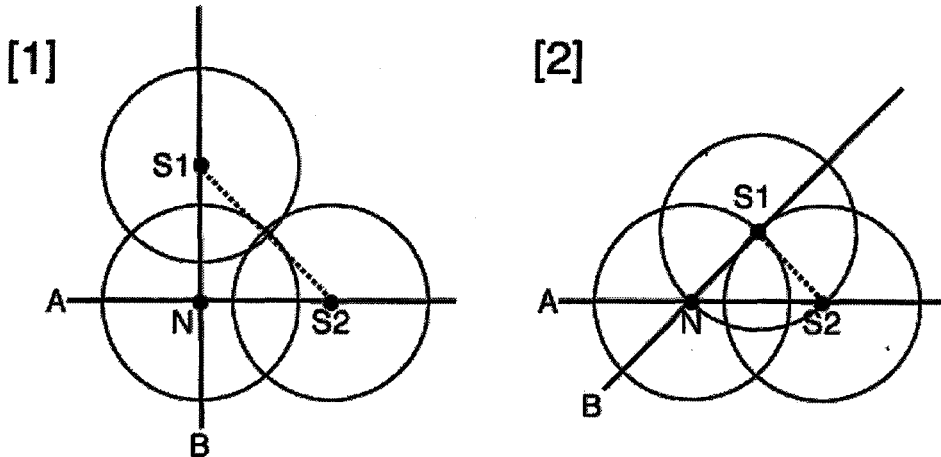


Figure 3.1: A schematic diagram of the sensory decision space underlying detection and discrimination judgments in the 'back-pocket' model [6]. A and B correspond to two sensory dimensions that are independent/orthogonal (panel [1]) and non-orthogonal (panel [2]). See the text for details.

In this model, the sensory distance between N and S1 can be estimated by measuring detection d' (where a smaller/larger d' corresponds to a smaller/larger sensory distance) for a stimulus that excites only dimension B. The sensory distance between N and S2 can be determined by the same method. Once d' for the two detection tasks has been measured, predictions can be made about the sensory distance between S1 and S2 (i.e., the dashed line in Figure 3.1) as measured by a discrimination task with the two stimuli S1 and S2. If the sensory distances between N and S1 and between N and S2 are both d' , then the orthogonal-channel model depicted in the left panel of Figure 3.1 predicts that the discrimination sensitivity between S1 and S2 is $\sqrt{2}d'$. If the underlying sensory dimensions A and B are not orthogonal/independent, then the model predicts discrimination sensitivity will be less than $\sqrt{2}d'$. It is these predictions that we set out to investigate in Experiment 3.1, where a double-judgement procedure was used to measure the detection and discrimination of visual noise simultaneously.

Experimental Motivation

The center-frequency of the classification images in Chapter 2 were very similar in both the 5 and 15 cy/deg center-frequency conditions, they were not identical however, as differences were found in the channel bandwidths. This result seems to run contrary to the idea that widely separated spatial frequencies are processed by independent channels

tuned to the stimulus center frequency [10]. If it is the case that the classification images that we measured in Section 2.4 are evidence against independence among channels, then a double-judgment task ought to reveal that observers' sensitivity in a detection can not predict their ability to discriminate patterns. This idea was tested in Experiment 3.1.

Experiment 3.2, uses the classification image technique to measure both detection and discrimination. The technique is able to address if the templates used for detection channel are related to those for discrimination, as would be predicted from previous work. Prior to the work in Experiment 3.2, the classification image technique had not been used to relate templates in the two tasks using the same stimuli.

3.1 Experiment 3.1

To investigate the relationship between detection and discrimination, this experiment measured sensitivity in both tasks simultaneously using a double-judgement procedure.

3.1.1 Methods

3.1.1.1 Observers

The two observers in this experiment were members of the McMaster University community and were paid for their participation. Both observers were unaware of the experimental hypotheses, had normal or corrected-to-normal visual acuity, and had extensive practice with this and other visual psychophysical tasks. The two observers were each 28 years of age.

3.1.1.2 Apparatus

A Macintosh G4 running MATLAB and the Psychophysics and Video toolboxes [5, 17] was used to generate the stimuli. The stimuli were presented with a Sony GDM-F520 monitor set to a resolution of 1024×768 pixels. The entire display subtended a visual angle of 10.8×8.3 deg at the viewing distance of 2 m. The frame rate of the display was 75 Hz and the mean luminance 45 cd/m^2 . The luminance of the display was calibrated using a PhotoResearch PR-650 photometer before each session. A Cambridge Research System Bits++ device was used to achieve fine grained (i.e., 14-bit) control of contrast. A chin/forehead rest was used to stabilize viewing position.

3.1.1.3 Stimuli

The stimuli were one-dimensional (horizontal) Gaussian white noise patterns that were spatially filtered with an ideal frequency filter that was centered on 5 or 15 cy/deg and that had a full bandwidth of two octaves. Stimulus contrast was modulated with a circularly-symmetric Gaussian envelope with a standard deviation of 1.08 degrees of visual angle. Examples of patterns similar to those shown in the experiment are shown in Figure 3.2. Stimuli were presented in a background of white noise with a contrast variance of 0.32. A new sample of signal noise was generated on each interval of every trial. Stimulus duration was 200 ms. The monitor provided the only illumination in the testing room. Stimuli were viewed binocularly through natural pupils.

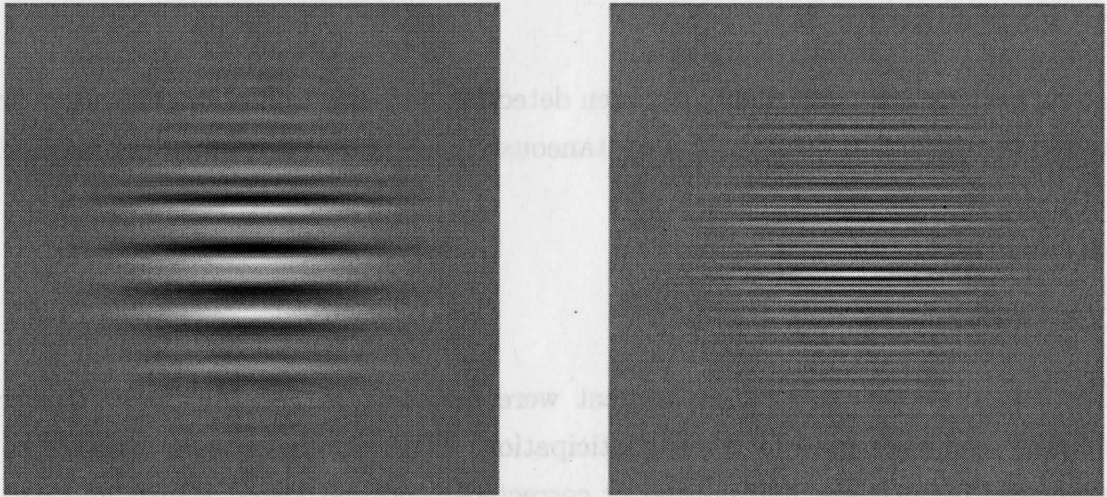


Figure 3.2: Two examples of the patterns of the type presented in this experiment. The pattern on the right has a center frequency three times that of the pattern on the left. Both patterns are filtered with to have a full bandwidth of two octaves.

3.1.1.4 Procedures

Observers were instructed to fixate a high-contrast dot located in the center of the display. The observer started each trial by pressing the spacebar on the keyboard. After a delay of 50 ms, the fixation point was removed and then, after another 50 ms delay, was followed by two 200 ms stimulus intervals separated by a blank, 300 ms inter-stimulus interval. Each stimulus interval was marked by a clearly-audible tone. To reduce spatial uncertainty, a 3×3 deg frame, drawn with a two-pixel wide line that was set to maximum contrast, was centered on the fixation point and was visible continuously during the entire duration of each trial. To reduce adaptation, the frame had a 50% probability of being

black or white on each trial.

Observers performed a double-judgment task [16] that consisted of detection and discrimination responses. For the detection task, a two-interval forced-choice (2-IFC) procedure was used: Observers reported whether they observed a noise signal in either the first or second interval. After the detection response, observers then reported whether the 5 or 15 cy/deg stimuli was presented.

Stimulus contrast variance was varied across trials using four interleaved staircases, two for each of the 5 and 15 cy/deg stimuli. For each pair of staircases, a 2-down/1-up staircase converged on the 71% correct point of the psychometric function and a 4-down/1-up staircase converged on the 84% correct point [22]. The detection responses were used to control the direction of the staircases. For example, two consecutive correct detection responses would cause the 2-down/1-up staircase to reduce stimulus contrast, irrespective of the correctness of the discrimination responses. The initial value of the staircase was set by thresholds measured for each observer in two practice sessions that were completed before the main experiment. Excluding practice, each observer completed ten sessions of 1200 trials, for a total of 12000 trials.

In the main experiment, the staircases could set stimulus contrast to one of only 4 values distributed around the detection thresholds measured in the practice sessions. By restricting the number of possible contrasts, we ensured that detection and discrimination d' at each contrast level was measured with a large number of trials. For each observer, d' was computed for the detection and discrimination tasks using standard formulae described by Macmillan and Creelman [14].

3.1.2 Results

Figures 3.3 and 3.4 show the results for two observers. Each observer's sensitivity in both the detection and discrimination tasks increased as stimulus contrast increased. Also, sensitivity in the discrimination task exceeds sensitivity in the detection task at all levels of stimulus contrast. The dashed lines in both graphs are the least squares fits to the detection data. The solid line in each figure indicates a fit to the discrimination thresholds that were *predicted* (the predictions are plotted with grey X symbols) from the detection data using signal detection theory [11, 14] assuming that the 5 and 15 cy/deg stimuli were detected by orthogonal channels. According to this model, d' for discrimination follows the equation:

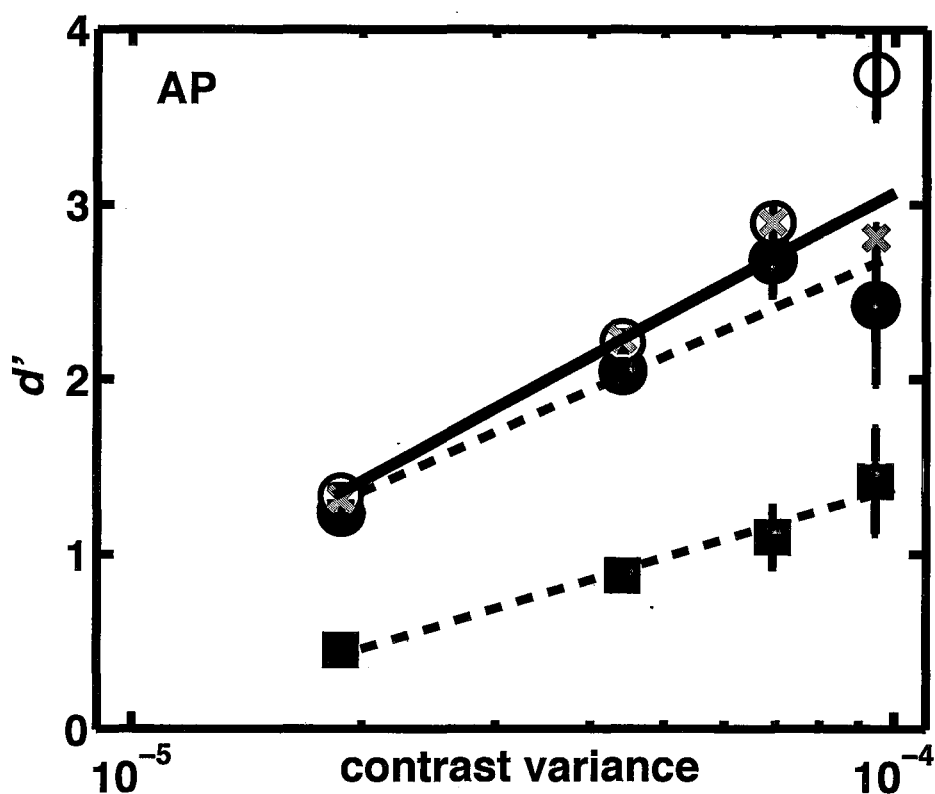


Figure 3.3: d' plotted as a function of contrast variance for observer AP in the double-judgement task. The error-bars represent 95% confidence intervals. The squares are thresholds for the 5 cy/deg detection condition, the filled circles thresholds for the 15 cy/deg detection condition, the open circles the observer's discrimination data, and the X symbols are the predictions from Equation 3.1. The dashed lines are fits to the detection data and the solid line a fit to the predictions.

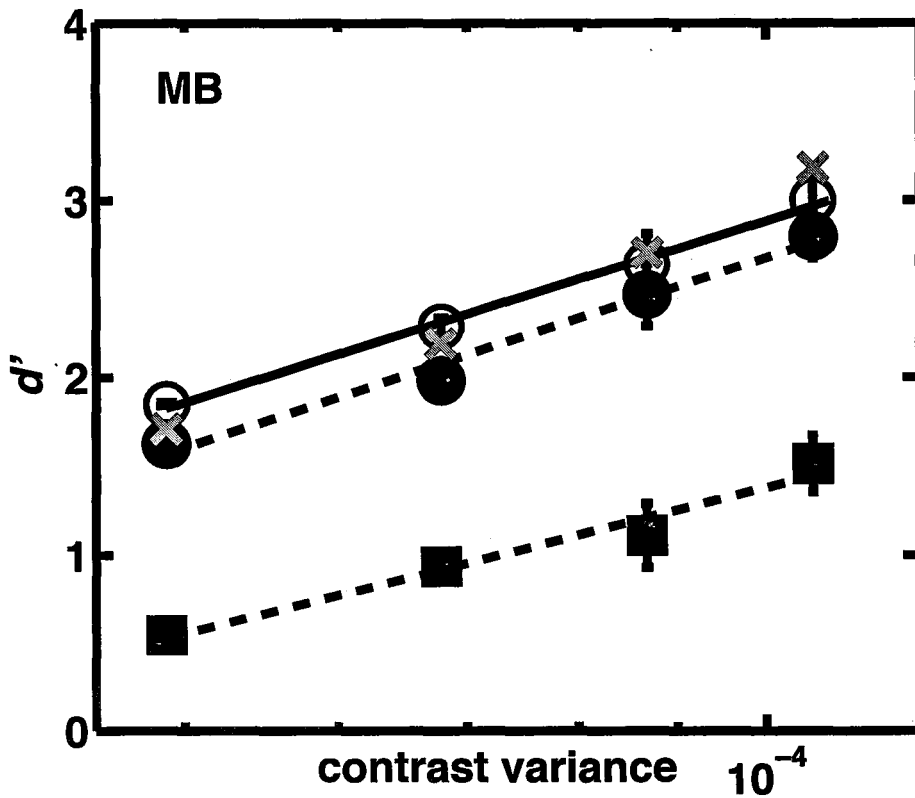


Figure 3.4: d' plotted as a function of contrast variance for observer MB in the double-judgement task. The error-bars represent 95% confidence intervals. The squares are thresholds for the 5 cy/deg detection condition, the filled circles thresholds for the 15 cy/deg detection condition, the open circles the observer's discrimination data, and the X symbols are the predictions from Equation 3.1. The dashed lines are fits to the detection data and the solid line a fit to the predictions.

$$d'_{pd} = \sqrt{d_a^2 + d_b^2} \quad (3.1)$$

where d_a and d_b are the detection d' for the two signals and d'_{pd} the predicted discrimination d' . The predicted discrimination thresholds were generated from the fits to the detection data. The detection and discrimination data were well fit by the Equation 3.1. Table 3.1 shows R^2 measures for the goodness of fit for the detection task as well as R^2 for the predictions generated by the signal detection model. The measures of goodness of fit measures show that a large portion of the variability in each observers' performance is captured by the signal detection model.

Table 3.1: R^2 for the fits to the detection data and the discrimination model.

Observer	5 cy/deg	15 cy/deg	Discrimination
AP	0.89	0.96	0.94
MB	0.98	0.94	0.98

3.1.3 Discussion

Recall that the results of classification image experiment discussed in Section 2.4, suggested that discrimination ought to be difficult because of the similarity of the classification images for the detection. The results of this experiment show that observers can not only discriminate 5 and 15 cy/deg patterns at or near detection threshold but they do so in a manner consistent with the labeled-line model proposed by Watson and Robson [21]. The model proposes that widely separated channels are labeled-lines – that once a stimulus is detected the channel also represents the spatial frequency of the stimulus, allowing discrimination to be predicted from detection.

3.1.3.1 The Wilson-Gelb Model

In Chapter 2 we used the Wilson and Gelb [23] model to successfully capture the results of both our threshold versus bandwidth (TvB) data and classification images. This subsection will describe the Wilson and Gelb model and its application to the Experiment 3.1 for both detection and discrimination.

The first model step applies a set of six weighted spatial frequency channels to the stimulus presented in each interval of the 2-IFC task. Figure 3.5 shows each of the six individual spatial-frequency channels and relative weights assigned to channel. The weights

were used for the modeling in this chapter were identical to one set of weights used in Chapter 2. The channel weights are based on the CSF measured for one observer (AMC) with a narrow-band (0.5 octaves wide) noise at a range of center spatial frequencies. The model then computes the response of the ensemble of channels. The channel responses are then summed via Minkowski summation. The decision rule for detecting both the 5 and 15 cy/deg patterns used by the model was to select the interval with the larger response after the Minkowski summation. The Wilson and Gelb detection model used here differs from that used in Chapter 2 because it includes a non-linear contrast gain control for each channel. The non-linearity adjusts the channel output such that the responses of the channels are adjusted to give the same output at detection threshold. Adding this model step was necessary as it was found that the model without the non-linearity added failed to capture observers' behavior at the higher contrast levels used in the current experiment. It should be noted, however, that adding the non-linearity has no significant effect on the model's responses in the conditions used in Chapter 2.

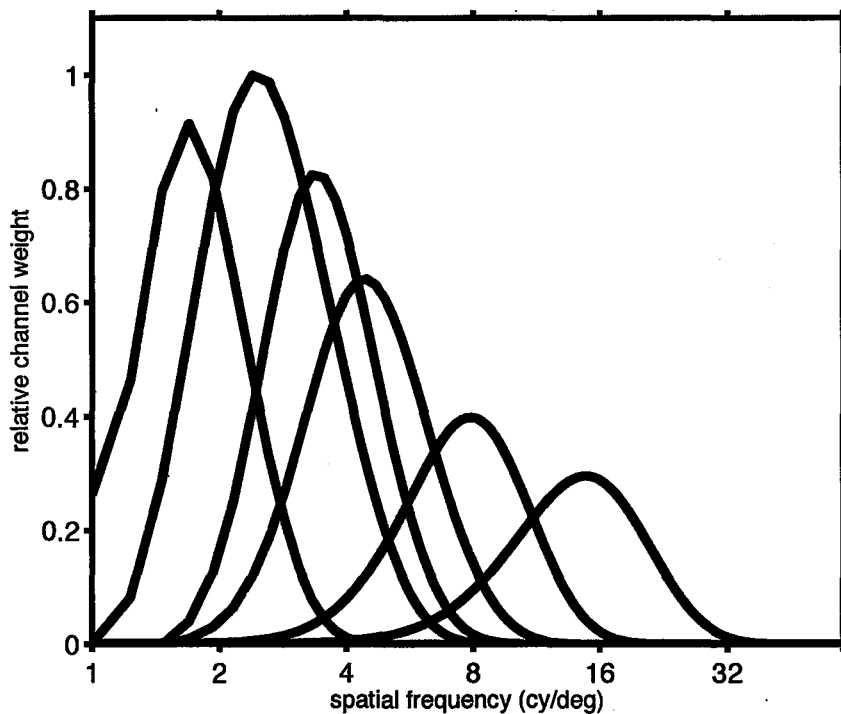


Figure 3.5: Relative channel weights for the Wilson and Gelb model measured as described in Chapter 2. The weights are from the CSF of observer AMC.

Implementing the Wilson and Gelb model for discrimination requires that the detection model be modified so that decisions can be based on the distribution of responses

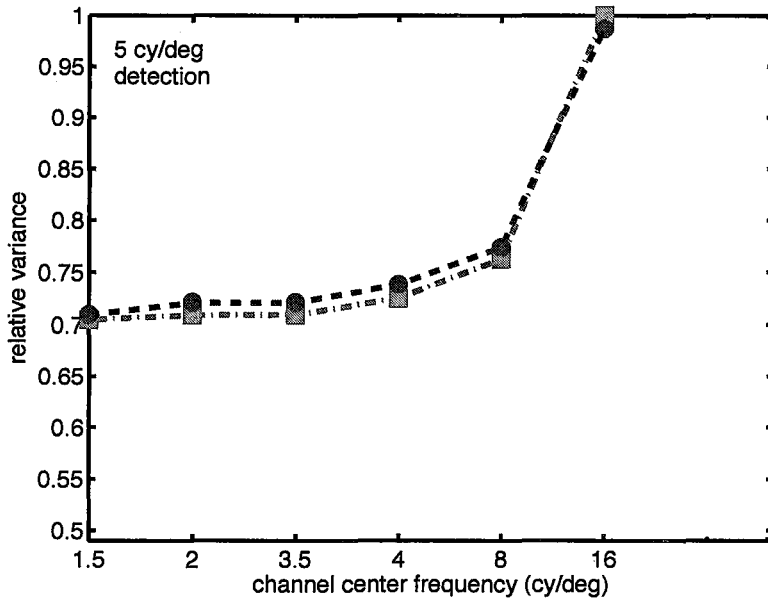


Figure 3.6: Channel variances for the six channels of the Wilson and Gelb model measured in a filtered noise detection task. The filtered noise was centered at 5 cy/deg with a one octave bandwidth set at a contrast equal to the contrast human observer MB required to obtain a d' of one. The points on the blue line represent the relative variance of each channel when responding to a filtered stimulus plus white noise. The points on the red line represents the variance of the response to the noise alone. After removing the mean response of each channel, the variance of the channels response were normalized by the maximum variance to determine their relative variance.

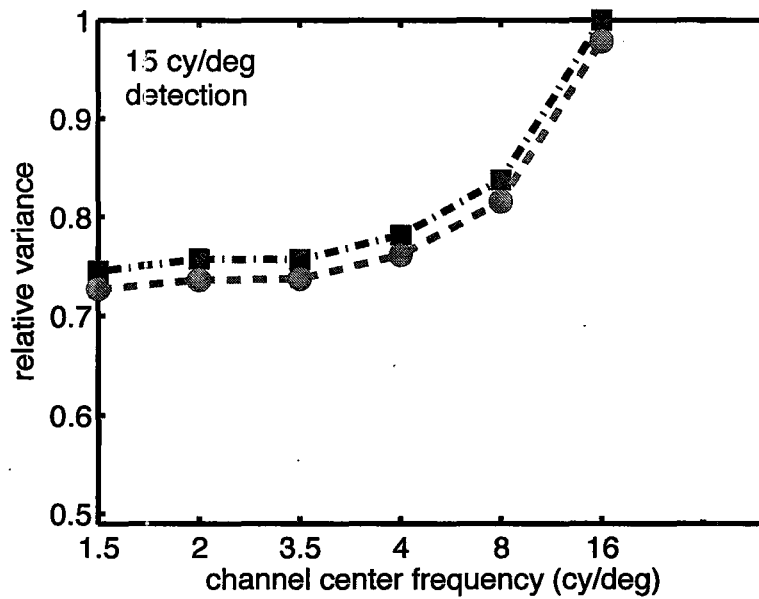


Figure 3.7: Channel variances for the six channels of the Wilson and Gelb model in for the detection of filtered noise in white noise. The stimulus was centered at 15 cy/deg with a bandwidth of one octave set at a contrast equal to the contrast human observer MB required to obtain a d' of one. Red and blue points represent conditions as in Figure 3.5

across channels, rather than a single decision variable formed by pooling responses across channels. In other words, the discrimination model must have access to the information carried by each channel and a rule that specifies how the set of responses should be used to determine a response. One possible discrimination rule is to use the output of the channel with a center-frequency nearest the center-frequency of the target stimulus. However, this rule is not optimal because it ignores information conveyed by the other channels. The stimuli used in this experiment stimulate multiple channels, and on some trials a channel other than the channel centered on the stimulus may contain information for discriminating the 5 from the 15 cy/deg patterns. That the channels differ in their sensitivities compounds the problem, as off-center channels carry information. For example, an off-center channel may respond much better than a less-sensitive channel closer to the center-frequency of the stimulus and carry more information to perform the discrimination. To address this problem, we applied a method that combines information across channels.

To determine each channel's response variability and consequently how to best combine information across channels, the individual channel detection responses were recorded. Figure 3.6 shows the relative variance of each of the six channel responses when the model detected a 5 cy/deg filtered noise. In both the signal-plus-noise interval (blue symbols) and the noise-alone interval (red symbols), the response variance of the channel with highest center spatial-frequency (16 cy/deg) was substantially greater than when the same channel was responding to noise alone. For the detection of both 5 and 15 cy/deg center frequency noise, relative variance increased as the channel center frequency increased. Comparing Figures 3.6 and 3.7 shows that the normalized variance in the 15 cy/deg condition is slightly higher overall than in the 5 cy/deg condition. The relative response variance is highest for the 16 cy/deg channel, which has the largest linear spatial frequency bandwidth.

Knowing the relative variances for each channel's output, a method for combining the information across the set of channels can be developed and a rule for discrimination implemented. Our model assumed that discrimination judgements were based on stimulus information contained in the *signal present* interval (i.e., the stimulus interval that was selected in the detection task). The set of channel responses in the signal present interval were used to calculate two quantities:

$$D_5 = \sqrt{\sum_{i=1}^6 \frac{(r_i - T5_i)^2}{\sigma_i^2}} \quad (3.2)$$

$$D_{15} = \sqrt{\sum_{i=1}^6 \frac{(r_i - T15_i)^2}{\sigma_i^2}} \quad (3.3)$$

The variable r_i represents the response of channel i in the signal present interval. $T5_i$ and $T15_i$ represent the expected (i.e., average) response of channel i if the 5 or 15 cy/deg pattern was presented, and σ_i^2 represent the variances of each channel computed from the blue curves in Figures 3.6 and 3.7. D_5 and D_{15} are the normalized Euclidean distances between the average and observed responses elicited by the 5 and 15 cy/deg stimuli, respectively. The discrimination response corresponded to the pattern that yielded the smaller distance.

Figure 3.8 depicts the sensitivity, in d' , for both the detection and discrimination models. The filled circles and squares represent d' for the detection 5 and 15 cy/deg detection task and the open circles d' for the discrimination task. The data were analyzed in the same way that the human observers' data were analyzed. A least-squares fit to the the detection data was computed for both the 5 and 15 cy/deg detection data (dashed lines). The solid line is the prediction for the discrimination data computed using the values from the least-square fits to the detection data as input to Equation 3.1. The prediction made by Equation 3.1 provides a good fit to the d' values of the discrimination model. That the model's discrimination data follow the prediction of Equation 3.1 means the model is consistent with the labeled-line hypothesis of Watson and Robson [21]. In summary, when the model can detect the signal, discrimination can be predicted from the model's detection performance for the two patterns.

Overall, for both the detection and discrimination tasks, the model requires less stimulus contrast variance to match the sensitivity (d') of the human observers. Figure 3.9 shows how much less contrast variance is required, by plotting the model efficiency for the two observers. Model efficiency was calculated by determining the contrast each of the observers required to obtain a $d' = 1$. Human observers were not as good as the model observer and ranged from one-quarter to one-half the model's performance. Consistent with the data presented in Chapter 2, observers performed best when the stimulus had a center-frequency of 5 cy/deg. That the model outperforms the human observers is ex-

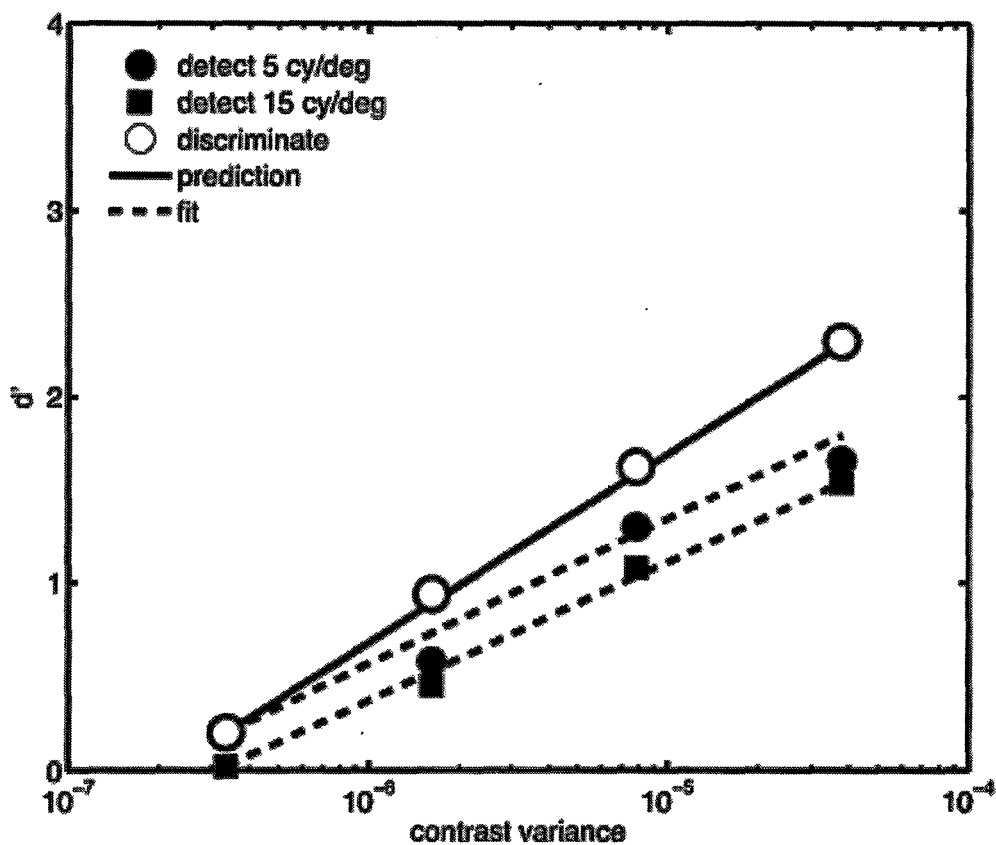


Figure 3.8: d' plotted as a function of contrast variance for the Wilson and Gelb model for the double-judgement task. Dashed lines are fits to the data in both detection tasks (filled squares and circles), the model was also made to produce discrimination data (open circles) the solid line is the prediction from Eq. 3.1.

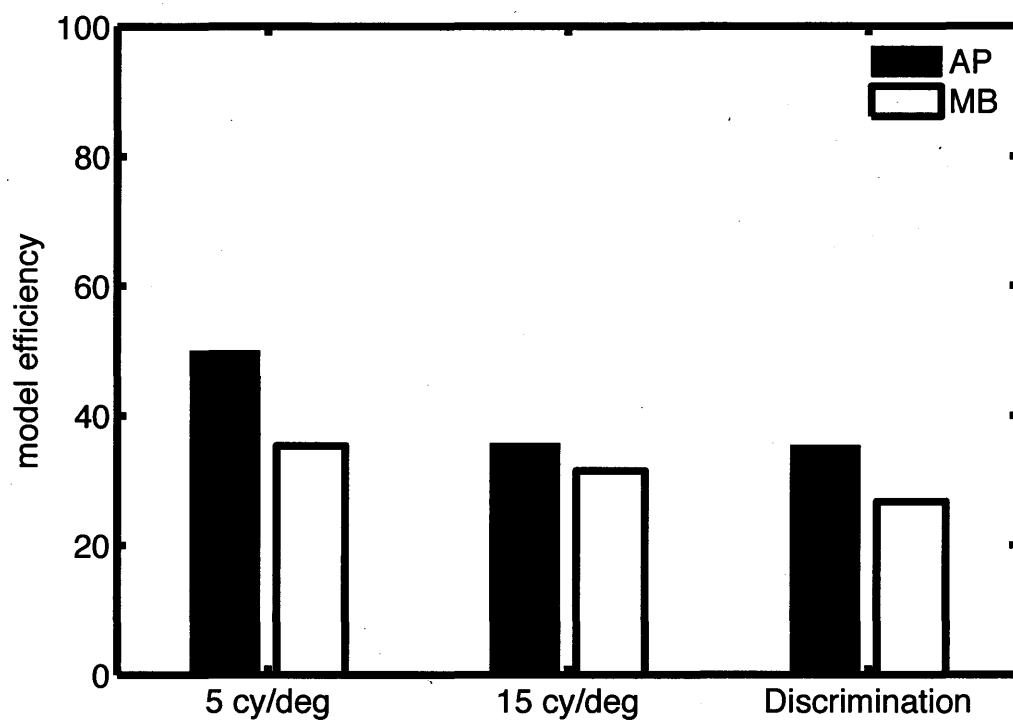


Figure 3.9: Model efficiency for observers MB and AP. Model efficiency was calculated as the ratio of contrast thresholds at $d' = 1$ for the model and human observers.

pected, as the model does not include any information losses, such as internal noise, that would be present in human observers. A second factor that could be contributing to the inability of the model to capture observers' behavior is that the weights used to set the sensitivity of the channels in the model were not set based on the individual observers' contrast sensitivity functions. The weights used were from a CSF collected in Chapter 2, perhaps some of the difference between model and human observer performance could be explained by the weights used in the model over-estimating each observers' sensitivity.

This experiment demonstrates that human observers can discriminate noise centered at 5 and 15 cy/deg, contrary to the suggestion of the classification images in Chapter 2. However, the question remains given the channel overlap in the 5 and 15 cy/deg detection conditions, what information are observers using to perform the discrimination. In the next experiment, we use the classification image technique to determine what information observers use to perform the discrimination task.

3.2 Experiment 3.2

Chapter 2 shows that measuring classification image data can substantially alter the interpretation of the results of an experiment that only measures observers' contrast thresholds. Thus, given the results of Experiment 3.1, we used the classification image technique [2] to measure observers' templates in detection and discrimination tasks. The classification image technique has been used to explore the visual templates that observers use to detect and discriminate a variety of stimuli in a wide range of tasks [2, 4, 8, 9].

3.2.1 Methods

3.2.1.1 Observers

The two observers who participated in Experiment 3.2 were the same observers who participated in Experiment 3.1.

3.2.1.2 Apparatus

The apparatus was identical to that used in Experiment 3.1.

3.2.1.3 Stimuli

As in Experiment 3.1, the stimuli in Experiment 3.2 were filtered, one-dimensional Gaussian white noise that varied in center spatial frequency. The center-frequency of the patterns was either 5 or 15 cy/deg and was blocked randomized for the detection tasks. The spatial frequency bandwidth, the Gaussian envelope, temporal duration, and the background noise were identical to Experiment 3.1.

3.2.1.4 Procedures

The algorithm used to calculate a classification image depends on the psychophysical procedure used [1] and the proper method for computing classification images in double-judgment tasks has not yet been determined. Therefore, in the current experiment, detection and discrimination judgments were performed in separate blocks during a single session. During a single session, observers performed two blocks of 2-IFC detection tasks. In one block they detected patterns with a center-frequency of 5 cy/deg and 15 cy/deg in the other. The block order was randomized. After the two detection blocks, observers discriminated 5 and 15 cy/deg stimuli. The task was 2-IFC procedure where a 5 and 15 cy/deg patterns were presented in random order on each trial and observers were required to select the interval that contained the 5 cy/deg pattern.

For each detection block the contrasts of the the 5 and 15 cy/deg patterns were set by a QUEST procedure [20]. This determined the level of contrast that each observer required on that day to obtain 71% correct which corresponds to a d' of approximately one. The QUEST procedure was used for the first 150 trials of each detection block, on the 151st trial, the contrast of the stimulus was fixed at the observer's contrast threshold for the remaining 500 trials in that block. For the discrimination block, the two stimuli were each set at the observer's detection threshold and a $\pm 5\%$ randomization factor was added to the contrast of each pattern independently on each trial to: i) allow for error in the estimate of contrast threshold; and ii) to make it difficult for observers to base discrimination on the perceived contrast of the stimuli. A session consisted of 1800 trials in total which took approximately 1.5 h to complete. Ten sessions were completed over approximately two weeks.

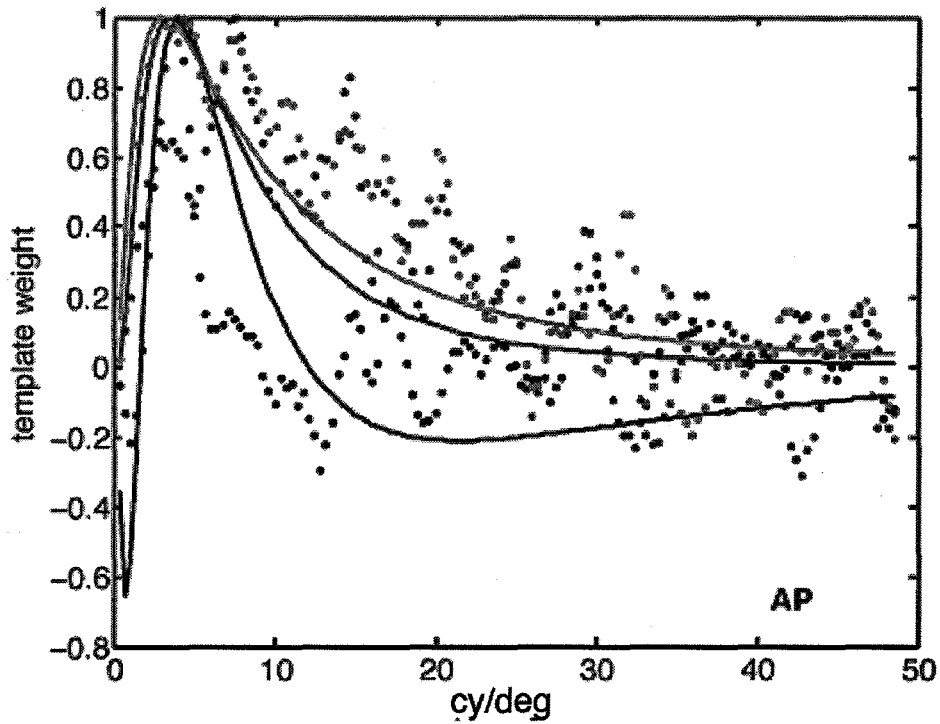


Figure 3.10: Classification images for observer AP for the detection of 5 cy/deg center-frequency noises (blue), 15 cy/deg noises (red) and discrimination (black). The points are the computed classification images. The solid curves represent the simultaneous fitting of three functions to the data (see text for details).

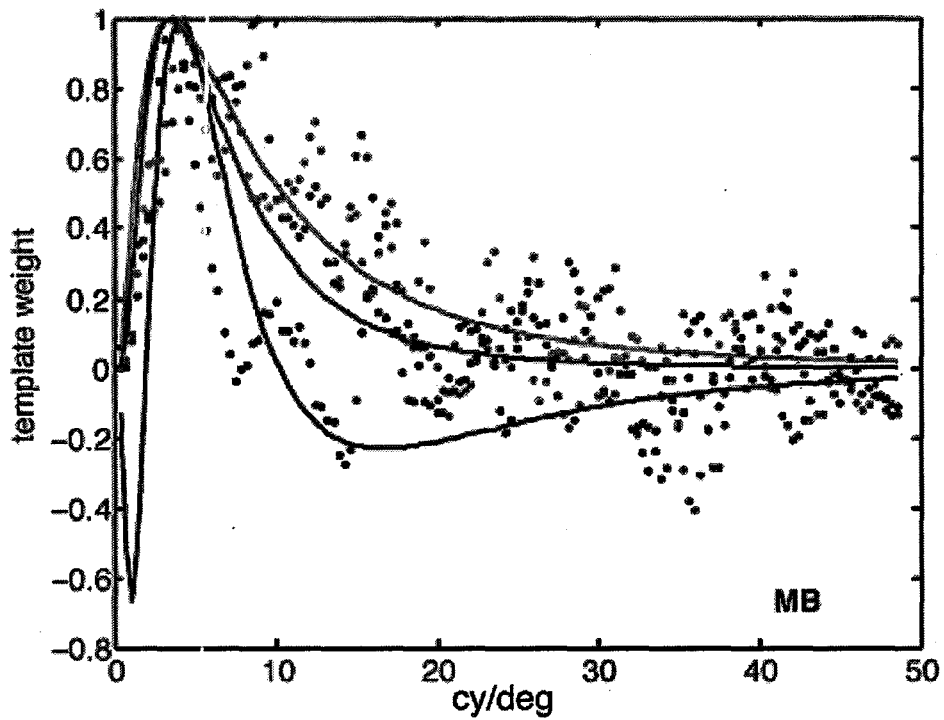


Figure 3.11: Classification images for observer MB for the detection (blue: 5 cy/deg condition; red: 15 cy/deg condition) and discrimination (black) tasks. The points are the computed classification images. The solid curves represent the simultaneous fitting of three functions to the data (see text for details).

3.2.2 Results

Fits to the classification images for the detection task were computed using the log-normal function:

$$Power(x|\mu, \sigma) = \frac{1}{\log(\sigma)\sqrt{2\pi}} e^{\frac{-(\log(x)-\mu)^2}{2\log(\sigma)^2}} \quad (3.4)$$

where μ is the center-frequency of the channel and σ the bandwidth. Classification images from the discrimination task were fit with a function that was the difference of the two lognormal functions fit to the detection data. It is important to note that the fitting procedures for the two detection tasks and discrimination task were not performed separately. Instead, the fits were calculated for all three channels simultaneously by adjusting the two center-frequencies of the detection channels ($\mu_{5cy/deg}$ and $\mu_{15cy/deg}$) and their two bandwidth parameters ($\sigma_{5cy/deg}$ and $\sigma_{15cy/deg}$).

Table 3.2 shows the best fitting center frequencies and octave bandwidths for the log normals fit to the classification images measured in the detection task. The values reported here replicate those we have reported previously (see Tables 2.3 and 2.4). Again, one key feature of the data stands out: the best fitting center-frequencies of the classification images in 5 and 15 cy/deg are the same (approximately 4 cy/deg), while the bandwidths of the classification images in the two conditions differ by ≈ 0.1 log units.

Unlike the data from the detection task, the discrimination data were not fit with Equation 3.4 but, instead, were described by the difference between the two curves fit to the detection data. The motivation for using a difference was as follows: if the two templates measured in the detection task are also being used for the discrimination task, then the spatial frequencies where the two detection templates differ maximally ought to be the frequencies where the most information is available to the observer to complete the detection discrimination task.

Table 3.2: Center-frequency, bandwidth and 95% confidence intervals, in cy/deg, for two observers in the detection tasks.

	stimulus (cy/deg)	AP	MB
center frequency	5	4.25 ± 1.02	3.91 ± 01.22
center frequency	15	4.15 ± 0.67	4.27 ± 1.06
bandwidth	5	13.5 ± 1.1	13.8 ± 1.7
bandwidth	15	28.8 ± 1.32	27.5 ± 1.4

The Median Absolute Deviation (*MAD*) provides a descriptive statistic that is robust to outliers for deviation of the observed classification images from the fitted classification image. For the detection tasks, the *MAD* for the residuals of our fitted normalized templates was less than 0.04 units of normalized power for both observers with both stimuli. For the discrimination task, *MAD* increased to 0.12 units for observer AP and 0.11 units for observer MB. To evaluate the fits, Hotelling's T^2 was calculated. This method was developed for the classification image technique by Abbey and Eckstein [1], the statistic tests whether the differences between the observed and predicted a classification images are significant. We used the fitted lognormal function for each observer, in each of three tasks, as the known profile. For both observers, in all tasks, Hotelling's T^2 statistic failed to find significant differences between the classification image and the lognormal profile: for the 5 cy/deg detection task $T^2 = 0.397(p = 0.527)$, for 15 cy/deg $T^2 = 0.288(p = 0.591)$ and $T^2 = 1.06(p = 0.303)$.

Table 3.3: Observer R^2 for two observers in the detection (5 and 15 cy/deg) and discrimination tasks.

	5 cy/deg	15 cy/deg	Discrimination
AP	0.757	0.750	0.647
MB	0.825	0.816	0.615

R^2 was also calculated for the two observers in the two tasks, and the values are shown in Table 3.3. In all cases, significant correlations ($p < 0.0001$) were observed. Each of the three statistics computed indicates that the lognormal channels fitted to the data provided a very good description of the shape of the template for noise detection and discrimination. While the fits are good, there are clear differences between the fit to the discrimination classification images and those produced by human observers. For example, the fit when the template weights are positive seems too broad. Also, for observer AP, the fit to the negative weights seems to be consistently under-predicted.

3.2.3 Discussion

The detection classification images obtained in the current experiment replicated those measured in Experiment 2.4. As before, the classification images had peaks that, surprisingly, were centered at ≈ 4 cy/deg when observers detected patterns centered at 5 and 15 cy/deg. We also replicated the difference in bandwidth between the classification images measured for the two classes of patterns. From these data we created a model that

used the difference between the two detection classification to compute a template that provided a reasonably good fit to the classification image measured in the discrimination condition.

Surprisingly, the difference in the bandwidths of the classification images in the 5 and 15 cy/deg conditions are enough to explain the discrimination of the two classes of patterns. This point can be seen in the classification images in Figures 3.10 and 3.11. The reason this is surprising is that the fits of the two detection tasks were computed separately for each observer. After fitting, the difference between the detection channels was then computed to produce a third curve for discrimination that was then correlated with the actual discrimination classification. The analysis is computed in a manner in which no additional free parameters, other than the fits to the detection data, are used to explain the results of the discrimination classification images.

The classification images for discrimination that we measured are inconsistent with a model that assumes that the discrimination of 5 and 15 cy/deg patterns are discriminated by two channels that differ in their center-frequency at 5 and 15 cy/deg. The ideal model would predict that the templates measured for detection ought to be centered on 5 and 15 cy/deg and have a bandwidth of one octave. The ideal model would also use the difference of the output of the two channels but the predicted channel from the detection channels would be very different.

The approach to modeling the discrimination data is a simple one and it could be argued that it is too simple to capture the difference between detection and discrimination. The goal was not to find the best fitting model, but to explore how much of the variance in the discrimination classification image could be captured by a very simple model. This does not rule out that a more complex discrimination rule is being used, however it does explain a surprising degree of the variance in the discrimination classification image. A natural question to ask at this point is if the difference between the two classification images for detection at 5 and 15 cy/deg center-frequencies can predict observers' d' in the tasks used in Experiments 3.1 and 3.2. The following subsection describes two simulations that explore whether a differencing model can predict observers' d' .

3.2.3.1 Do the classification images predict the d' of human observers?

To address whether the templates measured in the detection and discrimination task could predict each observers' d' , we performed simulations to determine whether d' could be predicted from the classification images shown in Figures 3.10 and 3.11.

Table 3.4: Observer and model d' for the 5 and 15 cy/deg detection and discrimination. Also included is the discrimination d' from Equation 3.1

	Detect 5 cy/deg	Detect 15 cy/deg	Discrimination	SDT Prediction
Model AP	0.60	0.46	1.6	0.91
Human AP	0.68	1.01	1.3	1.22
Model MB	0.5	0.55	1.7	0.75
Human MB	0.7	1.2	1.18	1.34

The procedures used to evaluate the classification image model d' were identical to that used with the human observers in this experiment. To generate a response in the detection task, the correlation of the classification images for the appropriate condition (5 or 15 cy/deg) and each interval in the 2-IFC task. The interval that produced the larger correlation was selected as the interval within which the signal appeared. In the discrimination task, the model used the observers' discrimination classification image to predict d' .

This model is important for the two following reasons. Firstly, the task was changed from Experiments 3.1 to 3.2 to simplify the collection of classification images, therefore a comparison of sensitivity between the blocked procedure used in this experiment and the double-judgement procedure used in the first experiment. Secondly, the fits of the discrimination classification image were relatively poor compared to those measured in the two detection tasks. The poor fits could indicate that the fitting procedure is not capturing important features of the classification image, specifically those features at high spatial frequencies or non-linear influences [15].

The d' for the template model and two human observers are shown in Table 3.4. The template model succeeds and fails in two important ways, the model captures d' for human observers for the detection of 5 cy/deg patterns and the discrimination of 5 from 15 cy/deg patterns, however, the model fails to predict d' for the detection of 15 cy/deg patterns. The failure of the model to capture the performance of human observers means that observers are using the information in a manner that is not represented in their

classification images. Murray, Bennett, and Sekuler [15] developed a method that could predict observer performance from their classification images, they found that classification consistently over-predicted observer performance, after investigating a number of non-linearities they concluded only phase-uncertainty explained the under-prediction of the classification images. The under-prediction in the classification images in this experiment is consistent with a number of studies that used grating and Gaussian blob stimuli. These studies found that classification images were often absent, despite observers being able to perform the task, when operating under the effects of uncertainty [3, 18, 19]. Perhaps, the under performance of the template model in our task for 15 cy/deg stimuli is the result of the effect of uncertainty at medium to high spatial frequencies.

3.2.4 The Wilson-Gelb Model and Classification Images

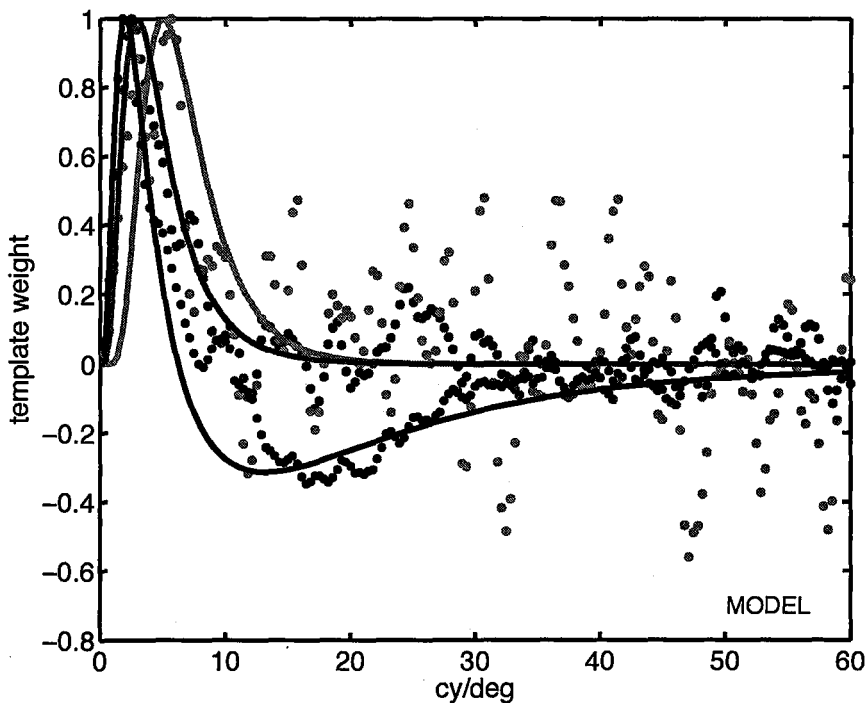


Figure 3.12: Classification images for the Wilson-Gelb model for the detection of 5 cy/deg center-frequency noises (blue), 15 cy/deg noises (red) and discrimination (black). The points are the computed classification images. The solid curves represent the fits of three functions to the data (see text for details).

In Chapter 2 the Wilson and Gelb [23] model was used to produce classification images. The Wilson and Gelb can produce classification images for both the 5 and

15 cy/deg detection conditions and the discrimination task. The model used in this experiment is described in Section 3.1.3.1.

Figure 3.12 shows the classification images produced by the Wilson and Gelb model. Figures 3.11 and 3.10 to Figure 3.12 show that the model produces qualitatively similar classification images to human observers. However there are differences: i) the detection classification images produced by the model have narrower bandwidths; ii) the peak frequencies in the 5 and 15 cy/deg conditions differ more than human observers and; iii) the negative weights in the model's discrimination classification image are approximately double the values obtained from human observers. It is possible that adjusting the parameters of the model could reduce these differences. However, the goal of applying the Wilson and Gelb model was not to fit human data precisely, but to see whether a simple model, with few free parameters could capture the major features of the classification images. In this context, the similarities between the classification images obtained from the model and human observers are, perhaps, more important than the differences.

This model directly addresses the puzzling result that motivated the work presented in this chapter – that the similarities in the detection classification images imply that discrimination of noise with center-frequencies of 5 and 15 cy/deg ought to be difficult to discriminate. The model produces detection classification images that are peaked around 5 cy/deg with bandwidth of approximately one and a half octaves which reproduces the puzzling result. However, the model also discriminates 5 and 15 cy/deg patterns and uses a template that is similar to that used by human observers.

3.3 Summary

Experiment 3.1 obtained results in a double-judgment task that are consistent with a labeled-line model of detection/discrimination [21]. The Wilson and Gelb model was applied to Experiment 3.1 and provided a good account of human observers' d' in both the detection and discrimination portions of the double-judgment task.

Experiment 3.2 measured classification images for the detection of 5 and 15 cy/deg stimuli. As seen in Chapter 2, the channels for the detection of 5 and 15 cy/deg patterns overlap considerably – the result that motivated the experiments in this chapter. The classification images for discrimination show that using the channels that we found in Chapter 2 could be fit with the difference of the classification images measured for detection.

Two models were applied to the data from Experiment 3.2. The first model, used observers' templates to attempt to predict d' in the detection and discrimination tasks. Human observers d' in the 5 cy/deg detection and discrimination conditions was explained by the template model, but the model failed on the 15 cy/deg task indicating that the templates fail to entirely capture how observers are using information in the 15 cy/deg detection task. The second model, the Wilson and Gelb model was used to produce classification images for detection and discrimination. The templates of the Wilson and Gelb model were qualitatively similar to those of human observers for both detection and discrimination, indicating that this simple model can capture the pattern of results shown in human templates.

3.4 Conclusion

The experiments in this chapter addressed the puzzling result from Chapter 2, that the classification images measured for noise filtered to have a center frequency of 5 and 15 cy/deg seem quite similar. The result was puzzling because spatial frequencies of 5 and 15 cy/deg are very discriminable. In this chapter we tested whether the thresholds and classification images we measured for detection could be related to discrimination. Surprisingly, we found that the small differences between 5 and 15 cy/deg classification images were enough to predict discrimination classification images.

References

- [1] C.K. Abbey and M.P. Eckstein. Classification image analysis: estimation and statistical inference for two-alternative forced-choice experiments. Journal of Vision, 2(1):66–78, Jan 2002.
- [2] A. Ahumada and J. Lovell. Stimulus features in signal detection. Journal of the Acoustical Society of America, 49:1751–1756, Jan 1971.
- [3] A.J. Ahumada. Classification image weights and internal noise level estimation. Journal of Vision, 2(1), 2002.
- [4] B.L. Beard and A.J. Ahumada. Detection in fixed and random noise in foveal and parafoveal vision explained by template learning. Journal of the Optical Society of America A, 16(3):66–78, Jan 1999.
- [5] D. Brainard. The psychophysics toolbox. Spatial Vision, 10:443–446, Jan 1997.

- [6] C. Chubb and M.S. Landy. Orthogonal distribution analysis: A new approach to the study of texture perception. In M.S. Landy and J.A. Movshon, editors, Computational Models of Visual Processing, pages 291–301, Cambridge, MA, 1991. MIT Press.
- [7] R.L. DeValois and K.K. DeValois. Spatial Vision. Oxford University Press, New York, 1988.
- [8] M.P. Eckstein, S.S. Shimozaki, and C.K. Abbey. The footprints of visual attention in the Posner cueing paradigm revealed by classification images. Journal of Vision, 2(1):25–45, 2002.
- [9] J.M. Gold, R.F. Murray, P.J. Bennett, and A.B. Sekuler. Deriving behavioural receptive fields for visually completed contours. Current Biology, 10(11):663–6, Jun 2000.
- [10] N.V.S. Graham. Visual Pattern Analyzers. Oxford University Press, New York, 1989.
- [11] D.M. Green and J.A. Swets. Signal Detection Theory and Psychophysics. Wiley, New York, 1966.
- [12] D. Kersten. Statistical efficiency for the detection of visual noise. Vision Research, 27(6):1029–1040, Jan 1987.
- [13] S.A. Klein. Double-judgment psychophysics: problems and solutions. Journal of the Optical Society of America A, 2(9):1560–85, Sep 1985.
- [14] N.A. Macmillan and C.D. Creelman. Detection Theory: A User's Guide. Cambridge University Press, Cambridge, 1991.
- [15] R.F. Murray, P.J. Bennett, and A.B. Sekuler. Classification images predict absolute efficiency. Journal of Vision, 5(2):139–149, 2 2005.
- [16] J. Nachmias and A. Weber. Discrimination of simple and complex gratings. Vision Research, 15:217–223, Jan 1975.
- [17] D.G. Pelli. The videotoolbox software for visual psychophysics: transforming numbers into movies. Spatial Vision, 10(4):437–442, Jan 1997.
- [18] J.A. Solomon. Noise reveals visual mechanisms of detection and discrimination. Journal of Vision, 2(1), 2002.

- [19] B.S. Tjan and A.S. Nandy. Classification images with uncertainty. Journal of Vision, 6(4), 2006.
- [20] A.B. Watson and D.G. Pelli. Quest: A Bayesian adaptive psychometric method. Perception and Psychophysics, 33(2):113–120, 1983.
- [21] A.B. Watson and J.G. Robson. Discrimination at threshold: labelled line detectors in human vision. Vision Research, 21:1115–1122, 1981.
- [22] G.H. Wetherill and H. Levitt. Sequential estimation of points on a psychometric function. British Journal of Mathematical & Statistical Psychology, 18:1–10, Jan 1965.
- [23] H.R. Wilson and D.J. Gelb. Modified line-element theory for spatial-frequency and width discrimination. Journal of the Optical Society of America A, 1(1):124–131, 1984.
- [24] H.R. Wilson and F. Wilkinson. Evolving concepts of spatial channels in vision: From independence to nonlinear interactions. Perception, 26:939–960, Jan 1997.

Chapter 4

Orientation summation in visual noise

Abstract

The results of the two previous chapters imply that human observers integrate a wide range of spatial frequencies efficiently but do so in a manner consistent with a standard model of pattern vision [33]. In this chapter, we investigate whether the results found for the summation of spatial frequency can be extended to the summation of information across orientations. Detection thresholds were measured in a background of white noise for two-dimensional filtered noise of increasing orientation bandwidth at two spatial frequency bandwidths. Detection thresholds increased with the quarter-root of stimulus bandwidth as predicted by the ideal detector. The classification image technique was applied in a manner similar to that of Chapter 2. Unlike what was found in the spatial frequency summation experiments, the classification images suggest that observers adjust the orientation bandwidth of internal channels to the orientation bandwidth of the stimulus.

Introduction

Standard models of the initial stages of visual processing include a stage in which stimuli are encoded by an array of multiple independent filters, or channels, that are tuned to relatively narrow ranges of spatial frequency and orientation [9, 30, 32, 33,

34]. Support for multiple channel models comes (in part) from spatial frequency and orientation summation studies [16]. In a typical orientation summation experiment, contrast detection thresholds are measured for two oriented component gratings of spatial frequency F and orientations θ and $\theta \pm \delta$ where δ represents the difference in orientation between the two stimuli. A compound grating comprising the two stimulus components is created, and the contrast threshold for the compound grating is measured. Summation is observed, if the detection threshold for the compound stimulus is lower than thresholds for either of the component gratings alone. Significant summation has been found only when the orientations of the two compound grating do not differ by more than 15-20° [16], suggesting that only orientations within relatively narrow range can be summed.

In this chapter, we applied an approach to orientation summation that was first used by Green to study auditory frequency summation [18, 19] and extended to spatial frequency summation by Kersten [20]. Kersten measured detection thresholds for static, one-dimensional bandpass noise stimuli presented in a background of white noise. The bandwidth of the noise was varied across conditions. Kersten found that detection thresholds, when expressed as root-mean squared (RMS) contrast, were proportional to the quarter-root of the stimulus bandwidth. Kersten also calculated the performance of an ideal detector that summed contrast power optimally across all spatial frequency components in the stimulus. Unexpectedly, the ideal detector's threshold (expressed as RMS contrast) was only 0.5 units lower than human thresholds and was proportional to the quarter-root of stimulus bandwidth. Hence, absolute efficiency, defined as

$$\eta = \left(\frac{c_{ideal}}{c_{observer}} \right)^2 \quad (4.1)$$

where c_{ideal} and $c_{observer}$ are the ideal and human observer's RMS contrast thresholds, was high ($\approx 50\%$) and nearly constant across stimulus bandwidth. To explain this result, Kersten [20] speculated that human observers used an internal channel, or filter, whose bandwidth could be adjusted to match the spatial frequency bandwidth of the noise signal. The finding that spatial frequency summation was highly efficient and nearly constant across a wide range of bandwidths suggests that the process of combining responses across channels appeared to be inconsistent with the results of spatial frequency summation studies using sine wave grating stimuli. Chapter 2 and 3 reconcile the findings of Kersten with standard models of spatial frequency summation [16].

Chapter 2 found the results of the experiments described above to be robust and replicable. However, using the conclusion that they support a model in which the vi-

sual system uses adjustable channels to sum spatial frequency proved to be premature. We measured classification images which clearly showed that adjustable channels are not necessary to produce efficient detection that is constant across stimulus bandwidth. Observers used information from a single template that did not exhibit any adjustment of bandwidth. Moreover, we showed that the detection and classification image results could be explained using a standard multiple channel model [33]. In short, the results of this chapter bring the work of Kersten in line with standard models of spatial frequency summation [16].

Spatial frequency and orientation are modeled similarly but to date measuring orientation summation with a visual noise stimulus using the methods [20] has not been done, nor has a classification image analysis been applied to orientation summation to determine what information is used when detecting orientation filtered noise. The ideal detector for orientation filtered noise is the same as that for spatial frequency when the stimulus is expressed as the number of Fourier components in the stimulus. Thus, we can use the same approach as we used previously [28] to determine whether orientation summation exhibits optimal summation and the shape of the template for the detection of orientation filtered noise. The goal was to determine whether the standard model can encompass the detection of orientation filtered noise signals as well as the narrow-band grating stimuli used to develop it.

Two experiments are described in this chapter. The first extends the findings of [20] to patterns that are band-limited in spatial frequency (two spatial frequency bandwidths, one- and two- octaves) but vary in their orientation bandwidth over a wide range. The second experiment uses the response classification technique [3, 11, 14, 26] to estimate the tuning characteristics of the internal filters used in this noise detection task.

4.1 Experiment 4.1

In this experiment we perform an experiment analogous to the one presented in Experiment 2.1, however, instead of only varying the spatial frequency bandwidth, the bandwidth of the stimuli varied in both their spatial frequency and orientation bandwidths.

4.1.1 Methods

4.1.1.1 Observers

The three observers in this experiment were members of the McMaster University community and were paid for their participation. Observers were naïve with respect to the experimental hypotheses, had normal or corrected-to-normal Snellen acuity, Pelli-Robson contrast sensitivity, and had extensive practice with this and other visual psychophysical tasks. Both observers were 27 years of age.

4.1.1.2 Apparatus

A Macintosh G4 running MATLAB and the Psychophysics and Video toolboxes [5, 23] was used to generate the stimuli. The stimuli were presented with an ATI Radeon video card on a Sony GDM-F520 monitor set to a resolution of 1024 by 768 pixels. The entire display subtended a visual angle of 10.8 by 8.3 degrees at the viewing distance of 2 m. The frame rate of the display was 75 Hz and the mean luminance 45 cd/m^2 . The luminance of the display was calibrated using a PhotoResearch PR-650 photometer before each session. A Cambridge Research System Bits++ device was used to achieve fine grained (i.e., 14-bit) control of contrast. Responses were recorded with a button box.

4.1.1.3 Stimuli

The stimuli were two-dimensional Gaussian white noise patterns that were spatially filtered with ideal spatial frequency and orientation filters. The spatial frequency filter was centered on 5 cycles per degree (cy/deg) with a bandwidth of either one or two octaves. The ideal filter applied to orientation and was centered on the horizontal orientation. In the one octave spatial frequency bandwidth condition, two-sided orientation bandwidths were 2° , 16° , 32° , 64° , 96° , 128° , and 154° . In the two octave spatial frequency bandwidth condition, the two-sided orientation bandwidths were 2° , 16° , 32° , 64° , 96° , 128° , and 154° . Bandwidth will also be expressed as the number of components in the stimulus in this chapter because while the measure of bandwidth is more intuitive, the ideal observer generates a quarter-root relationship with RMS contrast and the number of components.

Stimulus contrast was modulated with a circularly-symmetric Gaussian envelope with a standard deviation of 1.08 degrees of visual angle. Stimuli were presented in a background of white masking noise that had a variance of 0.32. A new sample of signal noise

and background noise were generated on each interval on every trial. Stimulus duration was 200 ms. The monitor provided the only illumination in the testing room.

4.1.1.4 Procedure

Observers viewed the stimuli binocularly through natural pupils. A 3×3 degree outline of a square of maximum contrast surrounded the stimuli to reduce spatial uncertainty. To reduce adaptation, the square had a 50% probability of being black or white on each trial. The square had a width of two pixels and was on the screen for the entire duration of each trial, from the presentation of the fixation point until the observer's response. A two-interval forced-choice (2-IFC) procedure was used. The observer was instructed to fixate a high-contrast dot located in the center of the display. The observer initiated each trial by pressing the spacebar on the keyboard. After a delay of 50 ms, the fixation point was removed, then after another 50 ms delay the first stimulus interval appeared. The first stimulus interval was 200 ms in duration and was followed by a 300 ms blank inter-stimulus interval. After the inter-stimulus interval a second 200 ms stimulus interval appeared. The stimulus intervals were marked by clearly-audible tones. The observer's task was to determine which of the two intervals contained the target. Auditory feedback, in the form of low- and high-pitched tones, indicated the accuracy of the response after each trial. Stimulus contrast variance was varied across trials using four interleaved staircases, two converging on the 71% correct point of the psychometric function and two on the 84% correct point [31]. The staircases were stopped when the observer had completed 75 trials in each staircase. The total number of trials run in each session was 2100, 300 trials per stimulus bandwidth, and seven stimulus bandwidths/session. Thresholds were estimated by fitting a cumulative normal to the combined data of all four staircases.

Two spatial frequency bandwidth conditions, of one- and two- octaves wide filtered noise, were run in separate sessions in reverse order for the two observers. The order of stimulus bandwidth was block randomized in each test session; the bandwidths were presented in separate blocks of trials and the order of blocks/bandwidths were randomized. Each observer completed all bandwidth conditions during a single session. Four sessions were run for each observer at each spatial frequency bandwidth.

4.1.2 White Noise Thresholds

In addition to the conditions listed, we collected contrast detection thresholds for a white noise stimulus after all the data were collected in the conditions previously

mentioned. The procedure for collecting white noise thresholds was the same as those for the filtered noise stimuli, with the one difference, the number of trials completed. The two observers completed four white noise detection sessions containing 300 trials. The thresholds from this portion of this experiment were run at the end of experiment 4.1, after all other conditions were completed.

4.1.3 Results

Figures 4.1 and 4.2 shows threshold versus bandwidth (TvB) functions for three observers in two spatial frequency conditions. Each symbol type (squares, circles, and triangles) represents the thresholds for a different observer. The filled symbols are the data from the one octave spatial frequency filtered stimuli and open for the two octave stimuli. Both Figure 4.1 and 4.2 show that as the orientation bandwidth of the stimulus increases, the RMS contrast threshold for each observer increases. There are no statistical differences between the thresholds in the one and two octave spatial frequency bandwidth conditions. The dashed lines in both Figure 4.1 and 4.2 have a slope of 0.25 and have been shifted vertically for visual comparison to the data. For all observers in both spatial frequency conditions, the slope of the TvB function appears to follow the prediction of the quarter-root-law. For more precision we computed best-fitting power functions for to each TvB function.

Although the sampling of the data points in this experiment does not make it readily apparent, TvB functions tend to be flat and then rise as the bandwidth of the stimulus is increased (see 2.1). The fitting routines therefore excluded the narrowest data point (2°). The slopes for the estimated TvB functions are shown in Table 4.1.

Table 4.1: Threshold vs. bandwidth slopes for three observers in Experiment 4.1 for the one and two octave conditions.

Noise σ^2	AP	MB	NS
1	0.24	0.24	0.26
2	0.26	0.24	0.25

Figures 4.3 and 4.4 show absolute efficiency (Eq. 4.1) as a function of orientation bandwidth for the two spatial frequency conditions. The absolute efficiencies found in this experiment are comparable to those found for spatial frequency summation in Chapter 2. Absolute efficiency for the three observers is approximately constant for

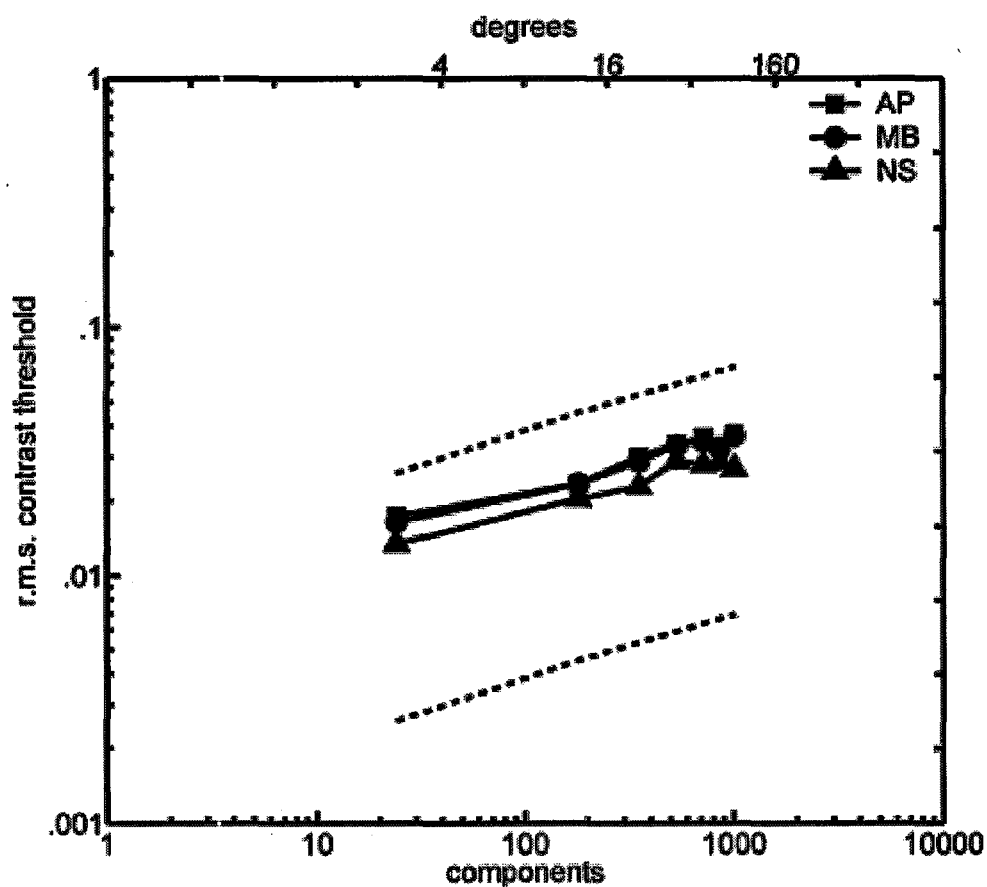


Figure 4.1: Noise detection thresholds versus stimulus bandwidth (top axis) and number of spatial frequency components (bottom axis) for three observers in Experiment 4.1 in the one octave spatial frequency condition. The different symbols represent thresholds for three separate observers. The dashed lines have a slope of 0.25 and are provided as a guide; they have been shifted vertically for clarity.

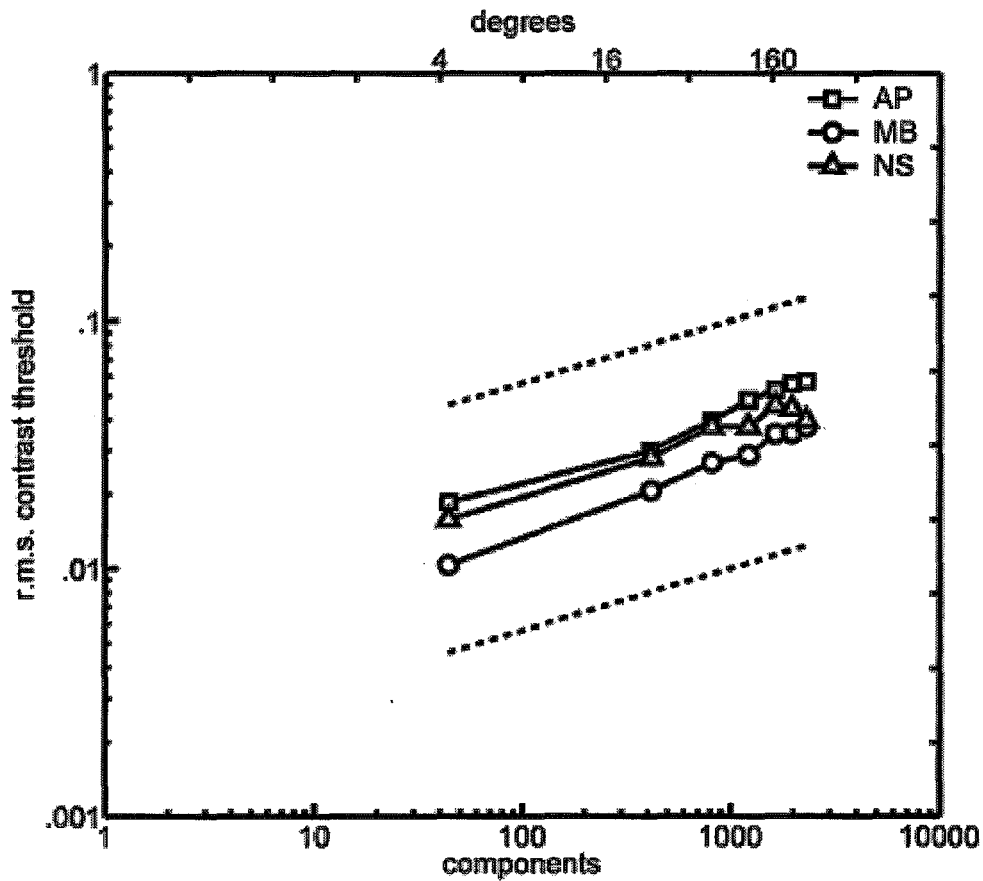


Figure 4.2: Noise detection threshold versus stimulus bandwidth (top axis) and number of spatial frequency components (bottom axis) for three observers in Experiment 4.1 in the two octave spatial frequency condition. The different symbols represent thresholds for three separate observers. The dashed lines have a slope of 0.25 and provided as guide and have been shifted vertically for clarity.

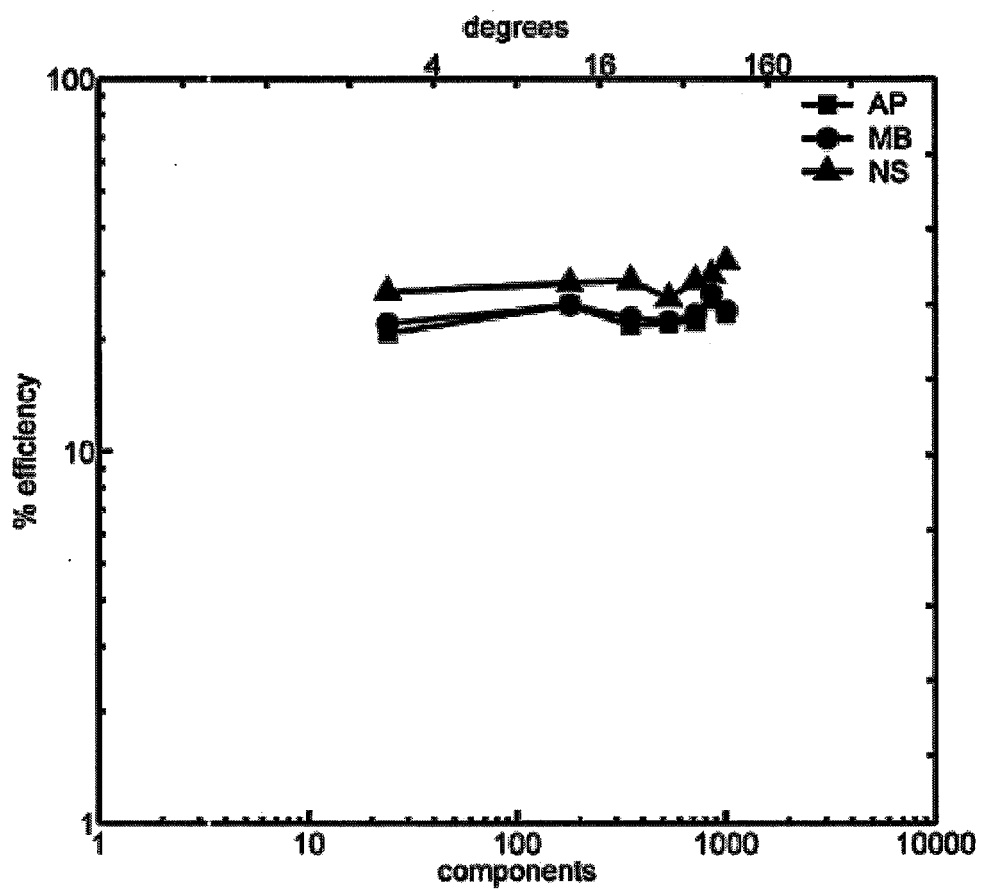


Figure 4.3: Absolute efficiency for three observers in the one octave spatial frequency bandwidth condition.

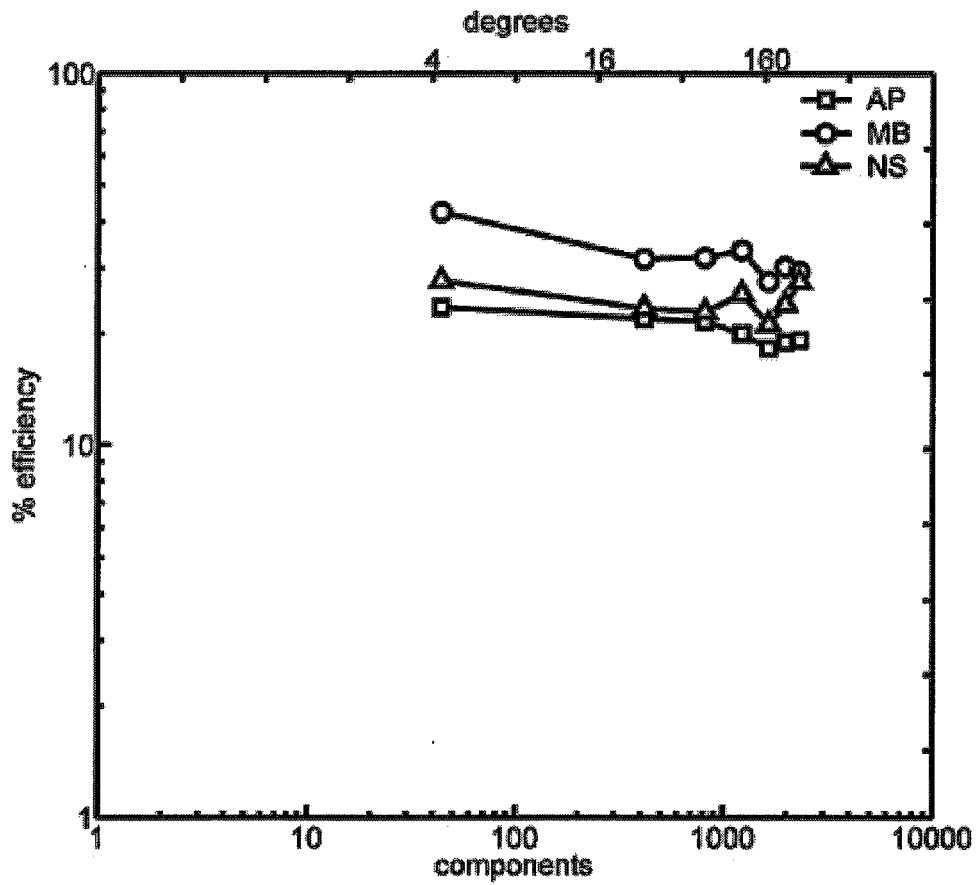


Figure 4.4: Absolute efficiency for three observers in the two octave spatial frequency bandwidth condition.

both the one and two octave spatial frequency conditions. The absolute efficiencies found in this experiment are lower than for spatial frequency summation. For spatial frequency summation efficiencies as high as 50% have been found, absolute efficiency for our observers in this experiment are also relatively high, ranging from 20% to 40%. The efficiency data exhibit another pattern, the absolute efficiencies found are not significantly different for the one- and two-octave spatial frequency conditions. This result indicates that the human visual system can optimally sum a wide band of orientations for spatial frequency bandwidths of both one- and two-octaves.

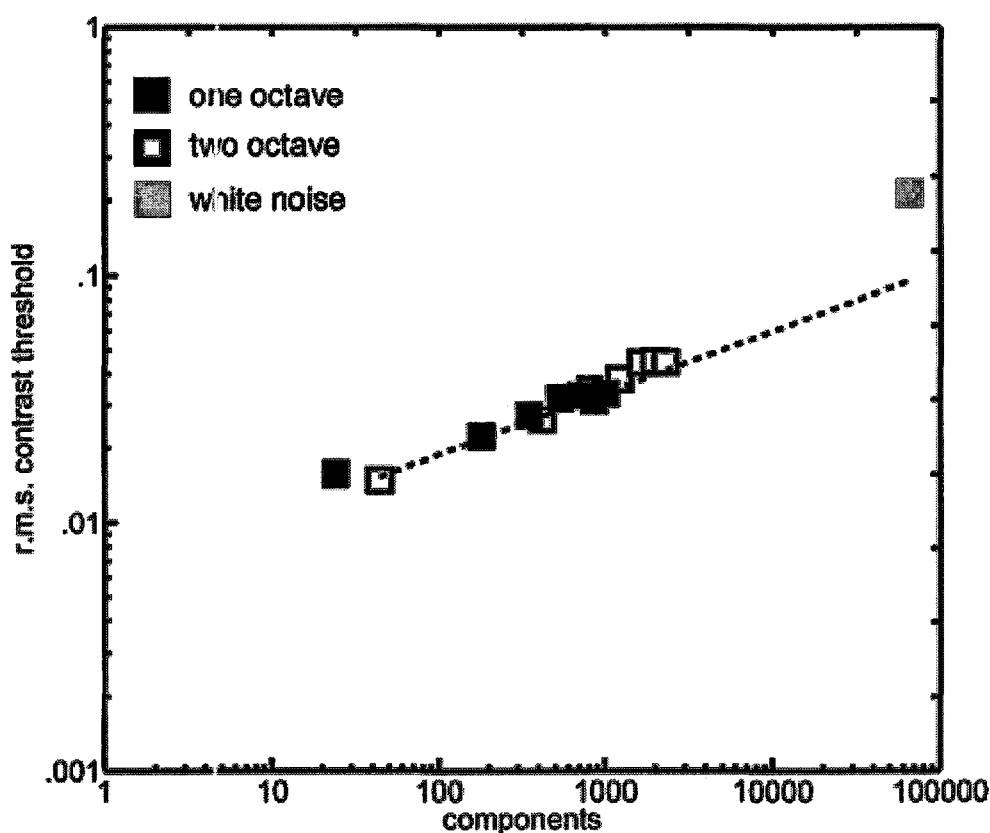


Figure 4.5: Threshold versus bandwidth data re-plotted from Figures 4.2 and 4.3. The grey square represents the observers' threshold for detecting white noise. The dashed line is the best fitting power function to the data excluding the threshold measured with white noise.

Figure 4.5 is a summary figure of the data of Figures 4.1 and 4.2. The filled squares represent the mean RMS contrast thresholds from the three observers in the one octave spatial frequency condition and the open squares the means for the two octave spatial

frequency condition. The data from each of the spatial frequency conditions is plotted as RMS contrast threshold versus the number of Fourier components in the stimulus (bottom axis) and stimulus bandwidth (top axis). The dashed line was fit to the data and has slope of 0.25.

The grey square represents the mean white noise threshold, expressed in RMS contrast, for the three observers. The white noise threshold data point was not used when the combined data were fit with a power function. Including this data point provides an instructive test as it demonstrates that the quarter-root law breaks down when the stimulus includes all frequencies and orientations. Although the quarter-root law breaks down for white noise, the number of components required to observe a breakdown of the quarter-root law has yet to be determined.

4.1.4 Discussion

The results of Experiment 4.1 echo the results for spatial frequency summation found in both Experiment 2.1 and [20] as well as the summation found auditory frequency [17]. The important similarity between the results of this chapter and previous work is that observer's TvB functions exhibit a slope of 0.25 in log-log coordinates. Recall that in Chapter 2, the result of a TvB function slope of 0.25 is important as it is the slope produced by an ideal observer. This result, along with the high absolute efficiencies we observed are consistent with the idea that orientation information is summed optimally. In addition to the results suggesting optimal summation, the results are consistent with the idea that noise patterns are detected by an internal filter that has an adjustable orientation bandwidth. This interpretation of the data is consistent with that provided by Kersten and in Chapter 2.

4.2 Experiment 4.2

The observation of a TvB function that follows the quarter-root law has led to the interpretation that a human observer could be using a strategy for detecting filtered noise patterns that resembles the adjustable template used by the ideal observer [17, 20]. In our previous work, TvB functions exhibited optimal summation, but we found the puzzling result that the classification images for the detection of visual noise did not change in a manner consistent with the ideal observer as the bandwidth of the stimulus was increased (see Chapter 2). This inconsistency leads to the conclusion that without data from classification images it is premature to conclude that TvB functions are an index of

optimal summation. In short, TvB functions with quarter-root slopes are not sufficient evidence to conclude that observers are summing information optimally or applying an adjustable template in a noise detection task.

4.2.1 Methods

4.2.1.1 Observers

The two observers who participated in this experiment were both 28 years old, were students at McMaster University, and paid for their participation. Both observers were naïve with respect to the experimental hypotheses, had normal Snellen acuity, and had extensive practice in psychophysical tasks. Both observers ran in the first experiment of this chapter.

4.2.1.2 Apparatus

The apparatus used to run this experiment was identical to that used in Experiment 4.1.

4.2.1.3 Stimuli

The stimuli were the same as those used in the one octave spatial frequency bandwidth conditions in Experiment 4.1. The level of external noise also was the same as in Experiment 4.1.

4.2.1.4 Procedures

The procedures were identical to those used in Experiment 4.1. Observers were highly practiced and ran 2500 trials per classification image.

4.2.2 Results

We used the 2-IFC variant of the response classification image task which has been described in detail previously[1, 2]. The major difference between the classification images measured in previous work and the classification images presented here is that our classification images were calculated using the *power spectra* of the noise masks, rather than the noise masks themselves. On each trial, the Fourier transform of the noise mask in each interval was computed. Next, the difference between the pair of power spectra

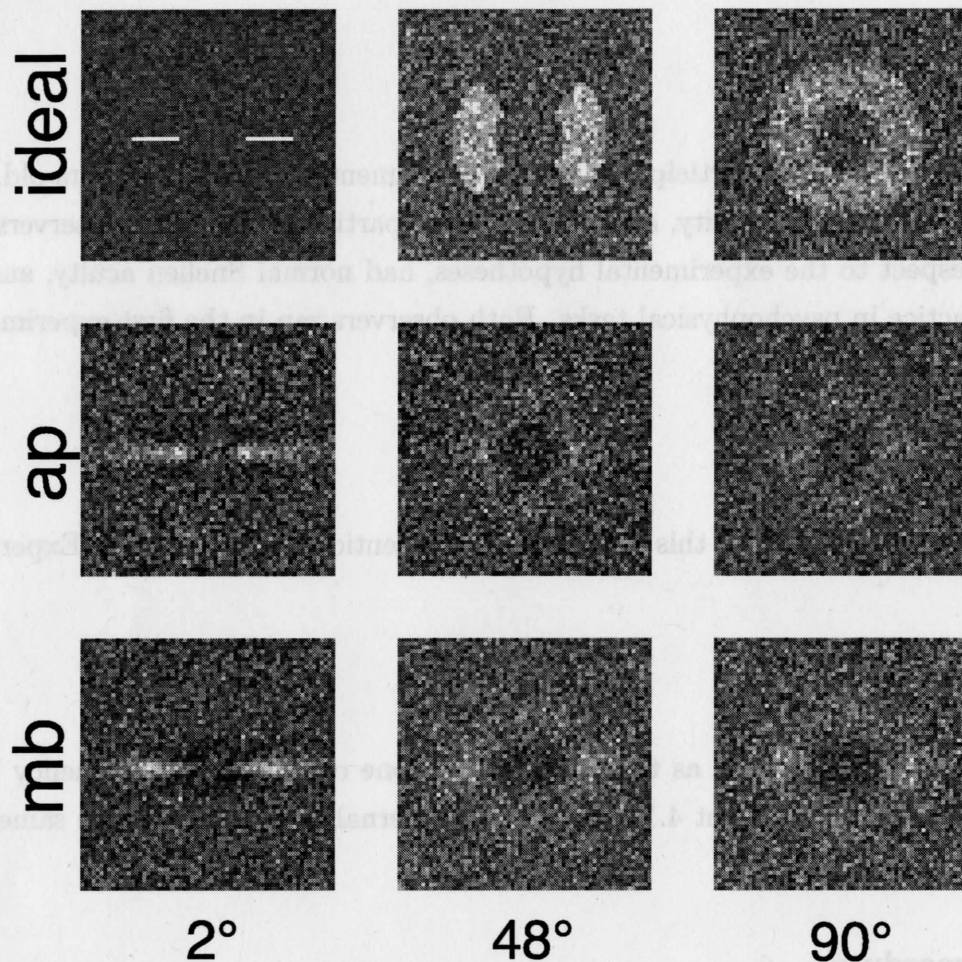


Figure 4.6: Classification images for the ideal observer and two human observers. The top row of images are classification images for the ideal observer and the second two rows are for two observers, MB and AP. The images in each column are from different bandwidth conditions, half-bandwidths of 2° , 48° and 90° . The images are rotate power spectra, spatial frequency increases the distance from the center of the image. Horizontal orientations are represented in the center row of the image and vertical orientations by the center column, intermediate orientations are represented as diagonals through the image between the two cardinal orientations. All images were collected with the same number of trials.

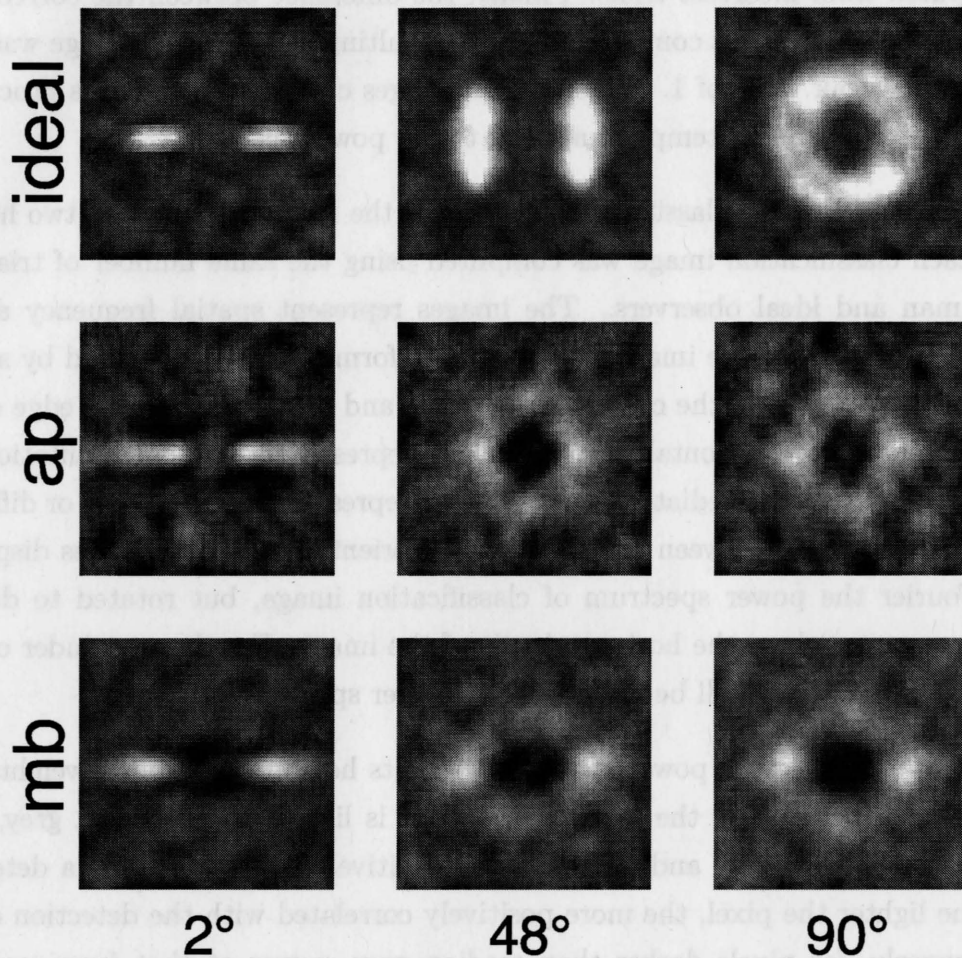


Figure 4.7: Smoothed classification images for the same conditions and observers presented in 4.6. The top row of images are classification images for the ideal observer and the second two rows are for two observers, MB and AP. The images in each column are from different bandwidth conditions, half-bandwidths of 2° , 48° and 90° . White regions represented where power in the stimulus is correlated with a correct detection response, whereas the dark or black regions indicate regions where incorrect responses are correlated with the stimulus. An important comparison to note is the difference between human and ideal observers at off-frequency and orientations in the stimulus. See the text for more detail.

was calculated at each spatial frequency, and the difference spectra was placed into one of four bins based on which interval contained the signal (1 or 2) and the observer's response (correct or incorrect). The power spectra within each bin were averaged: the two average spectra computed from correct responses were summed, as were the two average spectra computed from incorrect trials. Finally, the difference between the correct and incorrect averaged spectra was computed and the resulting classification image was normalized to have a peak value of 1. Classification images calculated using this procedure are proportional to the linear template applied to the power spectra [1, 2].

Figure 4.6 shows the raw classification images for the ideal observer and two human observers. Each classification image was computed using the same number of trials for both the human and ideal observers. The images represent spatial frequency as the distance from the center of the image. Orientation information is represented by sets of pixels in a line that begins in the center of the image and extends to the edge of the image. In the center row horizontal orientations are represented, vertical orientations by the center column, and intermediate orientations are represented as diagonals or different angles through the image between the two cardinal orientations. The images displayed are of the Fourier the power spectrum of classification image, but rotated to display horizontal frequencies along the horizontal axis of the image. For the remainder of this chapter, images of this sort will be referred to as power spectra.

The color of a pixel in the power spectra represents how information is weighted by the observer when performing the task. If the pixel is lighter than median grey, then noise power at that frequency and orientation is positively correlated with a detection response. The lighter the pixel, the more positively correlated with the detection of the signal. Conversely, for pixels darker than median grey, power at that frequency and orientation is negatively correlated with detection.

The power spectra shown are a subset of the full 512×512 power spectra for each condition. The images sample the 64×64 region of each power spectra classification image. The sampled regions include all the spatial frequencies presented in the stimulus. This region corresponds to spatial frequencies from DC to approximately 20 cy/deg. Figure 4.7 shows classification images that are smoothed to reduce spurious noise in the template that results from a limited number of trials. The convolution kernel used was a 5×5 two-dimensional triangle function whose value at each point in the function were set such that the area under the triangle function was one. Convolution with the triangle function had the effect of smoothing the template weights at the frequencies/orientations

in a 5×5 region covered by the triangle function.

Figures 4.6 and 4.7 exhibit important details of the templates measured in this experiment. First, the templates bear a resemblance to the ideal templates: the templates exhibit quite a narrow bandwidth with the smallest stimulus bandwidth and get larger with increasing stimulus. It is apparent from these images that the templates adjust their bandwidth as the stimulus bandwidth is increased from a half-bandwidth 2° to $48^\circ/90^\circ$. However, it is not obvious that the templates for the $48^\circ/90^\circ$ stimulus bandwidth differ for humans, although they do quite clearly for the ideal observer. Whether these two bandwidth conditions differ for humans will be addressed in the next section.

Additionally, there are two curious findings that can be observed in the templates. In the narrowest bandwidth condition, human observers tend to use off-stimulus orientations and spatial frequencies in a manner that negatively correlates with the presentation of the signal – the dark or black regions in Figure 4.7. The ideal observer does not make use of off-stimulus information (orientation and spatial frequency) but instead only uses information contained within the stimulus.

There are two clear examples of off-stimulus information, AP shows *off-orientation* information use and MB shows *off-frequency* information use. AP uses *off-orientation* information, unlike the ideal observer. This is shown in the figure in that the observer's classification image contains dark black regions at orientations other than the orientations contained in the stimulus. MB shows of *off-frequency* information use. MB uses spatial frequency information lower than that used by the idea observer, frequencies that are not contained in the stimulus. The interpretation of these findings will be addressed in the discussion.

4.2.3 Analysis

To relate the classification images to orientation channels found in orientation masking experiments e.g., [15], the two-dimensional classification images collected in this experiment were collapsed into one-dimensional classification images as a function of orientation. To collapse the classification images to one-dimension, they were summed in 1° steps across a band spatial frequencies (filter center-frequency 5 cy/deg and bandwidth of approximately 20 cy/deg) over a 180° range of orientations. The resulting functions, scaled to have a maximum value of 1, are shown in Figure 4.8. One feature of the data that is apparent is that orientations around 0° or horizontal had the strongest influence

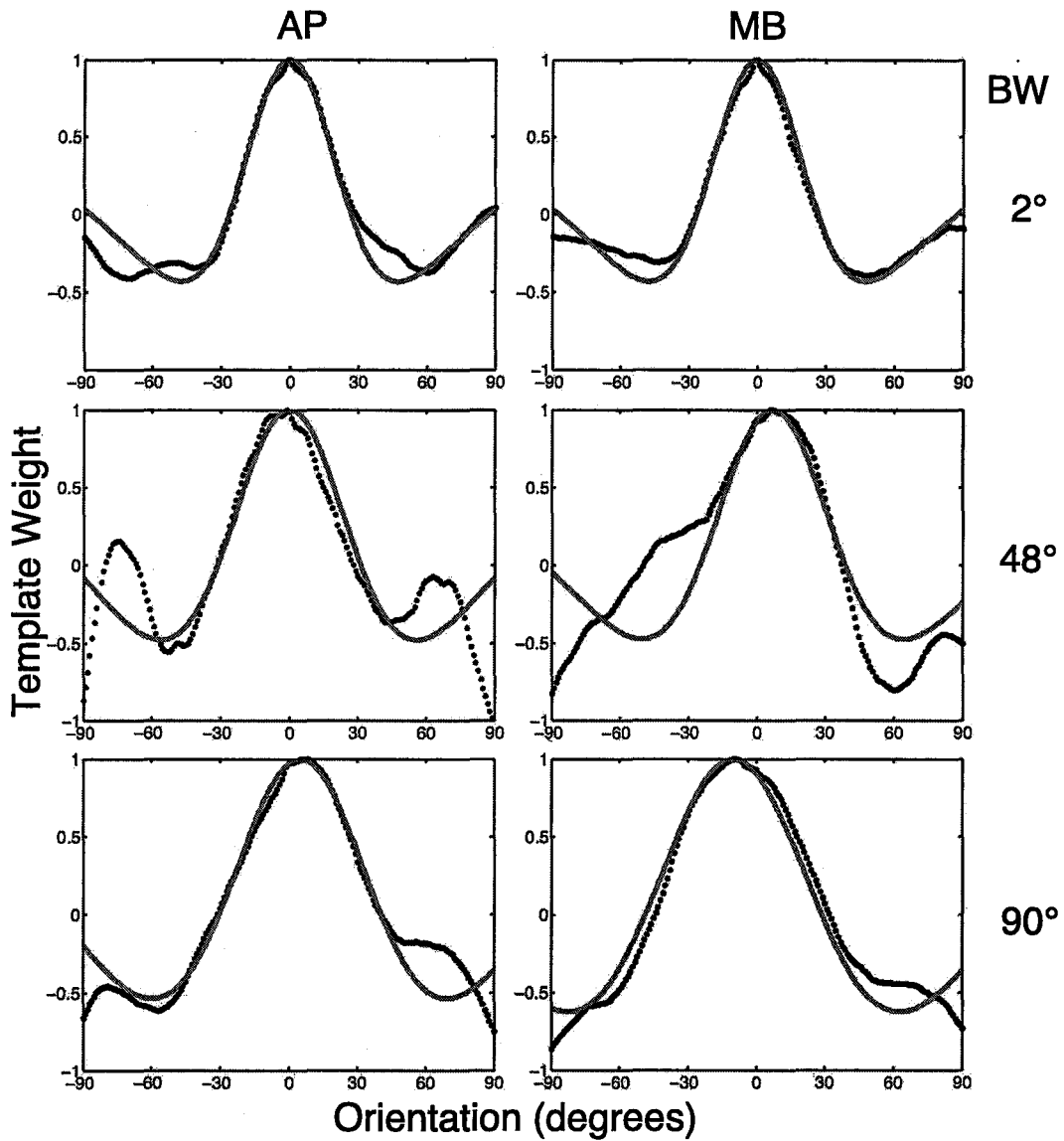


Figure 4.8: The black dotted lines show the one-dimensional circularly summed classification images for two observers in three conditions. The grey lines show the best difference of Gaussian fits to the empirical data. The x-axis in each sub-figure is orientation, the y-axis the weight of the template when scale to have a maximum of 1. See text for more detail.

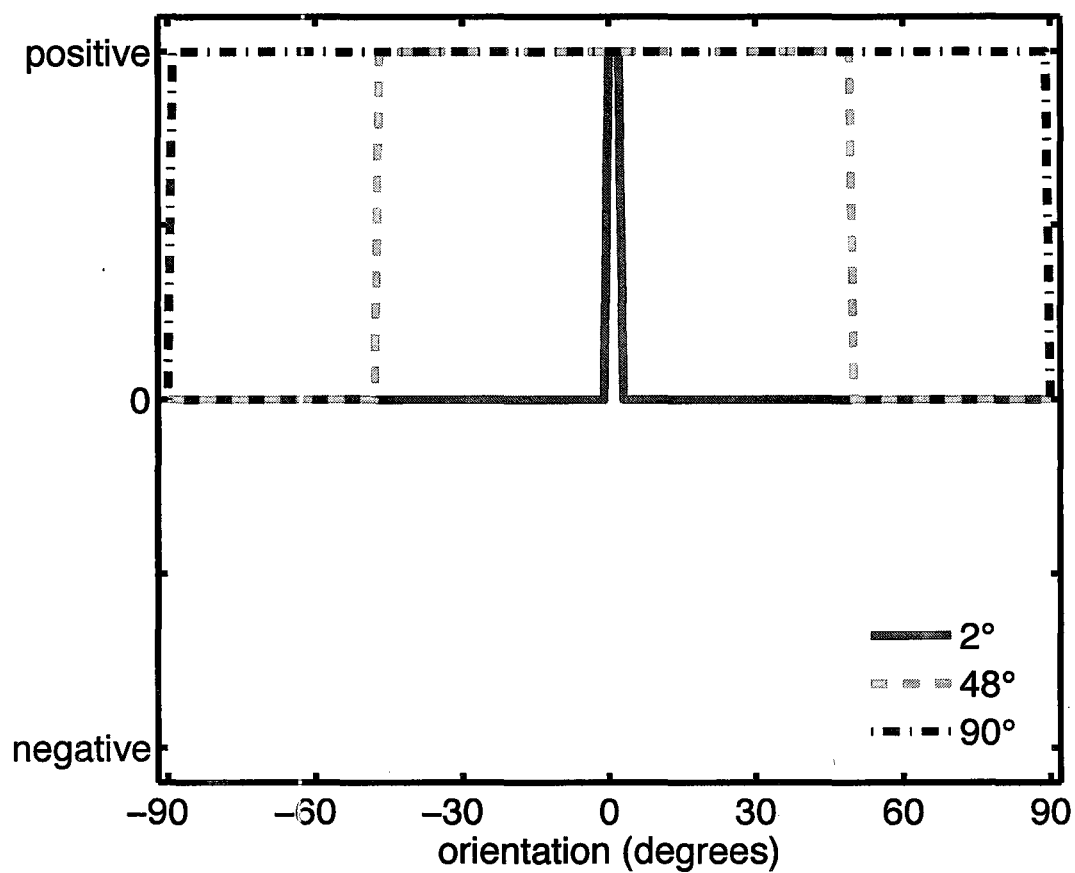


Figure 4.9: Ideal observer templates for the 2°, 48°, and 90° bandwidth conditions. The 2° bandwidth condition is shown in grey, 48° in green, and 90° in blue. The ideal observer weights all orientations positively and sums them equally.

on observers' decisions. Vertical orientations or other orientations far away from zero also had an influence, but in an opposite manner to that of horizontal frequencies. This can be contrasted with the classification images of the ideal observer which are shown in Figure 4.9. The ideal observer uses only the orientations contained within the stimulus and only positively weights the information which is quite different from the human observer data shown in Figure 4.7.

Table 4.2: Template half-width at half-height for two observers in three bandwidth conditions in degrees.

BW	AP	MB
2	17	15
48	18.5	22.5
90	27.5	27.5

Table 4.3: Equivalent Width as defined in Bracewell [4], for two observers in the three conditions measured in degrees.

BW	AP	MB
2	29.6	32.1
48	45.7	36.7
90	48.7	41.6

Tables 4.2 and 4.3 show descriptive statistics of the one-dimensional templates in Figure 4.8. Both the measures of half-width at half-height and equivalent width support the idea that observers adjusted the orientation bandwidth of the templates in response to changes in the orientation bandwidth of the stimulus. However, the increase in template bandwidth is noticeably smaller than one would expect if observers were using an adjustable template, like the ideal observer that matched the width of the stimulus exactly. The extent of the adjustment of the ideal observers template can be seen in Figure 4.8. The adjustment of the ideal observer template is much larger than the adjustment of the weighting of the human observers. The remainder of this section uses curve fitting to explore how this discrepancy could be explained.

To further analyze the results of our classification images, we fit a difference of Gaussians function which contains a positive gaussian component and a negative gaussian

component. In addition to providing better fits than a model that uses a single Gaussian to fit the data, another justification for using a DoG function is that it has been used previously to model orientation channels. [7, 9, 24, 27]. Typically, a DoG function has four free parameters: one for the center/mean and another for the bandwidth/standard deviation for each of the positive and negative Gaussians that comprise the function. In the DoG function fit to our data the number of parameters was limited to three by constraining the fitting routine to equate the centers of both Gaussian functions. A second constraint, namely that the bandwidth of the positive gaussian be narrower than the bandwidth of the negative Gaussian, was also applied. The constraints included in this model are consistent with previous models of orientation channels [6, 7, 24, 27].

Table 4.4: DoG center orientation parameters expressed as σ in degrees and 95% confidence intervals for two observers.

BW	AP	MB
2	0.64 ± 1.99	-0.01 ± 1.21
48	-5.1 ± 7.31	7 ± 5.33
90	5 ± 4.55	9 ± 6.41

Table 4.5: DoG orientation bandwidth parameters, for both the positive and negative Gaussians in the DoG function, expressed as σ in degrees and 95% confidence intervals for two observers.

DoG parameter	AP	MB
2 positive	20.1 ± 1.99	21.0 ± 1.21
2 negative	59.9 ± 0.11	60.8 ± 2.03
48 positive	25.3 ± 4.86	24.9 ± 5.33
48 negative	74.7 ± 0.11	75.8 ± 5.14
90 positive	30.1 ± 2.55	29.9 ± 2.41
90 negative	90.1 ± 0.11	89.8 ± 7.09

Tables 4.4 and 4.5 show the best-fitting (least-squares) parameters and confidence intervals computed via a bootstrap procedure [13]. The DoG fits provide a good description of the classification images. The fits of the bandwidth parameters shown in Table 4.5 are separated into the two components of the fitted DoG function; the positive Gaussian and the negative Gaussian. The fits of the positive Gaussian show an increase in bandwidth as the stimulus bandwidth was increased. This increase in the fitted param-

ters is consistent with the data for the measures half-width at half-height and equivalent width from the classification images. The width of the positive Gaussian increases as the bandwidth of the stimulus is increased.

The fits of the negative Gaussian of the DoG function also show an increase in bandwidth as the stimulus bandwidth is increased. The best fitting parameters of the negative Gaussian have a much larger bandwidth than the positive Gaussian. If the negative Gaussian is interpreted as an inhibitory effect, this result is consistent with previous research on psychophysical inhibitory effects for orientation [21, 24, 29, 34, 35]. The bandwidth of the negative Gaussian increases as the bandwidth of the stimulus increases. The increase in the bandwidth parameter for the negative Gaussian is larger than that observed for the positive Gaussian as it increases from approximately 60° to 90° as the stimulus bandwidth increases.

4.2.4 Discussion

The classification images lead to hypotheses about how observers used stimulus information to perform this detection task. While the templates measured by classification image technique adjust with increasing stimulus bandwidth, Figure 4.7 they do not match the templates used by the ideal observer. One hypothesis to explain this difference is that observers are performing implicit comparisons to perform the detection task. In other words, not only are observers using information within the stimulus orientation and frequency bandwidth of the signal, they are comparing the information content of the signal region to the non-signal region of this stimulus. Using this comparison, if the observer finds that there is a large amount power outside of the stimulus band, the interval with the power outside of the stimulus band will be rejected. Why observers would adopt this sub-optimal strategy remains to be explained.

4.2.5 The Cigars and Donuts Model

Olzak and Thomas [22] proposed a model of summation that could account for the classification images observed in this chapter — the Cigars and Donuts model. The first stage of the model consists of an array of channels that process a narrow band of information. For example, a first-stage detector could process a one-octave band of spatial frequency around 2 cy/deg and orientations 30° either side of vertical. While Olzak and Thomas have focused on spatial frequency and orientation, the model is general enough to describe early vision along any number of other dimensions (e.g., color, motion). To

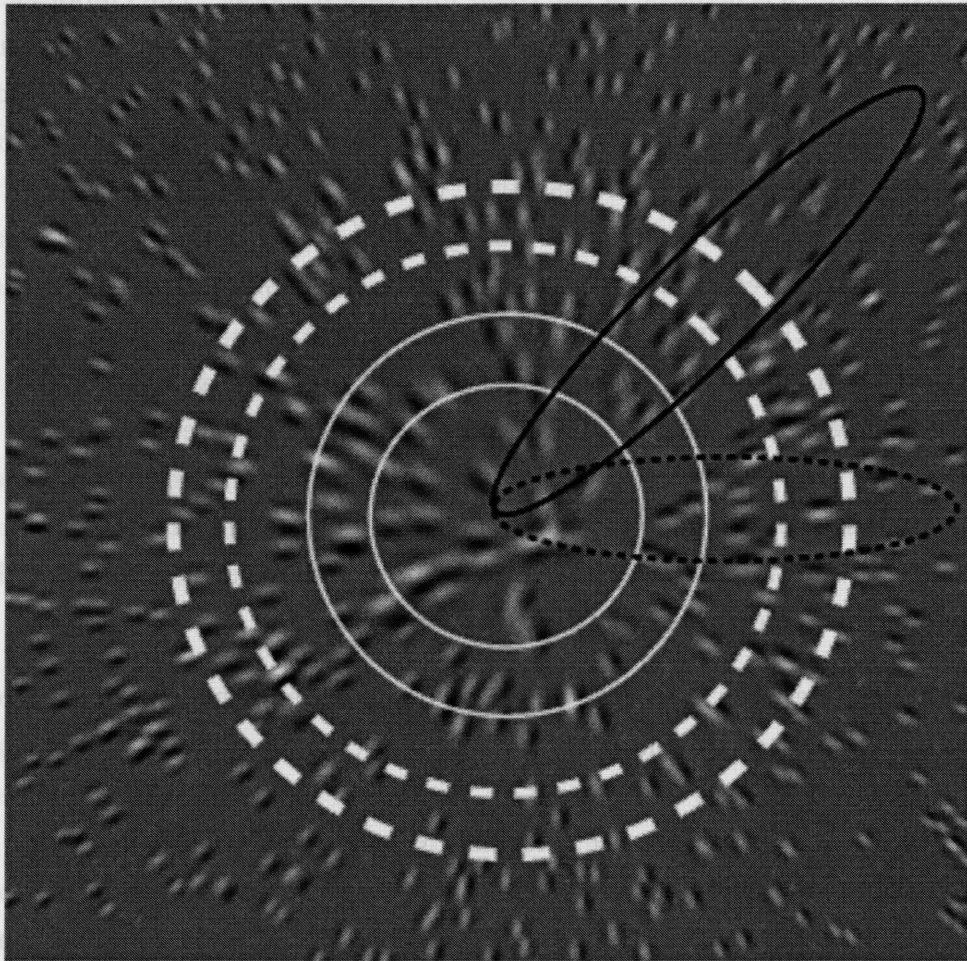


Figure 4.10: A schematic figure of the model, figure adapted from Olzak and Thomas [22]. The figure was originally adapted from the single cell recording work of [10]. In the figure, each Gabor represents a single unit tuned to spatial frequency, orientation and phase. The area within each of the regions bounded within black and white lines is summed by a single mechanism that sums over spatial frequency or orientation. See text for details.

add biological plausibility to their model, a second stage of the model contains two steps, within-channel response non-linearity and a pooled gain control. The next stage of the model sums information across a specified dimension and will be describe in detail as it is the model step that is most relevant to the work in this thesis.

Figure 4.10 shows a schematic of the Cigar and Donut model. The Gabor elements in the figure represent detectors tuned to spatial frequency and orientation. The summation mechanisms of the model are represented by the regions bounded lines. Information within a given region is summed to produce a response that is produced by combining information across the first stage mechanisms. For example, the region outlined in black in Figure 4.10 – a Cigar – sums signals at a small set of similar oblique orientations across a wide range of spatial frequencies. Another Cigar mechanism, represented by the dashed black line in Figure 4.10, sums signals across a wide range of spatial frequencies centered around horizontal orientations. A Donut mechanism sums signals across orientation at a narrow range of spatial frequencies. Two Donut mechanisms are represented by the solid and dashed white lines in Figure 4.10. Donut mechanisms sum across all orientations at low and medium spatial frequencies. The final stages of the Cigar and Donuts model are decision stages, first difference between the two summed dimensions is computed and the final response scaled to observer sensitivity as measured by d' .

The connection between the Cigars and Donuts model and the experiments presented in this chapter can be seen by comparing the schematic of the model shown in Figure 4.10 and the classification images in Figure 4.7. The space represented in the classification images represents how observers use information at a given spatial frequency and orientation. Continuous regions suggest that observers are combining information across spatial frequency or orientation. Thus, it is possible that the classification images are measuring the summation mechanisms proposed by Olzak and Thomas [22]. Specifically, the circular white regions in Figure 4.7 in the 48° and 90° bandwidth conditions could be the result of Donut summing mechanisms.

One limitation of the classification image technique, is that the images are the result of observer responses over a large number of trial, and, as such, do not give a direct measure of what information an observer is using on any given trial. This raises the possibility that observers may not be summing information across a broad range of orientations, as the 48° and 90° bandwidth conditions in Figure 4.7 show, but instead using horizontal/oblique/vertical information on a sub-set of the trials. However, there is reason to think that this hypothesis is unlikely, as any inconsistent use of information will

make the classification images more difficult to measure as observers' responses will not correlate well with the information at a given orientation/frequency in the image. The classification images shown in Experiment 4.2 were measured with 2500 trials, which is, in general a relatively small number of trial compared to other classification image experiments [1, 12, 14, 26].

Lastly, one other puzzling feature of the classification images might be explained by the Olzak and Thomas [22] Cigars and Donuts model: there are regions in the classification image that are negatively correlated with observers' responses, the dark/black regions in Figure 4.7. In other words, information in these regions, drives the observer to respond that a signal was not presented in a given interval of the 2IFC task. That human observers should respond in this fashion is puzzling, as Figure 4.7 shows that the ideal observer does not produce classification images with negatively correlated regions. The dark/black regions in the classification images might be explained by the fact that the final decision stage of the Olzak and Thomas [22] model contains a differencing mechanism that subtracts the responses of two summation mechanisms to model human observers' *discrimination* performance. This leads to the hypothesis that observers could be using a detection strategy that is related to one that would be useful for a discrimination task.

4.2.6 Inhibition/Response Suppression in the Classification Image?

The templates measured in this experiment are consistent with a previous experiment that used sinusoidal gratings and employed a sub-space reverse correlation technique [24]. In the orientation sub-space variant of the classification image technique, stimuli are presented successively in time at a range of orientations and observers were instructed to press a button when they detected a stimulus at the target orientation (e.g., vertical). Ringach found, once the baseline level of responding was corrected for, that observers responded more frequently than baseline to the target orientation, at baseline for orthogonal orientations, but oblique orientations were responded to less frequently than baseline. When Ringach plotted response rate as a function of orientation, the resulting tuning curves are very similar to those presented in the first row of Figure 4.8.

The analysis in the results section processed the two-dimensional classification images and summarized them as the one-dimensional orientation tuning curves presented in Figure 4.8. The negative template weights in Figure 4.8 are the result of the presence of the black/dark regions presented in the classification images shown in Figure 4.7.

Ringach's finding that observers respond less frequently than the baseline level of response to oblique orientations and the result, shown in Figure 4.8, that the template weights take on negative values at oblique orientations can be interpreted that observers are actively suppressing information at oblique orientations. The results from the two tasks raise the question whether the two techniques are tapping into inhibitory or response suppression mechanisms. Ringach, Shapley, and Hawken [25] measured single-cell firing in primary visual cortex (V1) in anesthetized macaques and found a tuning curve via the subspace reverse correlation technique. Ringach et al. found a wide diversity of orientation tuning in V1 including cells that responded below baseline outside of their orientation tuning. Perhaps, both the data of Ringach and the data presented in Figure 4.8 provide psychophysical measures of inhibitory/suppression mechanisms.

The data in Table 4.5 show that the bandwidth of the positive Gaussian or excitatory component ranges from approximately 20 to 30 degrees as the stimulus bandwidth is increased, the negative Gaussian or inhibitory mechanism increases in bandwidth from approximately 60 to 90 degrees. If it is true that the classification images presented here are tapping excitatory and inhibitory response mechanisms and if those mechanisms correspond to the positive and negative components of the DoG function then it raises the hypothesis that inhibitory mechanisms are more flexible/adjustable in their responses than excitatory mechanisms.

4.3 Conclusion

The goal of this chapter was to determine if the results we found in Chapter 2 extended to orientation. Unlike the work on spatial frequency channels in Chapter 2, this chapter provides evidence that is inconsistent with standard models of orientation summation [8, 16, 32, 34, 35]. Not only do the threshold versus bandwidth data exhibit a quarter-root law, but observers' templates exhibit flexibility that is consistent with the idea that the visual system could be using adjustable channels to detect and sum orientation information in a white noise background.

References

- [1] C.K. Abbey and M.P. Eckstein. Classification image analysis: estimation and statistical inference for two-alternative forced-choice experiments. Journal of Vision, 2 (1):66–78, Jan 2002.

- [2] C.K. Abbey, M.P. Eckstein, and F.O. Bochud. Estimation of human-observer templates in two-alternative forced-choice experiments. Proceedings of SPIE, 3663: 284–295, Jan 1999.
- [3] A. Ahumada and J. Lovell. Stimulus features in signal detection. Journal of the Acoustical Society of America, 49:1751–1756, Jan 1971.
- [4] R.N. Bracewell. The Fourier Transform and its Applications. McGraw-Hill, 2000.
- [5] D. Brainard. The psychophysics toolbox. Spatial Vision, 10:443–446, Jan 1997.
- [6] D. Burr, M.C. Morrone, and L. Maffei. Intra-cortical inhibition prevents simple cells from responding to textured visual patterns. Experimental Brain Research, 43(3): 455–458, 1981.
- [7] M. Carandini and D.L. Ringach. Predictions of a recurrent model of orientation selectivity. Vision Research, 37(21):3061–3071, 1997.
- [8] C. Chubb and M.S. Landy. Orthogonal distribution analysis: A new approach to the study of texture perception. In M.S. Landy and J.A. Movshon, editors, Computational Models of Visual Processing, pages 291–301, Cambridge, MA, 1991. MIT Press.
- [9] R.L. DeValois and K.K. DeValois. Spatial Vision. Oxford University Press, New York, 1988.
- [10] R.L. DeValois, D.G. Albrecht, and L.G. Thorell. Spatial frequency selectivity of cells in macaque visual cortex. Vision Research, 22(5):545–59, Jan 1982.
- [11] M.P. Eckstein, S.S. Shimozaki, and C.K. Abbey. The footprints of visual attention in the Posner cueing paradigm revealed by classification images. Journal of Vision, 2(1):25–45, 2002.
- [12] M.P. Eckstein, S.S. Shimozaki, and C.K. Abbey. The footprints of visual attention in the Posner cueing paradigm revealed by classification images. Journal of Vision, 2(1):25–45, 2002.
- [13] B. Efron and B.J. Tibshirani. Introduction to the Bootstrap. Chapman & Hall, 1994.

- [14] J.M. Gold, R.F. Murray, P.J. Bennett, and A.B. Sekuler. Deriving behavioural receptive fields for visually completed contours. Current Biology, 10(11):663–6, Jun 2000.
- [15] S.W. Govenlock, C.P. Taylor, A.B. Sekuler, and P.J. Bennett. The effect of aging on the orientational selectivity of the human visual system. Vision Research, 49(1): 164–172, 2009.
- [16] N.V.S. Graham. Visual Pattern Analyzers. Oxford University Press, New York, 1989.
- [17] D.M. Green. Auditory detection of a noise signal. In J.A. Swets, editor, Signal detection and recognition by human observers, pages 523–547, New York, 1960. Wiley.
- [18] D.M. Green. Auditory detection of a noise signal. Journal of the Acoustical Society of America, 32(1):121–131, Jan 1960.
- [19] D.M. Green. Auditory detection of a noise signal. In J.A. Swets, editor, Signal detection and recognition by human observers, pages 523–547, New York, 1960. Wiley.
- [20] D. Kersten. Statistical efficiency for the detection of visual noise. Vision Research, 27(6):1029–1040, Jan 1987.
- [21] T.S. Meese and D.J. Holmes. Spatial and temporal dependencies of cross-orientation suppression in human vision. Proceedings of the Royal Society B, 274(1606):127, 2007.
- [22] L.A. Olzak and J.P. Thomas. Neural recoding in human pattern vision: model and mechanisms. Vision Research, 39(2):231–256, 1999.
- [23] D.G. Pelli. The videotoolbox software for visual psychophysics: transforming numbers into movies. Spatial Vision, 10(4):437–442, Jan 1997.
- [24] D.L. Ringach. Tuning of orientation detectors in human vision. Vision Research, 38(7):963–972, 1998.
- [25] D.L. Ringach, R.M. Shapley, and M.J. Hawken. Orientation selectivity in macaque V1: diversity and laminar dependence. Journal of Neuroscience, 22(13):5639, 2002.

- [26] A.B. Sekuler, C.M. Gaspar, J.M. Gold, and P.J. Bennett. Inversion leads to quantitative, not qualitative, changes in face processing. Current Biology, 14(5):391–396, 2004.
- [27] M.N. Shirazi. Emergence of orientation-selective inhibition in the primary visual cortex: a Bayes–Markov computational model. Biological Cybernetics, 91(2):115–130, 2004.
- [28] C.P. Taylor, P.J. Bennett, and A.B. Sekuler. Spatial frequency summation in visual noise. Journal of the Optical Society of America A, 26(11):B84–93, Nov 2009.
- [29] P.C. Teo and D.J. Heeger. Perceptual image distortion. In SID International Symposium Digest of Technical Papers, volume 25, pages 209–209. Citeseer, 1994.
- [30] B.A. Wandell. Foundations of Vision. Wiley, New York, 1995.
- [31] G.H. Wetherill and H. Levitt. Sequential estimation of points on a psychometric function. British Journal of Mathematical & Statistical Psychology, 18:1–10, Jan 1965.
- [32] H.R. Wilson and J.R. Bergen. A four mechanism model for threshold spatial vision. Vision Research, 19(1):19–32, Jan 1979.
- [33] H.R. Wilson and D.J. Gelb. Modified line-element theory for spatial-frequency and width discrimination. Journal of the Optical Society of America A, 1(1):124–131, 1984.
- [34] H.R. Wilson and F. Wilkinson. Evolving concepts of spatial channels in vision: From independence to nonlinear interactions. Perception, 26:939–960, Jan 1997.
- [35] H.R. Wilson, D.K. McFarlane, and G.C. Phillips. Spatial frequency tuning of orientation selective units estimated by oblique masking. Vision Research, 23(9):873–82, 1983.

Chapter 5

Conclusion

This chapter has three goals, to briefly summarize the key results of the three preceding chapters, to discuss possible future directions that are related to the work in this thesis, and how the results of this thesis fit within the broader context of vision science.

Results Summary

Chapter 2: Spatial frequency summation in visual noise

Chapter 2 began by examining previous results that suggested human observers combine information across a wide range of auditory and spatial frequencies with high efficiency. Green [15, 16] found that noise was summed optimally in audition and Kersten [19] discovered that the visual system also appears to sum visual noise of increasing bandwidth optimally. Chapter 2 found this result to be replicable and robust to changes in external noise level, randomization, and temporal duration. Despite the robustness and replicability of the results in Chapter 2, the chapter provides an object lesson on the importance of using multiple methods to gather lines of converging evidence. The classification images in Chapter 2 illustrate this lesson, as without measuring classification images, the data from Experiments 2.1, 2.2, and 2.3 ostensibly support a model in which the visual system uses adjustable channels to sum spatial frequency in a noise detection task. The classification images collected in Experiment 2.4 clearly show that adjustable channels are not necessary to produce efficient detection that is constant across stimulus bandwidth. Moreover, the modeling contained within Chapter 2 demonstrates that a

standard multiple channel model [61] is enough to explain both the detection and classification image results. In short, the results of this chapter bring the work of Kersten [19], in line with standard models of spatial frequency summation [13].

Chapter 3: Detection and discrimination of visual noise

The experiments of Chapter 3 were motivated by a curious finding in Experiment 2.4. When classification images were measured in Experiment 2.4 at center frequencies (CF) of 5 and 15 cy/deg, the templates measured were centered on 5 cy/deg in both conditions. The visual system can easily discriminate 5 and 15 cy/deg; in experiment 2.4, however, observers were asked merely to detect a stimulus not to discriminate stimuli centered on those frequencies. To determine if the classification images would differ in a discrimination task, Experiment 3.1 used a double-judgement procedure to simultaneously explore noise detection and discrimination. Experiment 3.1 found that, as in the interpretation of results from Chapter 2 one-octave wide noise patterns each with a CF of 5 and 15 cy/deg are discriminated in a manner consistent with the idea of independent channels or labelled-lines [59]. The classification images for detection replicated the classification images found in Experiment 2.4. Surprisingly, the subtle differences between classification images for detection of 5 and 15 cy/deg patterns could explain the classification images measured for discrimination and could be modeled with the standard model of spatial frequency detection and discrimination [61].

Chapter 4: Orientation summation in visual noise

Chapters 2 and 3 explored the summation of spatial frequency in visual noise. Standard models of the visual detection and discrimination use multiple channel models to encode both spatial frequency and orientation [13, 55, 60, 61, 62]. With this similarity in mind, Chapter 4 examined the summation of orientation information with the methods used in Chapter 2. Detection thresholds for visual noise increased with increasing orientation bandwidth: specifically, detection threshold was proportional to the quarter-root of the number of spatial frequency components in the stimulus. Consequently, absolute efficiency was approximately constant as a function of orientation bandwidth. However, as Chapter 2 showed, TvB detection data consistent with a quarter-root law are not enough evidence to conclude that observers are summing information in a flexible or adjustable manner. Experiment 4.2 measured classification images to reveal the template that observers were using to detect orientation filtered noise. For orientation filtered

noise, unlike spatial frequency filtered noise, we found that the orientation bandwidth of observers' templates adjusted to orientation bandwidth of the stimulus.

Optimal summation is a key idea presented throughout this thesis. The motivation behind this work stems from work by Kersten [19] on the efficiency of detecting one-dimensional spatial frequency filtered noise. As mentioned previously, the results of Kersten may be interpreted as evidence for optimal summation of spatial frequency information across a wide bandwidth, quite unlike the probability summation ideas that are incorporated into standard models [13]. Similar to the approach taken in this thesis, investigating the summation of orientation as well as spatial frequency information in complex identification would address the question of how efficiently the visual system uses information and what information is used when detecting, discriminating, and identifying patterns. However, not only is it important to test a range of stimuli, for example, the comparison made between gratings and one-dimension noise in the Introduction, but a variety of methods are required as well. To be specific, if in Chapters 2 and 4, we had measured only thresholds, instead of also measuring the classification images, we would have missed important data that speak to the adjustable channels hypothesis and difference between the processing of spatial frequency and orientation.

Context Within Vision Research and Possible Future Directions

Levi, Klein and Chen (2005)

Levi et al. [25] measured classification images for noise detection in a background of white noise. In their experiment they varied the range of the highest and lowest spatial frequency in their noise. However, their stimuli differs from the noise used in this thesis; the experiments in this thesis used filtered one- or two-dimensional white noise, whereas Levi et al. summed 11 sinusoidal gratings of random contrast. To manipulate the center frequency and bandwidth of their noise they varied both the spatial frequency and step-size among the frequencies of the sinusoidal gratings. Consequently, unlike the stimuli used in Experiments 2.4 and 3.2, the stimuli used by Levi et al. [25] had center frequencies that increased as stimulus bandwidth increased (Figure 5.1).

Figure 5.2 shows the mean classification image from four observers collected by Levi et al.. If the results of Levi et al. were described using the terminology used in this thesis, their experiments could be summarized as showing that the template center frequency may be adjusted when the stimulus changes in both center frequency and bandwidth and

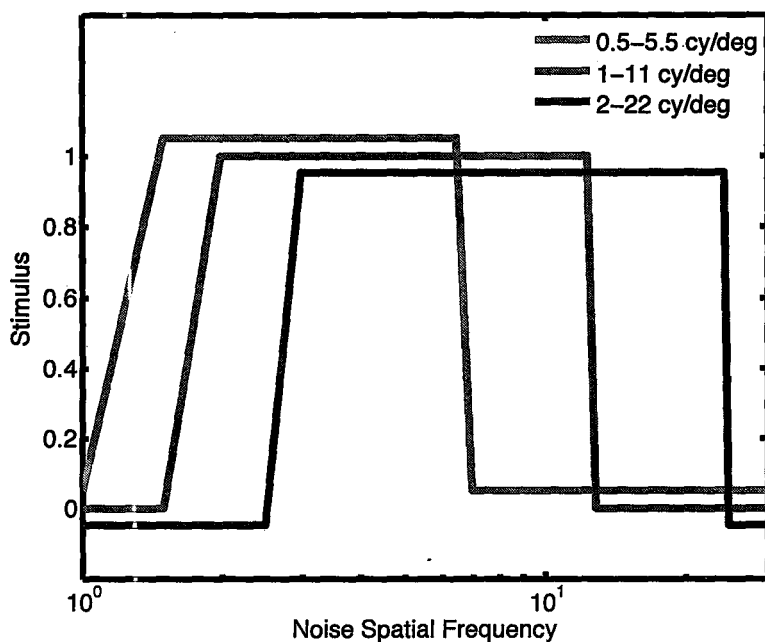


Figure 5.1: Idealized Fourier transforms of the stimuli from Levi et al. [25]. The red line represents the transform of the visual noise Levi et al. from 11 summed cosine gratings. To produce visual noise the contrast of each grating was randomized. This noise differs from the definition of noise in this thesis (see text for details).

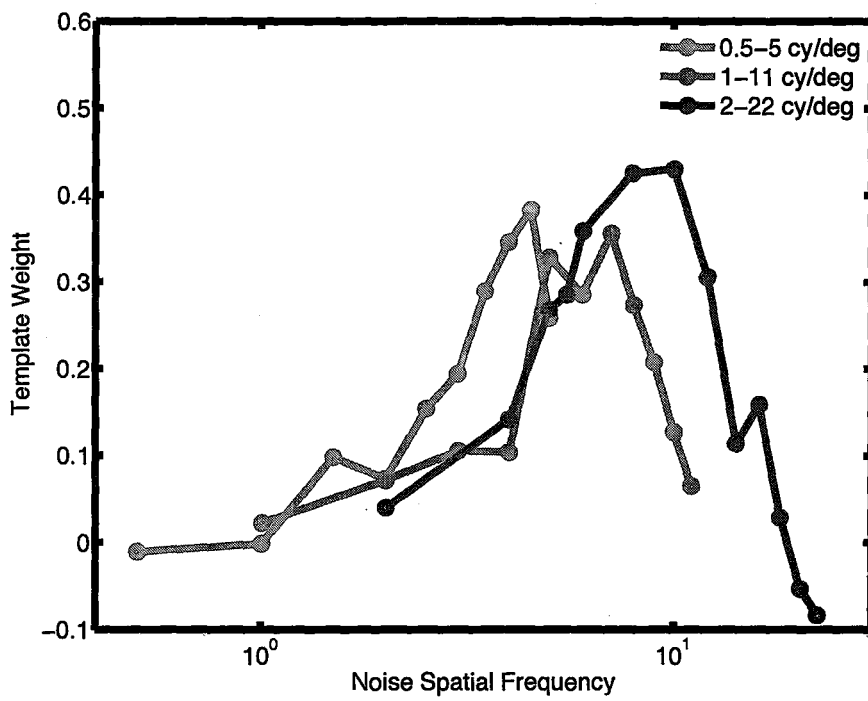


Figure 5.2: Classification images collected by Levi et al. [25]

the stimulus is composed of a small limited number frequency components. The data in Figure 5.2 show a shift in center frequency of the classification image as the range of the stimulus was increased from 0.5–5 to 2–22. Also, Green showed in derivation of the ideal detector that at least 30 Gaussian distributed random variables are required to be summed such that the noise is better described by a normal distribution, rather than a chi-squared distribution. On the surface, these results differ from the data found in Experiments 2.4 and 3.2 of this thesis where we found adjustment of the classification image bandwidth but no adjustment of template center frequency, but because of the substantial differences in the stimuli, it is difficult to interpret the differences between these two results.

To test whether the classification image would change if observers were uncertain about both the stimulus center frequency and bandwidth, the methods used in Experiment 2.2 could be combined with the classification image technique. Experiment 2.2 manipulated observer uncertainty by randomizing the stimulus bandwidth on each trial. Randomizing stimulus center frequency and bandwidth while simultaneously measuring classification images using stimulus centered on a given spatial frequency on log-axes would answer the question of whether the classification image would show adjustment of center frequency or bandwidth.

$\frac{1}{f}$ Amplitude Spectra

Chapter 2 measured detection thresholds for signals that varied in their spatial frequency bandwidth to investigate spatial frequency summation. The patterns used throughout this thesis had amplitude spectra that were, on average, flat across frequency. Research has shown that the amplitude spectra of natural images is $\frac{1}{f}$ on average [8]. The $\frac{1}{f}$ amplitude spectra of natural images has lead researchers to hypothesize that the human visual system may be optimized for natural $\frac{1}{f}$ amplitude spectra [2, 9, 46, 53, 54]. The difference in amplitude spectra in our stimuli and those of natural images, leads to the hypothesis that the difference between the amplitude spectra of natural stimuli and the filtered visual noise stimuli used in this thesis could alter our detection results. For example, using filtered noise stimuli with $\frac{1}{f}$ amplitude spectra could alter the results and interpretation of the curious result found in Experiment 2.4, where classification images for noise stimuli with a center frequency of 15 cy/deg were detected using a template centered at 5 cy/deg, especially if the visual system is utilizing the information contained at low spatial frequencies differently than higher spatial frequencies.

Using $\frac{1}{f}$ would not be a fishing expedition as in least one case the use of $\frac{1}{f}$ has produced results apparently inconsistent with a textbook psychophysical effect, the oblique effect [1]. The oblique effect refers to the finding that horizontally and vertically oriented gratings are perceived better than gratings with an oblique orientation. The oblique effect has been observed with a large variety of experimental methodologies (masking, adaptation, discrimination and detection) and is a classic result in visual psychophysics. However, a recent study by Essock, DeFord, Hansen, and Sinai [7] has shown that when an orientation filtered noise with a $\frac{1}{f}$ amplitude spectrum was used as a stimulus observers discriminate oblique orientation better. Computational models of the oblique effect [26] have yet to be applied to $\frac{1}{f}$ orientation filtered noise patterns and perhaps might be unable to explain this effect. This raises the question of what information are observers using when performing the task of Essock et al., the classification image methods presented in this thesis could answer this question by determining if observers use different aspects of the stimulus when performing tasks to measure the oblique effect with gratings and filtered noise.

One-dimensional Noise and Edges

Edge detection has been an important and long standing problem in vision research [30, 32, 40, 44, 49, 52]. Efficiency and classification images have been measured for edge detection [5, 6, 31]. Edge detection efficiency, depends on the spatial parameters (length and width) of the stimulus, but the maximum efficiency observed was approximately equal to the efficiencies found in Experiment 2.1 with spatial parameters that were similar to those used in this thesis [6].

To explain edge detection, energy models have been proposed [28, 29, 32, 56]. In general, an energy model encodes the stimulus as contrast values, applies a filtering mechanism such as template, and then pools information using Minkowski summation with an exponent equal to two. Such a model is equivalent to the the Wilson and Gelb [61] model used in Chapters 2 and 3. The modeling of edge detection, along with the efficiency and classification image data, also suggests a link between the two tasks.

One-dimensional visual noise and edges are related stimuli. The components in an edge stimulus have a $\frac{1}{f}$ amplitude spectrum and the zero crossings of the the Fourier components contained in the stimulus are aligned. If, instead of being aligned, the zero crossings of the components are randomized, the image will appear to be one-dimensional visual noise with a $\frac{1}{f}$ amplitude spectrum. Thus, one parameter, the alignment of the

phase in the signal can be used to vary a signal from an edge to a one-dimensional noise signal. Edge detection and noise detection lie at the extremes of a continuum of phase alignment. If the same mechanisms are used to detect edges and visual noise, the prediction follows that efficiency ought to be constant as the phase alignment of the signal is varied. Also, the template for edge and noise detection ought to be constant as the phase alignment of the signal is varied as well.

Two Possible Applications of Double-judgment Psychophysics

The work presented Chapter 3 shows that the small differences in the spatial frequency bandwidth of two classification images measured in a detection task are enough to support discrimination a model a labelled line or optimal model of detection/discrimination. The methods of Chapter 3 varied the center frequency of the stimuli used but kept the bandwidth of the stimulus constant. Another approach could be to vary the bandwidth and measure a double-judgement task [20, 33] where stimuli with the same center frequency but different bandwidths could be detected and discriminated.

Two possible results could be obtained from a double-judgement task where noise of two different bandwidths (e.g., one and six octaves) are detected and discriminated. First, if the data do not follow the predictions of Equation 3.1, then this result would indicate that the patterns of different bandwidths are not detected independently. Second, if the data do follow the predictions of Equation 3.1, this supports the assumption of independence. If the first result were found, it would support the idea that one and six octave patterns are not detected/discriminated by independent mechanisms. In either case, the results from this experiment would provide another route for testing the idea of spatial frequency channel independence. It is not unreasonable to believe that if observers are required to do two tasks simultaneously, as in a double-judgement task, they could use different mechanisms from those in summation experiments such as those described by Graham [13].

In addition to adding to the results presented on spatial frequency summation, applying double-judgement methods and the classification image technique could extend the findings of Chapter 4 on the summation of orientation information. In Chapter 3 a simple model based upon human observers' classification images for detection provided a good fit to discrimination data. A combining the classification image technique with double-judgement techniques could investigate the independence of orientation channels. If the orientation information is summed in a way that is suggested by the classification

images presented in Chapter 4, then it makes sense to predict, depending on the stimulus parameters, that Equation 3.1 which relates d' for detection and discrimination would fail.

If Equation 3.1 fails to predict the results of a double-judgement with an orientation filtered noise stimulus then it is possible that summation mechanisms, such as the Cigars and Donuts proposed by Olzak and Thomas [37, 38], and described in Section 4.2.5 of this thesis are being tapped by the task. This would provide another line of evidence that orientation information is not processed independently at detection threshold, which would be inconsistent with other experiments examining orientation summation[3, 12, 13].

ModelFest

The standard model of vision has been an important area in vision research for half a century [18, 47, 48]. Recently, many researchers from several labs have participated in an exercise known as ModelFest and have attempted to develop a standard model of vision using a standard set of calibrated visual stimuli [4, 56, 57]. The goal of ModelFest was to develop a standard observer that could be applied to a wide variety of tasks such as predicting outcomes of corrective laser eye surgery, evaluating display quality, the fidelity of compressed video, and the legibility of text [57, 58].

ModelFest is impressive in that identical experiments were conducted by several labs and detection thresholds measured for a variety of stimuli, from sinusoidal gratings to natural images. One of the series of ModelFest papers, a paper by Watson [56] evaluated the ability of a set of models to predict human observer thresholds. Three models are relevant to the work presented in this thesis, the contrast energy model, the generalized energy model, and the Gabor channels model. The three models will be described briefly and then links to the work presented here will be described.

The contrast energy model, encodes the stimulus as contrast values, then performs spatial filtering, and then pools the information in the stimulus by squaring and summing the contrast information. The squaring and summing operation is a Minkowski summation with an exponent equal to two, a value that was used in previous chapters to model the behavior of human observers. The contrast energy model is the ideal observer used in this thesis is Chapters 2 and 4 but with the addition of internal noise that limits detection thresholds. The contrast energy model fails in that human observer detection thresholds are much greater than the contrast energy model predicts. Specifically, the

contrast energy model under-predicts human observer thresholds by $\approx \frac{1}{2} - \frac{1}{5}$ for checker-board patterns, natural images, and white noise. The failure of the contrast energy model to predict white noise thresholds is the most relevant of the model failures, recall that in Experiment 4.1 white noise detection thresholds were under-predicted by the optimal summation. The model failure found by Watson is thus consistent with the work in this thesis.

The generalized energy model, is a minor modification of the contrast energy model that allows the exponent of the Minkowski summation to vary as a free parameter. The standard interpretation of Minkowski exponents greater than two is that they capture the effects of probability summation [45]. Watson varied the Minkowski exponent from $\approx 2 - 3$ in the generalized energy model to incorporate probability summation. Allowing the exponent to vary, reduced prediction errors for a variety of stimuli, however under-prediction error for natural images and white noise was unchanged. In the modeling work described in Chapter 2 the Minkowski exponent was varied. Values of two to five were used and the result was that the model fits were unchanged. The findings of Watson and the findings of Chapter 2 are consistent in that both models found that over a range of exponents, the Minkowski exponent does little to improve model fits for targets whose spatial frequency bandwidth exceeds that of Gabors and grating stimuli.

The Gabor channels model described by Watson is closest to the Wilson and Gelb model used in Chapters 2 and 3. Unlike Wilson and Gelb model the channels in The Gabor channels model has more free parameters. The model contains eleven spatial frequency channels, four orientation channels and a parameter to set the spatial phase of each Gabor. Watson notes the Gabor channels model does the best, of any of the models tested, in predicting thresholds for white noise stimulus, however the model predicts human observer thresholds of $\approx \frac{1}{2}$ to $\frac{1}{3}$ of their actual values.

None of the models discussed by Watson incorporate adjustable channels [17, 25, 50] or the failures of channel independence [37, 38]. ModelFest was a unique exercise and the researchers involved understandably sampled only a restricted space of stimuli and focused on detection tasks. It could be argued that the level of detail required to capture the results that challenge the standard models of detection used by ModelFest might not be necessary for the applications for which the standard spatial observer designed [57, 58]. On the other hand, for basic vision research, there is no reason not to progress toward a more detailed and complete understanding of pattern vision and such effects may become even more important as vision researchers understanding of primary visual

cortex moves forward in the future[35, 36].

Optimal Summation with Naturalistic Stimuli

Many psychophysical techniques have been used to study both letter identification and word recognition (e.g., [21, 22, 23, 24, 27, 39, 41, 42, 43, 51]). The techniques used and the conclusions drawn from experiments that investigate letters and words assume that the standard model applies to the perception of these complex patterns. That the standard model can explain results in letter/word recognition and identification, speaks to the usefulness of the standard model, that it can be used to solve problems that seem so different from those for which it was developed (e.g., the detection of Gabors and sinusoidal gratings). One recent result in the field of letter identification has found that spatially frequency information is summed optimally [34]. Given that the idea of optimal summation of information has been key throughout this thesis, it is worth describing this study in detail. Recently, efficient summation of spatial frequency information has been found in a 26-AFC letter identification task using filtered and composite stimuli in both the fovea and the periphery. The remainder of this subsection will discuss their experiment and foveal results and relate their work to that presented in this thesis.

Nandy and Tjan [34] used 26 lower-case filtered letters that were filtered to have a one-octave wide full width at half-height stimulus bandwidth at a range of center frequencies: $f_{center}=1.25, 1.77, 2.5, 3.54, 5$ and 7.07 cycles per letter. In their first experiment, Nandy and Tjan asked observers to identify a filtered letter of different center frequencies. From this task, they were able to determine a letter tuning function (LTF) by measuring contrast sensitivity for each of the filtered letters. The interpretation of the LTF is straight forward, it represents the spatial frequency tuning of letter identification and the tuning properties of the LTF can be described by two parameters, peak center frequency and bandwidth.

Once the LTF was measured Nandy and Tjan generated three filtered alphabets. Two alphabets were letters filtered with a low/high center frequency filter of one-half/twice the peak center frequency of LTF. The third alphabet was created by combining the two letter sets. Two aspects of this compound stimulus are important: the compound letters contained non-overlapping spatial frequencies and the contrast ratio of the two components patterns was set to be equal to what it would be in the unfiltered letter in the alphabet. This experiment is analogous to the one conducted by Graham and Nachmias [12] discussed in the Introduction. Nandy and Tjan quantify the amount of summation

using a measure summation they called the integration index, which equals:

$$S = \frac{CS_{low-high}^2}{CS_{low}^2 + CS_{low}^2} \quad (5.1)$$

where CS^2 for each of the conditions equals one over the contrast threshold (contrast sensitivity) squared. Depending on the value of S for a given observer, three interpretations are possible given the assumptions that the visual system behaves as if it were an ideal observer, limited by two factors, the match of a template to the signal, and a fixed additive internal Gaussian noise. The three interpretations are as follows: if S is less than 1, the information provided by the two component stimuli is not combined optimally. If S is equal to 1, then the information from the component stimuli is combined optimally. If the S is greater than 1, then this is evidence that the mechanisms used to encode the stimuli are non-independent. Specifically, given the two assumptions mentioned, a value of S greater than 1 could be the result of templates that do not closely match the width of the signal, but instead have a much broader bandwidth that overlaps in spatial frequency. A second possibility to explain a value of S greater than 1 is that the observers' behavior could be dominated by a noise in a model that is added after the computation for each of the two templates. The possibilities of a value of S greater than or less than 1 will not be discussed further as Nandy and Tjan found that the value of S was equal to 1, indicating efficient or optimal summation of spatial frequency information in a letter identification task.

The evidence from the experiments by Nandy and Tjan also provide evidence for optimal summation. However, as was the case with the results of Kersten the question of what information observers are using in the compound letter identification task remains. While the modeling work of Nandy and Tjan demonstrates that given a set of assumptions, an observer with two-octave wide templates that are summed optimally explains their summation results, a measure of the templates via the classification image technique could alter the interpretation of their results substantially as we found in Chapter 2. As yet, the question of whether orientation information in letter stimuli is summed and if so, whether the summation is efficient remains unanswered.

Another question about the results of Nandy and Tjan crops up when considered in light of work done by Gold et al. [11]. Gold et al. found that the efficiency of letter identification, unlike face identification, was approximately flat as function of stimulus center frequency for stimuli with both one- and two-octave wide spatial frequency bandwidths. While the results of Nandy and Tjan can be interpreted as evidence for optimal sum-

mation of spatial frequency information for letter stimuli, it remains to be determined whether this result generalizes to other patterns, such as faces.

A recent paper by Goffaux and Dakin [10] has found evidence that orientation information is summed optimally for face stimuli. In their work, they filtered faces with horizontal and vertical filters to produce two component stimuli. They also measured observers' sensitivity to the compound stimulus. The logic of this experiment is identical to the summation experiments described in the introduction. Goffaux and Dakin found that for these stimuli sensitivity to the compound stimulus was predicted by a detector that sums the horizontal and vertical component stimuli, and sums them optimally.

Optimal summation is a key idea presented throughout this thesis. As this section describes, researchers are currently following up on the idea of optimal summation and the flexible use of channels and the work presented contributes to this current trend in vision research.

References

- [1] S. Appelle. Perception and discrimination as a function of stimulus orientation: The oblique effect in man and animals. Psychological Bulletin, 78(4):266–278, 1972.
- [2] J.J. Atick and A.N. Redlich. What does the retina know about natural scenes? Neural Computation, 4(2):196–210, 1992.
- [3] C.R. Carlson, R.W. Cohen, and I. Gorg. Visual processing of simple two-dimensional sine-wave luminance gratings. Vision Research, 17(3):351, 1977.
- [4] T. Carney, C.W. Tyler, A.B. Watson, W. Makous, B. Beutter, C.C. Chen, A.M. Norcia, and S.A. Klein. Modelfest: year one results and plans for future years. In Proceedings: SPIE The International Society for Optical Engineering, pages 140–151. International Society for Optical Engineering, 1999, 2000.
- [5] J. Elder and Y. Morgenstern. Nonlinear pooling mechanisms underlying edge detection. Perception, 36 36 ECVF Abstract Supplement, 2007.
- [6] J.H. Elder and A.J. Sachs. Psychophysical receptive fields of edge detection mechanisms. Vision Research, 44(8):795–813, 2004.
- [7] E.A. Essock, J.K. DeFord, B.C. Hansen, and M.J. Sinai. Oblique stimuli are seen

- best (not worst!) in naturalistic broad-band stimuli: a horizontal effect. Vision Research, 43(12):1329–1335, 2003.
- [8] D.J. Field. Relations between the statistics of natural images and the response properties of cortical cells. Journal of the Optical Society of America A, 4(12): 2379–2394, 1987.
 - [9] D.J. Field. What is the goal of sensory coding? Neural Computation, 6(4):559–601, 1994.
 - [10] V. Goffaux and S.C. Dakin. Horizontal information drives the specific signatures of face processing. Frontiers in Perception Science, 1:1–14, 2010.
 - [11] J. Gold, P.J. Bennett, and A.B. Sekuler. Identification of band-pass filtered letters and faces by human and ideal observers. Vision Research, 39(21):3537–3560, 1999.
 - [12] N. Graham and J. Nachmias. Detection of grating patterns containing two spatial frequencies: a comparison of single-channel and multiple-channel models. Vision Research, 11(3):251–259, Jan 1971.
 - [13] N.V.S. Graham. Visual Pattern Analyzers. Oxford University Press, New York, 1989.
 - [14] D.M. Green. Auditory detection of a noise signal. In J.A. Swets, editor, Signal detection and recognition by human observers, pages 523–547, New York, 1960. Wiley.
 - [15] D.M. Green. Auditory detection of a noise signal. Journal of the Acoustical Society of America, 32(1):121–131, Jan 1960.
 - [16] D.M. Green. Auditory detection of a noise signal. In J.A. Swets, editor, Signal detection and recognition by human observers, pages 523–547, New York, 1960. Wiley.
 - [17] S. Haykin. Adaptive digital communication receivers. IEEE Communications magazine, 38(12):106–114, 2000.
 - [18] D.H. Kelly. Visual response to time-dependent stimuli. i. amplitude sensitivity measurements. Journal of the Optical Society of America, 51:422–9, Apr 1961.
 - [19] D. Kersten. Statistical efficiency for the detection of visual noise. Vision Research, 27(6):1029–1040, Jan 1987.

- [20] S.A. Klein. Double-judgment psychophysics: problems and solutions. Journal of the Optical Society of America A, 2(9):1560–85, Sep 1985.
- [21] G.E. Legge, D.G. Pelli, G.S. Rubin, and M.M. Schleske. Psychophysics of reading—I. Normal vision. Vision Research, 25(2):239–252, 1985.
- [22] G.E. Legge, G.S. Rubin, D.G. Pelli, and M.M. Schleske. Psychophysics of reading—II. Low vision. Vision Research, 25(2):253–265, 1985.
- [23] G.E. Legge, G.S. Rubin, and A. Luebker. Psychophysics of reading. V. The role of contrast in normal vision. Vision Research, 27(7):1165–1177, 1987.
- [24] G.E. Legge, T.A. Hooven, T.S. Klitz, J. Stephen Mansfield, and B.S. Tjan. Mr. Chips 2002: New insights from an ideal-observer model of reading. Vision Research, 42(18):2219–2234, 2002.
- [25] D.M. Levi, S.A. Klein, and I. Chen. What is the signal in noise? Vision Research, 45(14):1835–1846, Jun 2005.
- [26] L.M. Lin and H.R. Wilson. Fourier and non-Fourier pattern discrimination compared. Vision Research, 36(13):1907–1918, 1996.
- [27] N.J. Majaj, D.G. Pelli, P. Kurshan, and M. Palomares. The role of spatial frequency channels in letter identification. Vision Research, 42(9):1165–1184, 2002.
- [28] V. Manahilov and W. Simpson. Energy model for contrast detection: spatiotemporal characteristics of threshold vision. Biological Cybernetics, 81(1):61–71, 1999.
- [29] V. Manahilov and W.A. Simpson. Energy model for contrast detection: Spatial-frequency and orientation selectivity in grating summation. Vision Research, 41(12):1547–1560, 2001.
- [30] D. Marr and E. Hildreth. Theory of edge detection. Proceedings of the Royal Society of London B, pages 187–217, 1980.
- [31] Y. Morgenstern, J.H. Elder, and Y. Hou. Contrast dependence of spatial summation revealed by classification image analysis. Journal of Vision, 4(8):539, 2004.
- [32] M.C. Morrone and DC Burr. Feature detection in human vision: A phase-dependent energy model. Proceedings of the Royal Society of London B, 235(1280):221–245, 1988.

- [33] J. Nachmias and A. Weber. Discrimination of simple and complex gratings. Vision Research, 15:217–223, Jan 1975.
- [34] A.S. Nandy and B.S. Tjan. Efficient integration across spatial frequency channels in foveal and peripheral vision. Journal of Vision, 8(13):3, 2008.
- [35] B.A. Olshausen and D.J. Field. What is the other 85% of V1 doing. In T.J. Sejnowski and L. van Hemmen, editors, Problems in Systems Neuroscience. Oxford University Press, 2004.
- [36] B.A. Olshausen and D.J. Field. How close are we to understanding V1? Neural Computation, 17(8):1665–1699, 2005.
- [37] L.A. Olzak and J.P. Thomas. Neural recoding in human pattern vision: model and mechanisms. Vision Research, 39(2):231–256, 1999.
- [38] L.A. Olzak and T.D. Wickens. Paradigm shifts: New techniques to answer new questions. Perception, 28(12):1509–1531, Jan 1999.
- [39] İ. Oruç, M.S. Landy, and D.G. Pelli. Noise masking reveals channels for second-order letters. Vision Research, 46(8-9):1493–1506, 2006.
- [40] E. Peli. Feature detection algorithm based on a visual system model. Proceedings of the IEEE, 90(1):78–93, 2002. ISSN 0018-9219.
- [41] D.G. Pelli, B. Farell, and D.C. Moore. The remarkable inefficiency of word recognition. Nature, 423(6941):752–756, 2003.
- [42] D.G. Pelli, C.W. Burns, B. Farell, and D.C. Moore. Identifying letters. Vision Research, 46(28):4646–4674, 2006.
- [43] D.G. Pelli, C.W. Burns, B. Farell, and D.C. Moore-Page. Feature detection and letter identification. Vision Research, 46(28):4646–4674, 2006.
- [44] P. Perona. Steerable-scalable kernels for edge detection and junction analysis. Image and Vision Computing, 1992.
- [45] R.F. Quick. A vector-magnitude model of contrast detection. Biological Cybernetics, 16(2):65–67, 1974.
- [46] D.L. Ruderman. The statistics of natural images. Network: Computation in Neural Systems, 5(4):517–548, 1994.

- [47] O.H. Schade. Optical and photoelectric analog of the eye. Journal of the Optical Society of America, 46(9):721–39, Sep 1956.
- [48] R. Sekuler and L. Ganz. A new aftereffect of seen movement with a stabilized retinal image. Science, 139:419–420, 1963.
- [49] R.M. Shapley and D.J. Tolhurst. Edge detectors in human vision. The Journal of Physiology, 229(1):165, 1973.
- [50] H. Simon. Adaptive filter theory. New Jersey, Prentice-Hall Inc, 1996.
- [51] J.A. Solomon and D.G. Pelli. The visual filter mediating letter identification. Nature, 369(6479):395–397, 1994.
- [52] C.F. Stromeyer and S. Klein. Spatial frequency channels in human vision as asymmetric (edge) mechanisms. Vision Research, 14(12):1409–20, 1974.
- [53] D. Tolhurst, Y. Tadmor, and T. Chao. The amplitude spectra of natural images. Ophthalmic and Physiological Optics, 12(1):97, 1992.
- [54] J.H. Van Hateren and A. Van der Schaaf. Independent component filters of natural images compared with simple cells in primary visual cortex. Proceedings of the Royal Society B, 265(1394):359, 1998.
- [55] B.A. Wandell. Foundations of Vision. Wiley, New York, 1995.
- [56] A.B. Watson. Visual detection of spatial contrast patterns: Evaluation of five simple models. Optics Express, 6(1):12–33, 2000.
- [57] A.B. Watson and A.J. Ahumada Jr. Spatial standard observer for visual technology. In 2005 IEEE International Conference on Systems, Man and Cybernetics, volume 2, 2005.
- [58] A.B. Watson and A.J. Ahumada Jr. A standard model for foveal detection of spatial contrast. Journal of Vision, 5(9):717–740, 2005.
- [59] A.B. Watson and J.G. Robson. Discrimination at threshold: labelled line detectors in human vision. Vision Research, 21:1115–1122, 1981.
- [60] H.R. Wilson and J.R. Bergen. A four mechanism model for threshold spatial vision. Vision Research, 19(1):19–32, Jan 1979.

- [61] H.R. Wilson and D.J. Gelb. Modified line-element theory for spatial-frequency and width discrimination. Journal of the Optical Society of America A, 1(1):124–131, 1984.
- [62] H.R. Wilson and F. Wilkinson. Evolving concepts of spatial channels in vision: From independence to nonlinear interactions. Perception, 26:939–960, Jan 1997.