INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

Human Visual Speed Perception: Psychophysics and Modeling

by

David Ascher

B.Sc., Brown University, 1990

Thesis

Submitted in partial fulfillment of the requirements for the

Degree of Doctor of Philosophy in the Department of

Cognitive and Linguistic Sciences at Brown University

May 1998

UMI Number: 9830405

Copyright

by

David Ascher

1998

This thesis by David Ascher

is accepted in its present form by the Department of

Cognitive and Linguistic Sciences as satisfying the

dissertation requirement for the degree of Doctor of Philosophy

Date ...9/29/97....   .....*(signature)*.....

Leslie Welch, Advisor

Recommended to the Graduate Council

Date ...9/25/1957...   .....*(signature)*.....

James Anderson, Reader

Date ...9/26/1997...   .....*(signature)*.....

Norberto Grzywacz, Reader

Approved by the Graduate Council

Date ...Oct. 6, 1997...   .....*(signature)*.....

Peter J. Estrup

Dean of Graduate School and Research

ii

Vita

David Ascher

Born in France, October 12, 1968.

Education

- Ph.D. Cognitive and Linguistic Sciences, Brown University, May 1998.

- Sc.B. Physics, *magna cum laude*, Brown University, May 1990.

- Visiting Fellow, Functional Magnetic Resonance Imaging Program, Massachussetts General Hospital, MA, 1995.

- Cognitive Neuroscience Summer School, University of California at Davis, CA, 1995.

- Connectionist Models Summer, University of Colorado at Boulder, CO, 1993.

Teaching and work experience

- Teaching Assistant, Introduction to Neural Networks, 1995 and 1996.

- Teaching Assistant, Introduction to Cognitive Science, 1995 and 1996.

- Senior Software Engineer, Computing and Information Services, Brown University, 1990–1992.

Awards

- Research Fellowship, Brown University, 1996–1997.

- University Fellowship, Brown University, 1993–1994.

Publications

- D. Ascher and L. Welch and N.M. Grzywacz, *Integration across spatial frequency channels in speed discrimination.* Society for Neuroscience, 26th Annual Meeting, #349.6, (1997).

- D. Ascher and L. Welch, *Integration in speed discrimination across wide ranges of spatial frequencies and orientations.* Investigative Ophtalmology and Visual Science (Suppl.), **38**, (1997).

- D. Ascher and L. Welch and E.K. Festa, *Adaptation to stationary gratings results in an increase in apparent speed of moving gratings.* Investigative Ophtalmology and Visual Science (Suppl.), **37**, (1996).

- D. Ascher, *A Connectionist Model of Morse Code Perception.* Proceedings of the 1993 Connectionists Models Summer School. Mozer, Smolenksy, Elman and Weigand, Eds. (1994).

Preface

There are countless ways to approach a given area of scientific inquiry such as perception, and one hopes that the specific topic a given scientist devotes his or her energies to for years is one which is well-suited to their interests and avocations. My dissertation concerns the human ability to perceive the speed of objects moving in their field of view. While one might conclude from this that speed perception is a topic which has fascinated me for years because of its inherent interest, one would be mistaken. Instead, speed perception is a topic which I chose because it has properties which, I have found, are characteristic of problems I find interesting. These properties are:

Speed perception is a psychological phenomenon

I am fundamentally interested in *perception*, which I define to be the processes and structures responsible for the ability of animals to be aware of their environment. Speed perception in this regard is clearly one of myriad other tiny aspects of perception. But it sits squarely in the world of perception—while it is implemented by a biological substrate, one of the characteristics of speed perception which appeals to me is that this implementation is not obvious from what we know about biology. As will be discussed in detail later, the speed which my subjects perceive and base their judgments on is expressed in psychological units, not neuronal units or even physical,

"real-world" units. It is a mental construct, which while built from biological parts and with, most of the time, a fairly stable relationship to something about the world out there, lives only in our heads.

Speed perception is tightly bound to physiology

As just mentioned, speed perception is a psychological entity built from biological parts. As will be described in detail later, the considerable physiological data which has been accumulated on visual motion perception does not paint a clear picture of *how* the brain figures out how fast the things it sees are going. In fact, the mechanisms which the brain appears to use are quite unlike those an engineer would use when designing a speed sensor. This is no doubt due to the fact, first, that nature does not design task-specific systems but instead evolves processes which yield increased adaptability of the organism as a whole, and, more usefully perhaps, to the fact that visual neurons are not speed sensors. Instead, each neuron is a color sensor, shape sensor, distance sensor, friendliness sensor, and mother sensor, all at the same time. The paucity of careful studies of speed perception is an indication of our guess at the place of "speed" on the neuronal to-do list. Whatever data is available has shown that analysis of neuronal responses does not easily explain how speed is perceived. However, neurons *do* transmit all the information which is used (by later neurons) to compute speed. Hence the puzzle, hence the fun.

A field where theories and data mingle

Science mostly progresses in small, evolutionary steps. Experimentalists design and run experiments which produce data, and theorists read the data and dream up theories to explain existing data and predict the results of future experiments. In

some subfields of Physics, the theorists are ahead of the experimentalists by at least a lifetime—most experimentalists test old theories and most theorists will never see their theories put to the test. In vision research, on the other hand, one can do both theory and experiments. Furthermore, if one wishes to, one can do both theory and experiments which are very closely tied to one another. Speed is such an area. As will be clear in the next few chapters, the respective jargons used by theorists, psychophysicists and physiologists when discussing speed are easily mapped one onto another. The contribution of theory in shaping the experimental work is clear, as the predictions of speed models bring directly to bear on what experiments should be done. The reciprocal relationship holds, as psychophysical data can sometimes irrefutably disprove a model. This interplay is due to the fact that computational theories are models of the brain and are expected to produce the same data as the experiments. By studying a "low-level" process such as speed, one can be as grounded in data as one wants to be. Which suits me fine.

Acknowledgments

I would like to thank the following people for making my graduate tenure in general, and my dissertation work in particular, challenging and pleasureable both.

My first thanks go to the graduate students in the Department of Cognitive and Linguistic Sciences. Their companionship and friendships made it fun to go to work everyday. Special thanks go to the "old guard," who showed by example how a multidisciplinary collegial community can work. I hope to find again such a friendly group where the best academic traditions of respect for others' endeavors and enthusiasm for learning are made human by boundless generosity and kept alive by a healthy ordering of priorities. Paul Allopenna, Amit Almor, Sam Bayer, Gary Byma, Steve Finney, Silvia Gennari, Emily Pickett, Thanassi Protopapas, Eric Sklar, Gary Strangman, Jennifer Utman—you all were role models, colleagues and friends, and I am glad to have known you.

The faculty of the department also deserve thanks, for their teaching and research excellence, and perhaps even more for fostering an environment where graduate students are viewed as colleagues rather than subordinates. In particular, one of my advisors, Jim Anderson, in addition to being a reader of this dissertation, encouraged me and supported me throughout my years in the department. I thank him for this, for his generosity both intellectual and otherwise, and for always having a smile at the ready.

Norberto Grzywacz, in addition to carefully reading and critiquing the disserta-

tion, is to thank for repeatedly making me remember what I love about what I do, and making it all seem doable. I very much look forward to working with him in San Francisco.

Elena Festa, Nestor Matthews, Dawn Vreven, Vlada Aginsky and Cullen Jackson deserve thanks for being great colleagues, for their patience as subjects, and for accepting me as an "honorary psych student." I'll see you all at conferences, I'm sure. Dawn, I owe you subject hours.

Very special thanks to my main advisor on this disseration, Leslie Welch. She accepted my request to do "a project" in her lab, which quickly evolved into two projects, then three, then a dissertation. I've learned a great deal about psychophysics, vision, and academia from her, and I'm proud to have been one of her first students.

Finally, I'd like to thank my family. My parents, JacSue and Philippe Ascher, for more than I could ever express, but starting with teaching me that life is to be enjoyed and that learning is fun. My brother, Ivan, for being the kindest person I know and a wonderful friend. My wife, Emily Pickett, for the five great years we've spent together and in anticipation of all the years to come. Our son, Hugo, for making this summer not just the end of a graduate career but the beginning of a new phase in my life.

Contents

List of Figures

xvi

List of Tables

Chapter 1

Introduction

When walking through the world we, like all animals, use visual information to tell us where we are heading, where prey is (and where it is likely to be), where predators are (and whether they are coming our way). This information is typically categorized by scientists according to the physical dimensions along which we believe the world (the stimulus) varies. Take for example the case of a cheetah chasing an antelope. The cheetah is looking at the antelope it has chosen for its supper, while running as fast as its muscles will let it. The cheetah's task is quite simple in theory—it has to keep running while keeping the antelope straight ahead, so that it is likely to intercept it. To do this, its visual system must identify the antelope within its visual field—it probably uses *color* information and *spatial* information to do so. It must detect movements of the antelope away from the cheetah's path, probably by detecting the change in the position of the antelope on its retina away from the fovea—this is *visual motion* information, the topic of this dissertation. There are many levels at which the processes underlying the perception of visual motion can be studied. At one end of the motion system, one can study how the neurons in the early stages of the visual cortex are able to extract motion signals from their afferent connections. At the other end of the motion continuum, one can study how the

1

perceptual system is able to guide active movement using the information carried by the pattern of flow of objects in the visual world. The work reported in this dissertation concerns a level of processing which is between these two extremes. I am interested in the processes by which the visual system goes from the earliest level of neuronal coding of visual motion and obtains (computes) the perceptual quality we call perceived retinal speed—how fast something is moving across the visual field when the eyes are kept stationary[1].

The remainder of this chapter will present a brief overview of the first models of motion perception, and a very brief sketch of speed perception as the analysis of outputs of linear filters, a point of view common in contemporary models of early vision. These two sections are designed to provide an intellectual framework for the two review chapters which follow. The first of those chapters covers the experimental evidence concerning the type and characteristics of the first stages of motion perception and of the speed perception process, while the second chapter presents a more detailed review of the most significant contemporary models of speed perception, which in turn motivated the experimental work described later.

1.1   Basic Motion Perception

Visual motion (which will be referred to hereafter simply as motion) can be defined as a change in the spatial structure of a visual display over time. How this definition is interpreted depends on one's perspective, however. From the perspective of the world, visual motion is but the consequence of changes in the physical world, such as the motion of objects relative to an observer, or the motion of light sources or

---

[1]I am explicitely excluding the perception of object motion in real world coordinates, as most accounts of "real-world" object velocity perception start by assuming retinal speed estimates. It is this earlier stage which this dissertation focuses on.

2

occluders relative to a line of sight. This is in fact the definition of motion which is *behaviorally* useful, and which perceptual systems are built to decode. Decode, because while the source of the motion signal is most naturally expressed in terms of objects and light sources moving through space, the interface between the world and the perceptual system is a retina, which, like all measuring devices, loses information about the causal nature of the events it is measuring. For example, human retinæ contain a few million photoreceptors which are basically sensitive to one thing and one thing only—photons. There are different types of photoreceptors more or less sensitive to different types of photons, but from an individual neuron's perspective, there is no such thing as motion—there is light and dark. The visual system's task is thus what is often called an "inverse problem": given a set of photon landings on the retina (or more precisely, given a set of photon absorptions by retinal neurons), what, in the world ahead, is most likely to have happened? In other words, "what in the world happened that I see before me?" A discussion of how this problem is solved generally is clearly beyond the scope of this dissertation. We will consider instead how the visual system solves the smaller problem of "how fast on my retinæ is that object I see moving before me?" Note that while visual motion does not require the motion of an object, and can be elicited by, for example, ego motion, this dissertation will focus on the simpler problem of the translatory speed of objects in an otherwise stationary world. Introspection is often as good place as any to start a scientific enquiry. How might one decode motion information, for example how fast something is going, based on the responses of retinal neurons, which can only indicate whether they have just been hit by a few photons or many? Clearly the number of photons a given neuron absorbs (related to the luminance of the stimulus at that neuron's corresponding location in the visual field) is not *by itself* useful in telling us how fast an object is moving. However, if one can somehow pick up on the fact that a relatively bright light source sends photons through the eye to one neuron

3

(which is in a fixed relationship to a region of visual space) and then, after a given delay, to a different neuron, located nearby on the retina (and thus corresponding to a nearby region of the visual world), then one can say that it is likely that whatever caused the bright spot on the retina in the first interval *moved* so that it later caused the bright spot in the other location on the retina[2].

1.1.1 The Reichardt-Hassenstein Model

This insight is at the heart of the first published model of motion sensing, the Reichardt-Hassenstein model. The model as described by Reichardt (Reichardt, 1961) was designed to account for the physiological data observed in a set of insects, notably the beetle and two fly species, *Drosophila*, and *Musca domestica*. It is helpful to remember that the eyes of these animals are composed of arrays of ommatidia, receptors with very well-defined receptive fields, almost pointlike in nature. The gross physiological result the model addresses is that there are cells in the insect eyes which respond best to motion of bars across their receptive fields in a specific direction, but they respond poorly when there is no stimulus present or when the stimulus is moving in the opposite direction from the cells' preferred direction. In other words, the cells are motion-sensitive (they "prefer" motion), orientation-selective (they prefer specific stimulus orientations), and direction-selective (they prefer movement along one direction over motion in the opposite direction). These facts about insect ommatidia led Reichardt to formulate the model of motion detection depicted in Figure 1.1.

The simple Reichardt detector comprises two subunits, each subunit being built

---

[2]To be somewhat more rigorous, one can only conclude motion if one knows that the world is not full of twinkling spots of light which go on and off randomly—luckily the visual system made that assumption a long time ago (which is why twinkling "noise" appears to be moving even when it is just twinkling, why television and cinema fool us so convincingly, and why one can study motion perception without having to move physical objects around all day long!).

4
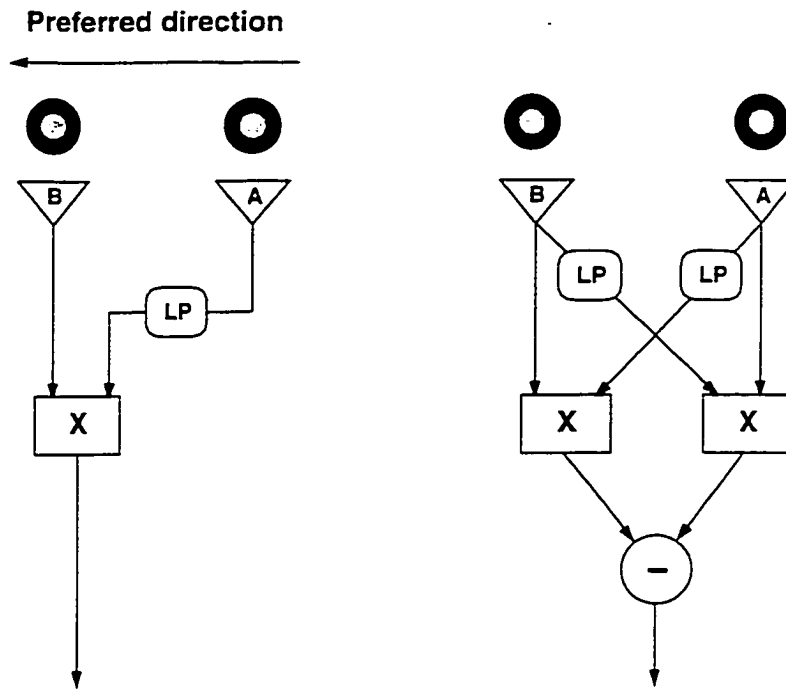
Figure 1.1: The Reichardt Model. A. A Subunit. The two receptors A and B are indicated below their receptive fields. LP refers to the low-pass temporal filter. X to the correlator. The correlator fires only if stimulus motion is along the preferred motion indicated at the top. B. A pair of complementary Reichardt subunits can be combined with a subtractive unit (-) to yield a detector whose output models the fly optomotor response.

of two luminance contrast detectors (schematized as A and B in Figure 1.1) and a correlator. The first detector is connected directly to the correlator, while the other is connected to it *via* a low-pass temporal filter.

When the detector is exposed to a stimulus at its receptor A at a time t, and then at its receptor B at a time $t+\Delta t$ (where the vector from A to B defines the correlator's *preferred direction*), the signal resulting from the activation of A correlates well with that resulting from B's activation and the output of the multiplicator is high. The response strength depends on the degree of correlation, which in turn depends on the match between the interval Δt and the characteristics of the low-pass filter. If on the other hand the motion is in the "null" direction (from B to A), then the two

signals are poorly correlated and the detector is not strongly activated.

Each subunit just described is matched to an subunit tuned for the opposite direction of motion. This matched subunit has equivalent wiring but with the temporal filter between the correlator and detector B as opposed to detector A. The output of the two subunits is then subtracted, yielding an output whose sign (positive or negative) indicates the direction of motion. The Reichardt model is therefore inherently an "opponent" model of motion perception.

The unit as described fits all of the requirements set by the physiological data which Reichardt meant to model. The two point-like receptors define an axis of motion to which the detector as a whole is sensitive. Any component of motion perpendicular to it will be invisible to the detector, and the response indicates the direction of motion of the stimulus (with some caveats which will be discussed in detail later).

The Barlow-Levick Model

Barlow and Levick (1965) were studying a different system, the rabbit retina, where they too found cells which responded differentially to motion in one direction as opposed to another. Their model, while based on a very different preparation, and expressed in slightly different terms, is functionally similar to the Reichardt model. Indeed, the schema they proposed contains two receptors, one directly wired to a comparator, and one relayed to it *via* a delay line. The comparator is specified as an instantiation of the logic gate AND-NOT (see fig 1.2). In the Barlow and Levick model, the comparator fires in the presence of a change in contrast over receptor A *unless* there was a change in contrast over receptor B at a time Δt earlier. Although this might seem a somewhat complicated wiring diagram, it accounts for the fact that the cells they recorded from. unlike the *Drosophila* cells, fired when presented with static flashing stimuli as well as with stimuli moving in their preferred direction.

6

If both detectors A and B receive simultaneous inputs. the "veto" action of the B pathway will not arrive at the AND-NOT gate in time to inhibit the excitatory action of the A pathway. This detector will then be activated by flashing stimuli. just as the rabbit retinal cells are (Barlow & Levick. 1965).

Preferred direction
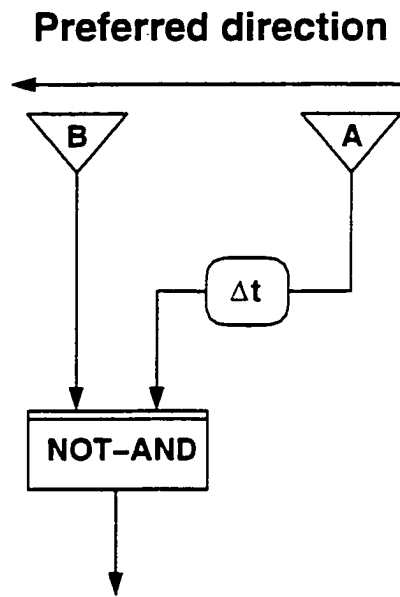
Figure 1.2: The Barlow-Levick model of direction selectivity (adapted from Barlow and Levick (1965))

1.1.2 The Elaborated Reichardt Detectors

A feature of both the Reichardt and Barlow-Levick model is that they predict that under certain circumstances, presenting periodic patterns on the receptors will cause maximal firing indicating the direction *opposite* to the real direction of motion— this is a form of aliasing, and like all aliasings, is due to the fact that there is a discrete sample (in this case a spatial one). This surprising prediction has been shown to be correct for fly eyes, but it is not valid in the case of humans—there is no speed at which continuously moving stimuli will be perceived by humans as going

7

in the opposite direction (Hildreth & Koch. 1987)[3]. The basic correlator model just described can be adjusted so as to eliminate this aliasing effect by modifying the punctate nature of the receptive field, and instead use receptive fields of finite extent, as done for example by Fermi and Reichardt (1963), Götz (1965), Van Santen and Sperling (1984. 1985).

This "elaborated Reichardt detector" can be shown to be equivalent two other motion detector models, which come from a different computational heritage, the motion detector by Watson and Ahumada (1985), and the motion-energy model of Adelson and Bergen (1985)[4]. Both of these models (which will be discussed in detail in Chapter 3) are based on the principles of linear systems analysis. Linear systems are used throughout this dissertation, so a brief introduction will be given, to setup some terminology and basic results.

1.2   Linear Systems

The general hypothesis upon which the use of linear systems in visual modeling is based is that the behavior of some cells in the visual system can be modeled to a first approximation as a linear filter with both spatial and temporal dimensions, so that the response of a unit early in the visual processing stream can be modeled as the convolution of the input (in space-time) with a linear filter which has a spatial component (typically a bounded. oriented, bandpass function such as a Gabor function), and a temporal component (which empirical results argue should be lowpass or bandpass). This convolution by filtering operators can be thought of as a weighted

---

[3]The only time when this occurs is when the stimulus is stroboscopic (i.e. sampled), as in a projected cinematographic movie or under artificial pulsed lighting.

[4]It should be noted that this computational equivalence is true only regarding the time-averaged responses—the dynamics of the detector responses can be quite different.

sum of the stimulus contrast over the spatial neighborhood of the receptor and recent history. Each cell, therefore, performs a spatially and temporally localized filtering operation on the stimulus. Because of its spatial extent and temporal integration property, a unit's response is spatially and temporally imprecise—any given filter response can be due to stimulus contrast at one of many location within the filter's spatiotemporal "receptive field". These filters are assumed to cover the entire visual field with overlapping spatial receptive fields, and have spatial and temporal tuning functions which span the range of visibility. Furthermore, within the scope of this dissertation, the output of these filters is assumed to be the *sole input* to the visual system[5]. The process of speed perception is thus viewed as the analysis of these filter outputs by the rest of the visual system so as to extract local image speed.

1.2.1 Why linear systems?

The power of linear systems analysis lies, not surprisingly, in the properties of linearity. A system is said to be linear if its response to compound inputs is additive. If a linear system has a response R_A to an input A, and a response R_B to an input B, then it must be true that the system's response to the compound stimulus $A + B$ will be $R_{A+B} = R_A + R_B$. While this property may seem at first trivial, it has a powerful consequence when combined with Fourier analysis. Fourier analysis is a procedure by which any pattern can be decomposed into the sum of a set of sinusoidal patterns. The response of a linear system to a pattern can thus be exactly computed from the decomposition of the pattern and the responses of the system to the component sinewaves. In general, if one knows the response of a linear system to a set of basis functions (e.g. all sinewaves), one can predict the response of

---

[5]Extra-retinal signals such as eye movement information are ignored, and all the experiments described throughout this dissertation are designed to minimize both their presence and their use in processing.

9

that system to any pattern. Thus the system is fully characterized by its response to sinewave patterns. While this result is exactly true only for periodic patterns over infinite extent and fully linear systems, predictions using this method degrade gracefully as the assumptions of infinite extent and full linearity are replaced with the realities of finite receptive fields and slight non-linearities. Furthermore, physiological data (e.g. Movshon, Thompson, & Tolhurst, 1978) have shown that so-called simple cells in visual cortex are to a first order linear, making linear systems analysis appropriate for the study of early visual processing. Finally, as will be reviewed in the next chapter, many psychophysical data have been usefully analyzed within the framework of linear systems.

It should be noted that the property of linear addition applies in the temporal domain as well, so that the response of a system is the convolution in time of the temporal filter with the temporal history of the input. This point is difficult to illustrate in general but will be made clearer within the context of motion processing below.

1.2.2 What do these filters look like?

Linear systems theory is a general theory, independent of any specific domain of application. The filters used to model the early stages of visual processing, however, are quite constrained by the data. The neuronal stage of processing which is widely considered to be the input stage for speed perception is visual cortical area 17 (also called V1, or visual area 1 (Zeki, 1969)). Cells in this area receive input from lateral geniculate (LGN) cells which themselves receive inputs from the retinal cells. While considerable processing occurs before V1, in humans this processing appears not to be motion-specific. Thus within the context of this dissertation, V1 cells are considered to be the first stage of processing.

10

This section will present a brief review of the characteristics of these cells, both in the spatial and temporal domain. Along the way, graphical representations of the cellular or behavioral filtering characteristics will be presented to clarify the exposition.

Spatial Profiles and Tuning Functions

V1 receptive field sizes depend on the eccentricity of the receptive field for that cell (the position in the visual field to which the cell is sensitive, relative to the point of fixation, equivalent to the position of the cell's retinal afferents relative to the fovea). Because of its central role in human vision, almost all of the work reported in this dissertation focuses on central, or foveal vision. Foveal cortical cells have receptive fields which have diameters on the order of 0.1 to 1 deg (Dao, 1994).

V1 cells are known to have two dimensional spatial receptive fields and most have orientation tuning (Hubel & Wiesel, 1962). In order to apply the linear systems framework just described, one needs to know the cells' spatial frequency tuning— that is, what is the amplitude of the cell's response after presentation of a sinewave pattern as a function of the grating's spatial frequency[6]. Once we know how a cell responds to sinewave gratings (its amplitude and phase lag), if it is linear, one can in theory compute its response to any pattern. Given the orientation tuning just mentioned and the fact that the cell response depends on the temporal history of the stimulus, this tuning cannot be measured in isolation. However, it can be measured under conditions which minimize or control for the effect of these other variables.

---

[6]Gratings are two-dimensional patterns modulated in one direction by a sinusoid and not modulated in the orthogonal direction. They have a spatial frequency, corresponding to the period of the sinusoid, and expressed in cycles per degree of visual angle (cpd). If a grating is moving, it is defined by its velocity along the axis of modulation, expressed in degrees of visual angle per second (°/sec), or by the rate of modulation at a point in the visual field (temporal frequency), expressed in Hertz (Hz).

The spatial frequency tuning of cells is generally measured for gratings which have the cell's preferred orientation (that which yields the largest response), and for fixed stimulus durations.

The field of spatial vision is too broad to review here, but a general feel for the spatial tuning functions and the corresponding filter shapes is easy to illustrate. Figure 1.3 shows a typical spatial frequency tuning curve for a set of six striate cells (DeValois, Albrecht, & Thorell, 1982). It shows that the cells responds best to stimuli with spatial frequencies in an intermediate range, and respond poorly or not at all to either high or low spatial frequencies. This type of tuning function can be used (along with phase information) to derive a filter shape which will yield this kind of behavior. Such a filter must clearly be bandpass in SF. Various proposals have been made for mathematical models of this filter, but a commonly used one is the Gabor function. A Gabor is a sinewave modulated by a Gaussian, as illustrated in the bottom panel of Figure 1.4. The sinewave is the most important factor in determining what the peak spatial frequency of the cell will be, while the Gaussian falloff determines the bandwidth—an infinitely wide Gaussian would produce a perfect sinewave detector, with infinitely narrow tuning, while an impulse function envelope would produce a broadly tuned analyzer.

The Gabor just described is a one-dimensional function. Receptive fields are two-dimensional and orientation-tuned. This is usually modeled by modulating the sinewave carrier with a two-dimensional Gaussian. The orientation of the sinewave function yields the preferred orientation of the modeled receptive field, while the falloff of the Gaussian determines the receptive field size (see Figure 1.5).

Temporal Profiles and Tuning Functions

What are the temporal properties of V1 cells? Given that the general topic under study is speed perception, it should be clear that the answer to this question will

12
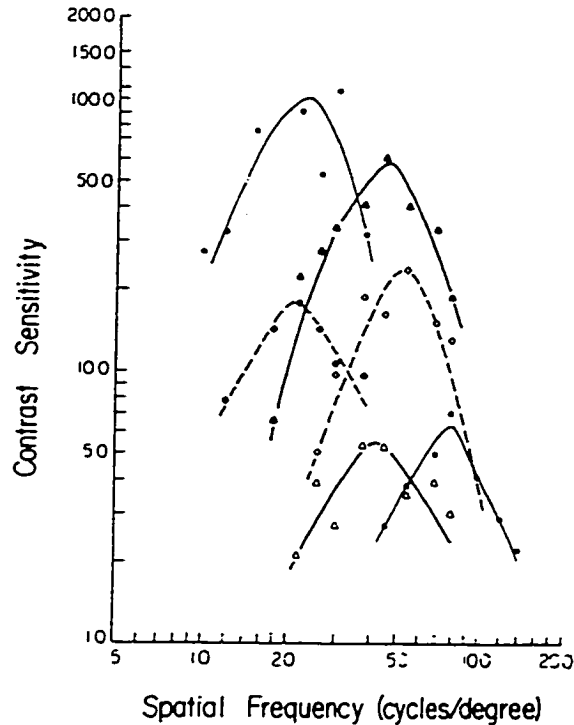
Figure 1.3: Spatial frequency tuning of six striate cells recorded during the same electrode penetration. Smbols indicate the contrast sensitivity of each cell (the reciprocal of the contrast required to reach a constant response criterion) as a function of spatial frequency; the curves were fitted by eye. Note the variation in peak tuning, bandwidth and sensitivity for this sample of cells (all of which pick-up from the same retinal locus) (DeValois et al., 1982).

have profound impact on how speed can be computed from the output of these cells. One method for obtaining cellular tuning functions is to measure the reponse of a cell to a stimulus to which it is sensitive (e.g. using its preferred spatial frequency and orientation), and modulating the stimulus with a sinusoidal temporal modulation function (thus creating a counterphase grating), recording the cellular response as a function of the modulation frequency. As will be reviewed in detail in the next chapter, both the physiological and psychophysical data seem to classify the temporal responses of V1 cells as one of two kinds, low-pass and band-pass, as illustrated in Figure 1.6. A low-pass cell shows a temporal tuning function which has a high

13
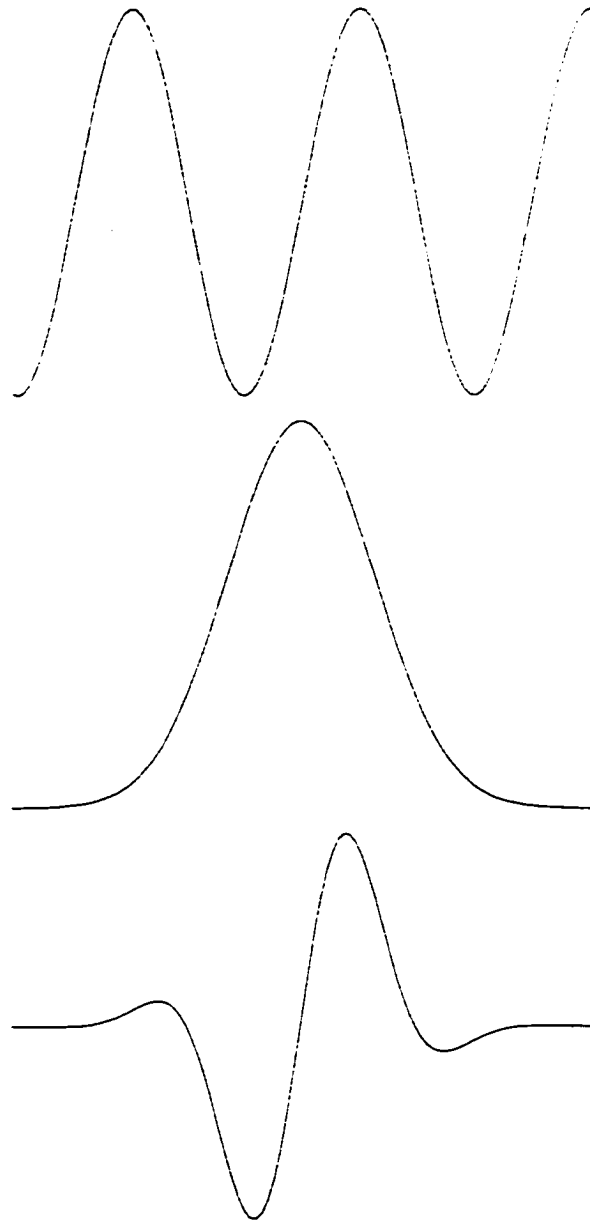
Figure 1.4: Explanation of a Gabor in terms of the carrier sinewave and the Gaussian envelope. The top panel depicts a sinewave function. The middle panel depicts a Gaussian function. The third panel depicts the convolution of these two functions, the Gabor.

14
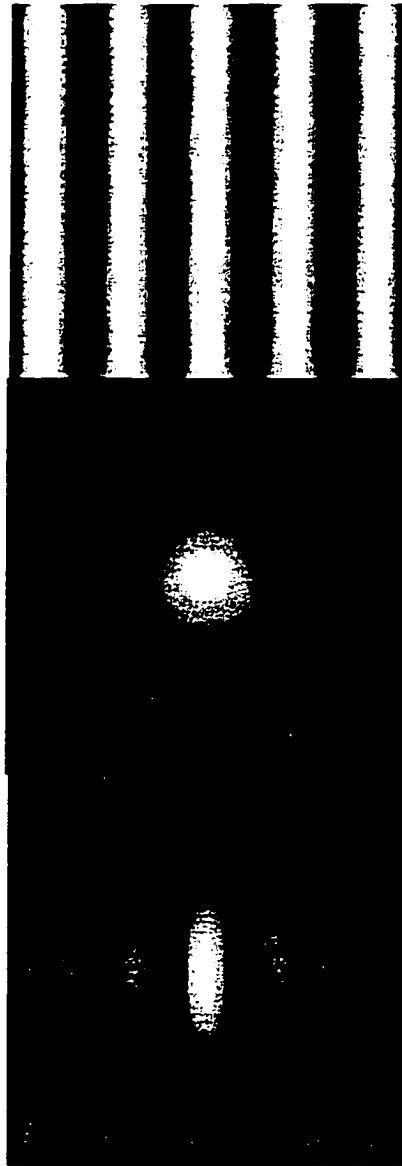
Figure 1.5: An oriented bandpass filter as a two-dimensional Gabor: The top panel presents a two-dimensional sinewave, the middle panel a two-dimensional Gaussian, and the bottom panel a two-dimensional Gabor. The orientation of the sinewave defines the orientation of the filter.

15

temporal frequency cutoff (the dynamics of the cells are not fast enough to keep up with very rapidly changing stimuli, so very high temporal frequency stimuli are equivalent to mean-field (blank) stimuli), and the sensitivity of the cell increases as the temporal frequency is decreased, and stays constant down to stationary (0 Hz) stimuli. This means that these cells give *sustained* responses to the onset of stationary stimuli, explaining why these cells are also called "sustained" cells. Band-pass cells, on the other hand, have tuning functions which are more similar in general shape to the spatial frequency tuning. Figure 1.6 shows tuning curves from three V1 neurons, including lowpass and bandpass types.

A note should be made regarding the use of the terminology lowpass and bandpass. In the physical sciences, a lowpass filter is one which transmits a non-null proportion of any DC input. In the visual domain, this would correspond to a mechanism or cell which responded to a completely stationary stimulus. Due to retinal adaptation, stabilized images are invisible to the post-retinal processing stages. It is therefore impossible to test (at least *in vivo* the formal definition of lowpass behavior for a visual cell. The term lowpass will therefore be used to refer to cells or mechanisms who give significant response down to very low temporal frequencies, such as the lowest ones which a subject viewing the world might see, given normal eye drift behavior. Similarly, bandpass filters are formally defined as yielding no response outside of a well-defined range. As is clear from the tuning functions in Figure 1.6, within the domain of visual motion, bandpass cells merely have a preferred temporal frequency range, and no absolute cutoff.

1.2.3 What are these filters good for?

So far, the spatial and temporal filters have been viewed as independent characteristics of V1 cells. It is important to remember that each of these is but an aspect of the
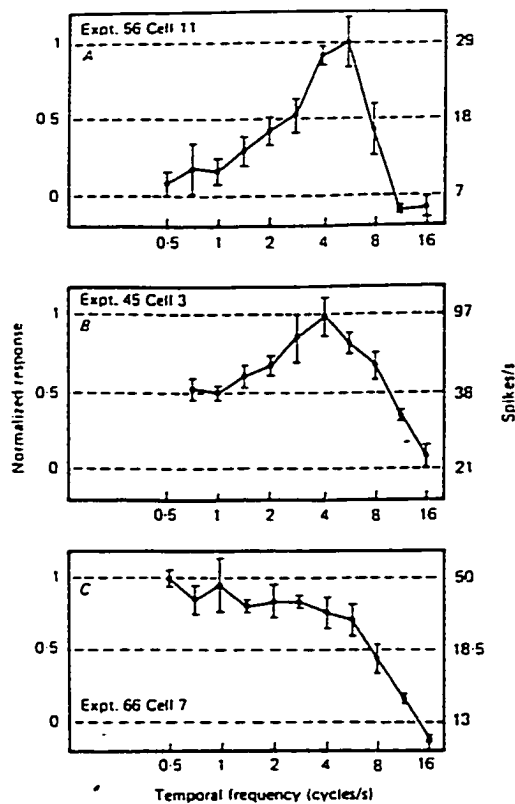
Figure 1.6: Temporal frequency tuning curves obtained from three neurones in Macaque visual cortex (reproduced from Foster et al. (1985)). The top two would be called band-pass using our terminology, while the bottom one would be called low-pass.

behavior of the entire filtering operation. In fact, cells are tuned to a variety of dimensions simultaneously, including spatial and temporal frequency, spatial position, and color. The complete tuning function of a cell is therefore at least six-dimensional. For the sake of speed perception, however, only two dimensions are needed: spatial frequency tuning along the preferred orientation of the cell, and temporal tuning. The filter sensitivity of the cell can therefore be efficiently represented in a space-time diagram, as in Figure 1.7, where spatial position is along the horizontal axis, time is down the vertical axis, and the grey-level indicates the response of the filter to that space-time coordinate, with dark greys representing inhibitory responses, and light greys representing excitatory responses.

17
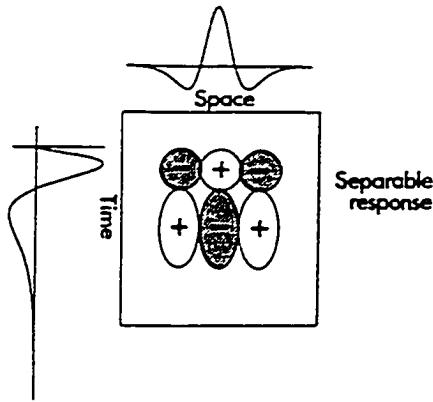
Figure 1.7: A separable spatiotemporal impulse response function. Reproduced from (Adelson & Bergen, 1985).

The unit just described has what is called a spatiotemporally separable impulse response. Its response can be exactly characterized as the product of a spatial function with a temporal function. There is no interaction between these two dimensions. While it is clearly easy to model, it suffers a significant flaw when it comes to motion perception: its response to two gratings moving in opposite directions at the same speed will be identical. Thus it is incapable of determining the direction of stimulus motion. As will be seen in the next chapter, the standard solution has been to use combinations of these simple units to build more complex units which are *not* spatiotemporally separable, and can thus be motion sensors. What about speed, the topic of this dissertation? We know that the basic units just sketched are spatially bandpass, and temporally tuned as well. Thus the unit responses will depend on the stimulus' spatial and temporal spectrum. If a stimulus spectrum overlaps almost not at all with a unit's tuning functions in either dimension, the unit will not respond to that stimulus. If the two are well matched, however, the unit will respond. It should be clear then that the strength of the response is directly related to, among other things, the speed of the stimulus (since given a stimulus with a fixed spatial structure, the temporal spectrum will depend only on the stimulus speed). This dependence is not simple. however. Response strength depends on the stimulus' "physiological

18

spectrum"[7], which varies with such factors as contrast and color, which have no relation to speed. It depends also on the filters' tuning characteristics, which vary across cells and which are not directly observable. Understanding speed perception is thus understanding what the nature of the underlying filters is, and how the properties of these filters can affect perceived speed.

1.3 Outline of the thesis

This chapter outlined the fundamental theoretical assumptions regarding how visual signals are processed by the first stages of the motion processing system, and sketched the general framework for speed perception based on the output of this first stage of filters. Two general aspects of this process are yet unspecified. The first is the characteristics of the filters, both their number (along both the spatial and temporal dimensions), and their tuning properties. These tuning properties include the receptive field size, the temporal integration window, and their spatial and temporal frequency tunings. The next chapter will present a review of the existing psychophysical literature which bears on these questions. The second general area which needs to be specified is how, from the output of these filters, the visual system computes retinal speed. Chapter 3 will present the existing models of speed perception, and compare them. The next five chapters present results from experiments which tested various aspects of speed perception. The first experiment concerns the requirement, stated implicitely at least by most models of discrimination, that speed discrimination is affected by contrast levels, a requirement which previous researchers have found not to be met. The next two experimental chapters report studies of the spatial and temporal integration properties of the speed mechanism respectively. Chapter 7

---

[7]By *physiological spectrum* is meant the physical spectrum of the stimulus, filtered by the physiological optics and photoreceptor tuning functions.

19

reports the results of an experiment which tested the influence on perceived speed of selective adaptation of the temporal low-pass mechanisms, which is predicted by many models of speed perception. The last experimental chapter, Chapter 8 reports on experiments focussing on the ability of the visual system to use information from a variety of spatial scales in computing speed. Chapter 9 summarizes the new requirements which a speed model needs to account to be comprehensive in its coverage of the data and presents a modified model which aims for a somewhat broader scope than the previous models. Finally, the Conclusion presents a summary of the results presented in the dissertation and some closing comments.

Before starting on the main content, a note on terminology. In this chapter, the words filters, channels, cells and mechanisms have been used interchangeably. Each has a specific meaning, grounded mostly in the framework of different disciplines (signal processing, psychophysics, physiology). *Filters* refer to "black boxes" which take a one-dimensional input signal and produce an output which is a scaled version of the output, with the scaling a well-specified function of the properties of the input. *Cells* refer to neurons, whose behavior one can only estimate within the context of a living brain, exposed to specific visual stimuli. That the neuronal firing rate after presentation of visual stimuli can be viewed as the response of a spatiotemporal filter is the linking step which bridges psychophysics and physiology. *Channels* are theoretical constructs from psychophysics, best likened to sets of units which share a tuning property and tile the visual field—thus one may refer to two temporal channels, and six spatial channels. What is meant by this is that at each retinal location, the visual input is processed in parallel by units whose temporal tuning functions can be well categorized as fitting one of two profiles, and whose spatial tuning functions can be well-categorized as fitting one of six profiles. In the following chapters, starting with the reviews of psychophysical data and of the models, the terms appropriate to each study will be used. The fact that so many words mean

more or less the same thing is an encouraging sign. as it seems to indicate that the various methods of study of visual processing must be dealing with the same "things."

Chapter 2

Speed Perception and

Spatiotemporal Channels

This chapter presents a review of the psychophysical and physiological data which characterize the filters described in the Introduction. The little reliable data available on speed perception *per se* will also be reviewed. The former set of data describes the building blocks of the computational models of speed perception, while the latter define the expected behavior of those models. Detailed presentation of the speed models is deferred until the next chapter.

As described above, speed perception is assumed to be recognition of patterns of of activity of an array of spatiotemporal filters. Given the historical lack of interest in speed perception, it is helpful to point out that these filters are also believed to underlie other low-level visual tasks, such as static spatial vision and other aspects of motion perception. Thus the data on the spatial properties of the filters comes from the literature on spatial pattern detection and discrimination, and the temporal tuning functions come from speed and other temporal processing experiments.

This said, it is important to remember when reading about the experiments described in this chapter that in most cases the results obtained may or may not

apply to the speed mechanism. Indeed, the psychophysical characteristics of the speed perception process may or may not be the same as those of, e.g., the direction perception process, studies of which are used to infer filter characteristics. The extent to which the speed mechanism shares characteristics with the other early visual perceptual mechanisms is the subject of some of the experiments described in the later chapters.

2.1 Spatial Characteristics of the filters

The review of the filters used in speed discrimination will start with their spatial characteristics. These include number, receptive field size, orientation tuning, and spatial frequency tuning.

2.1.1 Receptive Field Size

Estimates of the spatial extent of the visual input processed by an individual analyzer come from both physiology and psychophysics.

Physiological data is the simpler to analyze in this case, since it is obtained from single-neuron recordings. In the landmark study of Hubel and Wiesel (1959, 1962), the response of a neuron is measured for stimuli of increasing sizes. The smallest size which elicits the largest response of the neuron is defined as the receptive field size for that neuron (at least cells which have no inhibitory surround). Recently, this simple picture of the receptive field of cells has been challenged by physiological recordings finding that the behavior of cells can be influenced by stimuli well outside the "classically defined receptive field" (Gilbert & Wiesel, 1990; Gilbert, 1993; Das & Gilbert, 1995; Zipser, Lamme, & Schiller, 1996).

Psychophysical data has been obtained both using stationary and moving displays. One technique used consists of measuring the detectability of stimuli as a

function of their size. As the size of the stimulus grows, the detectability increases. A key aspect of this relationship however is that this increase in detectability is at first relatively rapid, and slows down for larger stimulus sizes. This is interpreted to mean that during the early phase of the increase in detectability, the signal is integrated both *within* the detecting receptors, using very efficient "synaptic" or "physiological" summation, and *across* receptors, using the less efficient, inter-receptor probability summation[1] (Robson & Graham, 1981; Olzak & Thomas, 1986; Anderson & Burr, 1987). Once the stimulus size exceeds the receptor size, within-receptor integration yields no improvement, and the integration which remains is probability summation. The point at which the rate of integration changes can be viewed as one definition of the receptive field size[2]. The size estimates for receptive fields vary drastically with eccentricity (e.g. vanEssen, Newsome, & Maunsell, 1984), but this dissertation limits its scope to foveal vision, so these changes will not be described. A further source of change in receptive field size is the stimulus. While the earlier psychophysical work on estimating filter sizes used filled spots of light (Blackwell, 1946), later work using sinewave gratings showed that receptive field size depends on the stimulus spatial frequency (McCann, Savoy, & Hall, 1973; Hoekstra, Goot, Brink, & Bilsen, 1974; Howell & Hess, 1978; Quinn & Lehmkuhle, 1983; Tootle & Berkley, 1983), with most estimates of the receptive field size at about 1.5–2.5 cycles (thus estimates for a 1 cpd grating correspond to 1.5–2.5 degrees of visual angle).

Anderson and Burr (1987) pointed out that this estimate would indicate that gratings of 0.025 cpd (which are visible) would be detected by units with receptive

---

[1]Probability summation is less efficient than physiological summation because a given receptor output is necessarily noisier than the component signals within a receptor.

[2]Estimates of receptive field sizes depend on the stimulus used, on the recording technique, and in some cases on the configuration of the stimulus over a large area of visual space.

fields of 60–100 deg. far exceeding the largest field sizes recorded electrophysiologi-cally. Using moving (8 Hz) gratings at a variety of spatial frequencies (from 0.01 cpd to 30 cpd), they performed a motion detection task, measuring the minimum contrast needed for detection as a function of the stimulus size. By plotting the obtained contrast sensitivity as a function of stimulus area, they were able to record the size at which the integration exponent changes from indicating within-receptor integration to cross-receptor integration. Their results are in agreement with the previous data, except that for low spatial frequency stimuli, the receptive field size corresponds to smaller fractions of the stimulus SF (down to approximately 0.07 cycles for 0.01 Hz stimuli), in good correspondence with the maximum receptive field sizes reported physiologically (Dao, 1994). To summarize, psychophysical experiments appear to yield estimates of receptive field sizes which are on the upper end of those measured physiologically, since V1 cells in the fovea have receptive field sizes on the order of a fraction of a degree of visual angle, and they have approximately circular shapes.

2.1.2 Orientation Tuning

While there is considerable data available on orientation tuning of motion sensors, it can be summarized for the purposes of speed perception as follows. Physiological data indicate that motion-sensitive units in the first visual cortical area are orientation-tuned, and that orientations are roughly equally sampled. Computational models of spatial vision such as Wilson, McFarlane, and Phillips (1983), use orientation channels with bandwidths at half-height on the order of 30°, consistent with both the physiological data and psychophysical reports (Mostafavi & Sakrison, 1976; Thomas & Gille, 1979). Physiological evidence argues that the orientations are somewhat densely sampled, with neighboring cortical columns having preferred orientations separated by 5–10° (Hubel & Wiesel, 1974)

2.1.3 Spatial Frequency Tuning

An important aspect of spatial frequency channels in the context of speed perception, as will become clear in the discussion of the various computational models of speed perception (Chapter 3), is what the number of these channels is, and what their spatial frequency tuning curves are like. Luckily, there has been a considerable amount of work done on the processes responsible for of spatial pattern perception, and general consensus can be found regarding these two issues.

The number of filters which process a given stimulus is clearly large, given that at least a set of filters is necessary at each location, for each direction. As is usual, the number and shape of spatial channels is evaluated experimentally at a given orientation (e.g. vertical) and at a given location (usually the fovea), and then extrapolated from there to the entire visual field and to all orientations[3].

How are tuning properties of channels estimated? Reviewing this topic is best left to others (Olzak & Thomas, 1986; Watson, 1986), but a brief summary of the techniques and basic results is useful to understand the subsequent chapters. The discussion up to this point has focused on filters and channels as purely linear systems. Perception, however, is nonlinear, especially at low stimulus intensities. For very low stimulus intensities, the stimulus is invisible. For the stimulus to be detectable, it has to elicit a response in a channel which is higher than the threshold for that channel. What that threshold stimulus energy is for a given stimulus depends on the match between the tuning function of the channels and the stimulus—if a channel is very sensitive to that stimulus, a relatively low stimulus energy will result in detection. If all the channels are blind to that stimulus, no amount of stimulus energy will result in detection (e.g. ultraviolet light and the human visual system). Thus by measuring

---

[3]This extrapolation is inaccurate in the peripheral visual field, but as mentioned earlier, we are focusing on foveal vision.

the physical stimulus energy needed for detection. one can infer the tuning function of the set of channels as a whole[4]. Various techniques can then be used to further identify the channels which are responsible for this "envelope" sensitivity, as well as characterize their individual sensitivity curves. Foremost among these is the use of *discrimination thresholds*. Discrimination refers to the ability of an observer to reliably tell two stimuli apart. Assume that a given stimulus is at detection threshold, so that it "barely" excites one channel. For the subject to be able to distinguish between this stimulus and another at detection threshold. the two stimuli have to be exciting *different* channels. Thus by measuring the discrimination threshold for a stimulus as a function of the, e.g., spatial frequency of a second stimulus, one can partially determine the tuning functions of the (in this case) spatial frequency channels.

To review the terminology used: detection threshold is the physical stimulus intensity which results in reliable detection. Discrimination threshold is the minimal separation between two stimuli which affords reliable discrimination by an observer (detection and discrimination thresholds are quantified by an index of sensitivity d' index (Green & Swets, 1966)). Sensitivity refers to the relation between stimulus energy and detectability either by an organism as a whole, or by an individual channel. Threshold and sensitivity are inversely related—a high sensitivity corresponds to low thresholds, and vice versa.

Many methods have been used to identify channels and channel shapes. Two will be discussed briefly (adaptation and subthreshold summation), since they allow the

---

[4]It should be pointed out that the theoretical views of channels used here, while it is the standard psychophysical one, is not devoid of assumptions. Thus while channels can be related to cells, they should always be remembered as theoretical constructs, the properties of which are most probably only grossly mirrored by reality. For example, the properties of fixed thresholds and linear summation within a channel are likely not exact in either cells or behavior

27

introduction of concepts which will be useful in understanding the later experimental chapters.

Adaptation

In an adaptation study, the sensitivity function is measured as described above to establish a reference, against which the effect of diminishing the output of a small number of channels is measured. This diminution in channel output is obtained by having the observer view a given stimulus (e.g. a 3 cpd grating) for a relatively long period of time before measuring the sensitivity function anew. When such an adaptation is performed, the sensitivity function generally shows a dip for stimuli identical or similar to the adapting stimulus. The physiological mechanism underlying this loss in sensitivity is still ill-specified, but it is an empirical fact that channel responses decrease after their adaptation. This is useful because adaptation of a channel affects the entire sensitivity range of the channel (thanks to the principle of univariance (Naka & Rushton, 1966)), not just for stimuli whose characteristics match those of the adapting stimulus exactly. Thus, if a channel is very broadly tuned and is the most sensitive channel over a large part of its bandwith, then adaptation of that channel will affect the organism's sensitivity over a large range. If the channel is either narrowly tuned or if it is not the most sensitive over a large range, then detection sensitivity will only be affected over a small range of stimulus parameters.

Subthreshold summation

Summation studies also use use the principle of univariance of psychophysical channels. The output of a channel can be thought of as proportional to the number of stimulus energy quanta which have been absorbed by the channel, with the channel sensitivity defining the relationship between the nature of the stimulus energy and

28

the likelihood of quantum absorption by the channel. Channel output only yields a response, however, if it is above a threshold (which corresponds to a certain number of quanta). Thus whether two spatial frequencies are processed by the same channel can be tested by measuring the detection threshold for each spatial frequency, and then testing whether the presence of one spatial frequency at subthreshold contrasts affects the detectability at the second spatial frequency. If it does, then both stimuli are processed by the same channel (as the quantal numbers add within the channel). The amount of subthreshold summation between one grating and another is a further indication of the relative sensitivity of that channel to those frequencies. Subthreshold summation does not occur across channels because the output of each subthreshold channel is indistinguishable from noise.

Summary of Spatial Frequency Filter Shapes

Methods such as those just described as well as other methods including masking (Stromeyer & Julesz, 1972; Legge & Foley, 1980), two-pulse summation and others (see review in Olzak & Thomas, 1986) have found that spatial frequency channels are bandpass, with bandwidth at half-height of between 1 and 2 octaves. To tile the spatial frequency contrast sensitivity function thus requires a fairly large number of channels, with for example the models of Wilson et al. using *at least* 6 channels, spaced approximately one octave apart (more channels might exist—psychophysical techniques only allow the setting of a lower bound).

2.1.4 Summary of Spatial Filters

The various techniques above have mostly converged on a fairly noncontroversial view of analysis by spatial filters. These filters have receptive fields on the order of a fraction of a degree in diameter depending on the SF tuning, are orientation tuned

with orientation bandwidths of about 30°. and have narrow spatial frequency tuning. thus tiling the fovea, the 360° of orientations and spatial frequency space. Spatial frequencies up to 60 cpd are processed by a set of mechanisms with fairly constant bandwidths (proportionally), with a half-height bandwidth estimated to be between 1 and 2 octaves depending on the study (Olzak & Thomas. 1986). At least six such channels are needed to account for the available data. Physiological evidence, not surprisingly, does not show segregation of best spatial frequency into any discrete set of clusters.

2.2 Temporal Characteristics of the filters

As can be seen in the preceding section, the spatial channels are fairly well characterized by a relatively large number of narrowly tuned bandpass filters[5]. The study of the temporal channels is much less simple to summarize, because of fundamental reasons having to do with the shape of the sensitivity functions (as will become clear), because of methodological problems which will be described in detail, and because of the technical difficulties associated with carefully controlling moving stimuli with display technology with much more "information bandwidth" in the spatial domain than in the temporal domain.

2.2.1 Filter number and shapes

While in the spatial domain, adaptation to a given spatial frequency results in changes in sensitivity for a restricted range of spatial frequencies at and around

---

[5]One should remember that this characterization is mostly based on the use of foveal, stationary, luminance-defined displays under photopic lighting conditions, and that generalizing beyond this domain is fraught with risks. Luckily such a generalization is not needed for this dissertation.

the adapting frequency (Blakemore & Campbell. 1969). the results in the temporal domain indicate much broader bandwiths. The range of temporal frequencies to which humans are sensitive depends on aspects of the stimulus such as contrast and size, but extends from stationary stimuli up to 50–70 Hz. It should be noted again that absolutely stationary stimuli (for example. as obtained by using optical apparatus affixed to the cornea which stabilizes the stimulus on the retina) will quickly result in large decreases in sensitivity (Kelly. 1979; Tulunay-Keesey & Jones. 1980) The first studies explicitely testing the number and bandwidth of the temporal channels (Smith, 1970, 1971) report that after adaptation to sinusoidal full-field flicker, the temporal contrast threshold is elevated for a broad range of test temporal frequencies. Since then, many different psychophysical tasks have been used to determine the minimal number of channels required to explain human performance, and their respective sensitivity functions. Additional adaptation studies (Smith & Edgar, 1994), masking studies (Pelli, 1981), threshold discrimination experiments (Watson & Robson, 1981) and others (see a review in Watson, 1986) seem to converge on requiring between two and four channels to fit the data. When two channels are invoked (e.g. as in Smith & Edgar. 1994), one has a relatively low-pass tuning function, and one has a more band-pass tuning function. It should be noted however that these two functions must necessarily overlap within the ranges of temporal frequencies at which TF (or speed) discrimination is possible, since discrimination requires the output of two mechanisms at least. When more channels are invoked (e.g. Mandler & Makous, 1984; Wilson & Gelb, 1984; Lehky, 1985; Bowne, McKee, & Levi, Personal Communication), the additional channel(s) are proposed to be band-pass, providing sensitivity especially in the high TF range. Finally, Jackson and Welch (personal communication) report evidence for three or maybe four channels using masking speed discrimination experiments. Examples of fairly typical filter shapes are depicted in Figure 2.1

31
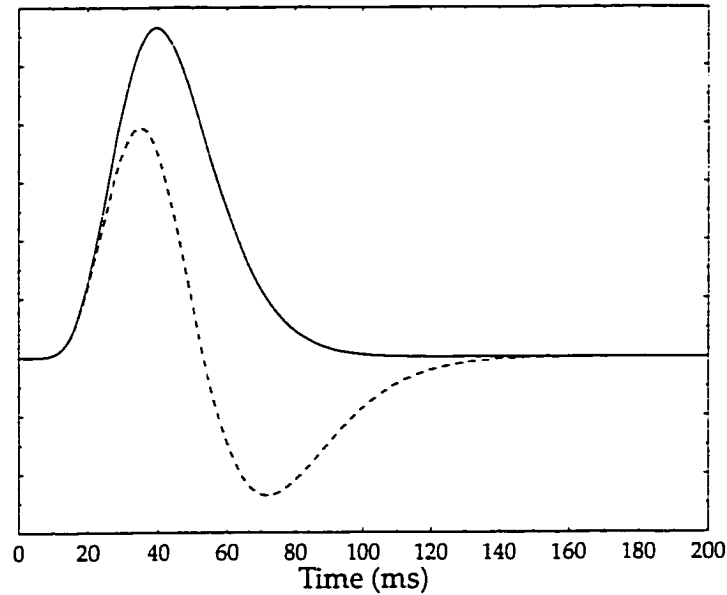
Figure 2.1: The Alpha functions h_0 (solid line) and h_1 (dashed line) used by Watson (1986).

2.2.2 Temporal Integration Window

An additional measure of the temporal properties of the filters is what can be called the temporal integration window. It is the temporal equivalent of the spatial receptive field, the minimal duration which must separate two events so that the response of the filters to the second event is independent of the filter response to the first event. When two events (stimuli) are closer together in time than this minimal separation, the two events *interact* within the filter. The linear nature of the putative filters (as discussed in the Introduction) means that the nature of this interaction is assumed to be strictly linear. Thus the total stimulus response is a linear combination of the responses to the two individual stimuli, weighted by the amplitude (positive or negative) of the temporal impulse response function. The temporal response function is,

32

by definition, the function which specifies the effect of a second impulse stimulus on the output of the filter which has been exposed to a first impulse stimulus at time $t = 0$. The region of time where the temporal impulse function is *positive* identifies times at which a second stimulus would result in an *increased* output compared to if the first stimulus had not been presented, while the region of time where the temporal impulse function is *negative* identifies times at which a second stimulus would result in a *decreased* response. Thus positive and negative lobes in a temporal impulse function are the temporal equivalent of the excitatory and inhibitory spatial subunits within a receptive field (see Figure 2.1). The two functions used by Watson (1986) extend over 90–140 msec respectively. Those used in Watson and Ahumada (1985) are somewhat briefer, extending only 80 msec.

2.3   Spatiotemporal Separability

The spatial and temporal filters have so far been presented as independent mechanisms. An important question regards the relationship between the spatial filtering operation and the temporal filtering operation. This question can be asked at two levels. At the first level, the question is whether the spatial and temporal functions are separable—whether the filter response is simply the convolution of the spatial filtering function by the temporal filtering function. As will be clear in the next chapter, most models of speed perception assume such a separability in their first stages of processing. The second question is whether the spatial and temporal frequency tuning functions of cells or of individuals are independent. A true "speed-tuned" mechanism would answer this question in the negative, since for such a cell the best SF would depend on the TF and on the cell's speed tuning. Obtaining psychophysical data on this topic is especially difficult, since, as mentioned earlier, to identify a temporal channel one must fix the spatial properties of the stimulus, and vice

33

versa. There is some evidence from the physiological literature that V1 cells have *independent* SF and TF tuning functions, however: Foster et al. (1985) and Holub and Morton-Gibson (1981) found that the cells they investigated had indepedendent SF and TF tunings in Macaque and cat respectively.

Recently, Schrater, Knill, and Simoncelli (1997) demonstrated using noise stimuli that in a translation detection task, observers were more sensitive to stimulus configurations which had stimulus energy aligned in a single velocity plane (in other words corresponding to a single speed) rather than stimulus configurations which had stimulus energy scrambled away from this plane (corresponding to several speeds). Schrater et al. (1997) concluded from these experiments that translation detection is performed by velocity-sensitive units, such as those described in Simoncelli and Heeger (1997). Related recent work from Reisbeck and Gegenfurtner (1997) appears to provide further evidence from the orientation of just-noticeable-difference shifts in spatiotemporal space for the existence of speed-tuned (as opposed to SF-tuned or TF-tuned) mechanisms in speed perception.

2.4 Speed Psychophysics

Having reviewed the status of spatial channels and temporal channels, one may wonder where hypothetical "speed channels" would fit with respect to these "orthogonal" dimensions of SF and TF. The relationship between temporal frequency discrimination and speed discrimination is subject to a great deal of debate, due to a set of apparently conflicting results regarding the effect of contrast on speed and temporal frequency perception. This section will review this literature and attempt to synthesize its main results.

The first result on the relationship between contrast and speed perception is the report by Thompson (1982) that grating contrast can have profound effects on the

34

perception of grating speed. Thompson measured. using a matching technique as well as a magnitude estimation technique, the perceived velocity of gratings of various contrasts relative to the velocity of a fixed-contrast grating. His displays used 1. 2. 4 and 8 cpd gratings, with TFs ranging from 1 to 16 Hz. The contrast range he covered spanned the 4.5%–25% range. His main conclusion was that low-contrast gratings generally appear to move slower than high-contrast gratings. This effect was greatest (40% slowdown in perceived speed) for slow gratings. and reversed at high TFs (8 Hz and above). Importantly, the effect is independent of grating SF. An important conclusion one must draw from this result is a methodological one— since apparent contrast affects perceived speed. it is important when testing speed discrimination that the apparent contrast of the stimulus not covary with speed. This is especially significant when testing at high TFs, since due to the falling contrast sensitivity function at high TFs, small changes in stimulus TF will yield large changes in apparent contrast, when the stimulus contrast is kept constant. We will return to this point later. Thompson (1982) explained his data by sketching out a preliminary two-channel model of velocity computation, where at low TFs, changes in stimulus contrast would affect the output of a "fast" channel more than that of a "slow" channel. thus yielding an decrease in apparent velocity. This sketchy model is not considered in greater detail because it does not make explicit the computational mechanism by which speed is computed. nor explain much data.

In apparent contradiction with this result is the important report by McKee, Silverman. and Nakayama (1986) which showed that observers are able to make precise speed discrimination even in conditions where the contrast, spatial frequency and temporal frequency were randomized across stimuli. McKee et al.'s data touch on two separate issues, which will be described in turn. The first, the major point of the paper, has to do with the preeminence of velocity coding over temporal frequency coding. The second has to do with the apparent conflict between the McKee et al.

35

data and the Thompson data.

2.4.1 The status of speed perception relative to that of temporal frequency perception

It is worth reviewing the experimental design and major results of the McKee et al. (1986) paper. In one experiment, observers performed a speed discrimination task with displays where the speed was chosen from one of five velocities around a mean velocity, and observers were to indicate whether a given stimulus was faster or slower than the mean speed of the stimuli seen so far. While the feedback to the user was based on the speed of the stimulus, the spatial frequency of the stimuli varied over a range of up to 1.6 octaves (with the temporal frequency varying over the corresponding range). The results indicate that the Weber fraction for speed discrimination was basically unaffected by the randomization of the spatial and temporal frequencies in the display. If the perception of speed were the result of a computation based on an first-stage estimate of temporal frequency, then the jittering of temporal frequency should have had drastic effects on discrimination performance, effects which were not found. Furthermore, the point of subjective equality (i.e. the criterion point used by the subjects) was not significantly affected by the spatial or temporal frequency variations, arguing again that the sensory variable which the subjects responded to was indeed velocity and not spatial or temporal frequency. In a control experiment, McKee et al. (1986) tested the ability of observers to make temporal frequency discriminations with the same stimuli (thus asking the observers to discount changes in speed). Interestingly, performance was worse than that for speed discrimination for one observer, but equal to that for speed discrimination for another. To determine whether the observers were using estimates of velocity to compute temporal frequency, they repeated the TF discrimination task with counterphase gratings, which

36

offer no velocity signal (due to the opponent nature of motion processing. (Levinson & Sekuler, 1975; Adelson & Bergen, 1985; Van Santen & Sperling, 1984)). In this condition, TF discrimination thresholds were significantly impaired. McKee et al. (1986) conclude from these data that:

> What our results show is that, for briefly presented targets, this temporal [frequency] signal is so tightly embedded in the neural machinery for velocity that it cannot be extracted from the velocity signal with the same precision as the velocity signal itself.

This first conclusion of McKee et al. (1986) has been challenged when extended to different stimulus characteristics. For example, Smith (1987) has used longer duration stimuli moving at a range of velocities (4 to 24 deg/s). His results indicate that unlike in the McKee et al. (1986) experiments, randomization of spatial frequency can have drastic effects of velocity discrimination. In a later experiment, Smith and Edgar (1990, 1991) have shown that not only can random changes in temporal frequency affect speed discrimination performance, but random changes in speed can also affect temporal frequency. Smith and Edgar's major conclusion is that this lack of separability (the fact that affecting one variable influences the perception of the other) argues that temporal frequency is available as a "raw" perceptual attribute, rather than derived from velocity estimates. In general, it appears that temporal frequency can be estimated without first estimating velocity, for example when using long duration stimuli, as McKee et al. (1986) acknowledge. In my own experience, the subjective impression of the task can shift dramatically depending on the stimulus configuration. Even with a constant duration, if the spatial extent of the stimulus is of the same spatial scale as the spatial period of the stimulus or smaller, then the percept is of "faster or slower flicker" rather than "faster or slower motion." This should not come as a surprise. Clearly, the speed of object motion is, like every

other aspect of the world, computable by a variety of algorithms and methods. each of which has a range of applicability. For very slowly changing stimuli, such as the position of the sun (or stars!) in the sky, motion sensors in the early visual pathway are not going to provide very reliable data, and cognitive strategies and reference to external timing devices will be more accurate. Without going to such extremes. it appears logical that the efficient computation of speed by low-level motion mechanisms is limited to a behaviorally relevant range, especially considering the fact that the same motion system is also responsible for other tasks such as direction discrimination, grouping and segmentation, etc.

Regardless of the "first-class" or "second-class" status of temporal frequency among perceptual attributes, McKee et al.'s data argue convincingly that, as stated in the Preface, under the conditions of interest here, *speed is a perceptually "real" attribute*, not derived by higher-level cognitive processes but "directly" available to the earliest stages of cortical processing. It is worth keeping in mind however that whatever conclusions are drawn regarding speed perception with these stimuli may or may not apply to other stimuli. For example, Seiffert and Cavanagh (1997) showed recently that the detection of motion of second-order displays (displays which have no net motion energy but still elicit a motion percept) is most likely based on a comparison of position information and temporal information rather than a direct motion signal.

2.4.2 Speed and Contrast

The second major point of the McKee et al. study has to do with the relationship between stimulus contrast and perveived speed. In an experiment analogous to the one described above, McKee et al. measured speed discrimination thresholds for a wide range of contrasts (from 6.4% to 82%). Their results show that speed

38

performance was independent of contrast. Importantly, discrimination ability was unaffected by within-block randomization of contrast. This is in apparent conflict with the results of Thompson (1982) described earlier. If the Thompson results had carried over, then one would have expected perceived speed to change over the range of contrasts used in the McKee et al. (1986) study, which should have made the speed discrimination task harder in the "mixed contrasts" condition. One possible reason behind the apparent mismatch between these two reports is that while Thompson showed the contrast effect with 2 cpd gratings with TFs between and 1 and 8 Hz, the gratings used in the mixed-contrast condition in McKee et al. (1986) were 1 cpd gratings moving at speeds of 10 or 15 dps, i.e. with TFs of 10 or 15 Hz. Thompson's data showed that the contrast dependency nulled between 8 and 16 Hz and reversed sign at 16 Hz. Thus one explanation for the discrepancy is that the TFs used by McKee et al. were high enough that the contrast dependency was not present. This may also explain the similar lack of contrast dependency found by Pantle (1978), but that author does not report enough details in the footnote which mentions the contrast results to be able to know for sure.

In my hands, as reported in Chapters 4 and 7, stimulus contrast affects both speed discrimination thresholds and perceived velocity. While the motivation for these experiments and the conclusions one can draw from their results are left to their respective chapters, it is appropriate to make a methodological point at this time. The studies discussed above consider the effect of changes in contrast on perceived speed or speed discrimination. The converse relationship is also true. That is, stimulus speed can affect perceived contrast. That this is true is trivial in the extreme. A very rapidly moving stimulus is blurred, and gratings for example become indistinguishable from blank fields if their temporal frequency exceeds the critical flicker fusion frequency, by definition. This can have profound consequences regarding the controls necessary for speed studies. At high temporal frequencies

especially, perceived contrast may be a more reliable cue to solving the experimental task than speed. Thus if the stimulus contrast is held constant while speed varies, the *perceived contrast* may be used by observers to perform the task under conditions where their response has no relation (or only a limited relation) to their motion system's performance. It is therefore important to control for this cue when doing speed discrimination and matching tasks. One technique is to jitter stimulus contrast. Another technique which can be used sometimes is to use a contrast compensation technique (Bowne et al., Personal Communication; Mandler, 1984), which explicitely compensates for the shift in perceived speed by adding a pedestal contrast.

As has been mentioned before, the speed of a sinewave grating (the most commonly used stimuli in the studies reported here) is the ratio of the grating's temporal frequency by its spatial frequency. Speed is also the distance traveled by the stimulus divided by the duration of the stimulus. What these relationships imply is that there are many correlates of speed, which can, depending on the experimental design and conditions, be more accessible to the subject than speed. Thus while the feedback to the subject is based on stimulus speed, in poorly designed studies subjects responding to one of these correlates will perform the task well, while not basing their behavior on perceived speed at all. For example, most studies of speed perception hold the spatial frequency of the stimuli constant within an experiment, and vary the speed of the stimulus by modifying its temporal frequency. As has been discussed in detail earlier, temporal frequency is a stimulus attribute which subjects are notably poor at discriminating. Furthermore, when trying to discriminate or match temporal frequencies, subjects often discriminate or match speeds. This, however, assumes that the cues of duration and spatial displacement are not available. Most studies using gratings use constant duration stimuli, and changes in speed (and TF) result in changes in the spatial extent traveled by any feature of the stimulus (e.g. a peak or a trough in a grating). The ability of humans to distinguish between distances is quite

40

accurate (e.g. McKee & Welch, 1989), and in conditions where speed estimation is hard, subjects may well switch strategies and use such positional cues to lower their error rates. A good control for this positional cue to speed is to randomize the duration of the stimuli, so that the positional cue is unreliable *on a trial-by-trial basis* as to the stimulus speed. The amount of randomization necessary for this technique to be effective depends on the threshold estimation procedure used. Further details will be provided in the experimental chapters when this duration jittering is used.

Alas, many studies of speed perception did not incorporate these controls, and thus are vulnerable to the criticism that the speed performance they report is in fact performance on a relative position task or a perceived contrast task.

2.5 Summary

Speed perception is a psychologically real percept, based on the output of early motion sensors. These motion sensors have narrow spatial tunings which are well characterized by six-channel models, and broad temporal tunings, requiring at least two and maybe up to four temporal channels. At least one temporal channel is low-pass, and the others are band-pass. Importantly, the spatial and temporal responses of the filters are independent of one another. As the next chapter will make clear, this property of the early filters is crucial in explaining why speed perception is complex enough that it cannot be understood in terms of the "simple" readout of a bank of filters, the way spatial frequency is often viewed. Briefly, if the output of cells is defined by both spatial and temporal frequency tuning, then a given stimulus velocity will result in the activation of mechanisms tuned to that stimulus' spatial and temporal frequency, but not to other mechanisms which, while tuned for the

41

same speed, have different spatial and temporal tunings[6].

This analysis of speed perception in terms of its *inputs* leaves a great deal to be desired. Indeed, so far very little has been said about *how* speed is computed, if a fair bit has been said about what the inputs to the speed computation are. This issue of the process by which speed is computed is the topic of the next chapter, which reviews the existing models of speed perception. Once the models have been presented, experimental questions raised by their designs will be raised, and some of them will be addressed by the later chapters.

---

[6]While there is a preliminary report in the literature of true "velocity-tuned" cells in area MT (Newsome, Gizzi, & Movshon, 1983), this result has apparently not been replicated, although efforts are underway to do so (Simoncelli, 1997, personal communication).

Chapter 3

Computational Models of Speed Perception

This chapter will review the existing models of speed perception, compare their coverage of the data, and outline some of the testable predictions they make.

There are three broad classes of models in the literature: models of motion perception which address the topic of speed computation explicitely; models of motion perception which do not mention the topic of speed perception, and models of motion perception which, while they touch on speed computation, do so from the perspective of computer vision and not of human vision. Models from the first group will, naturally, be the main concern of this chapter. Models from the second group will be mentioned only in as much as they influenced the development of current thought on motion sensing, and will not be reviewed extensively. The last group of models will not be covered, since their requirements are driven by engineering concerns, and comparing them to psychophysical data would be inappropriate. For example, several such models assume first that features of the objects moving in the display have been identified and "tagged." As the preceding chapter makes clear, this type of approach is quite orthogonal to that apparently employed by biological visual

43

systems.

Models of human motion sensing come from two traditions. each based on a different fundamental idea behind how motion might be detected. The first (chronologically). based on the biological intuitions of Reichardt described in the Introduction. aims to detect correlations of the stimulus across space and time. and views motion as the fact that a stimulus at position x_1 at time t moves to position x_2 at time $t + \Delta t$. After a review of the models which are based on this "correlational" model. the models based on *gradient methods* will be reviewed. These models are derived from the mathematical observation by computer vision researchers that under some conditions, velocity can be obtained by comparing the local rate of change of intensity with the local spatial gradient of the image (the spatial variation in image intensity). The quality of the fit between psychophysical results and each class of model will then be discussed, along with some predictions each makes regarding psychophysical experiments.

3.1   Correlational Models

According to the correlational point of view. the essence of motion perception is to detect correlations in the spatiotemporal structure of the stimulus—that something moved from point A to point B in a certain time interval. The first such model used the bilocal detectors described in the Introduction (Reichardt, 1961). To briefly review the Reichardt. model, a schematic of the detector[1] is reprinted here (see Figure 3.1. As described earlier. the input is processed by two opponent subunits in parallel. each of which multiply the output of one of the sensors with a delayed version

---

[1]As per the common usage, the term *detector* is used to refer to an individual motion-sensing unit, and the term *model* is used to refer to the entire theoretical framework, which may include multiple detectors.

44

Preferred direction
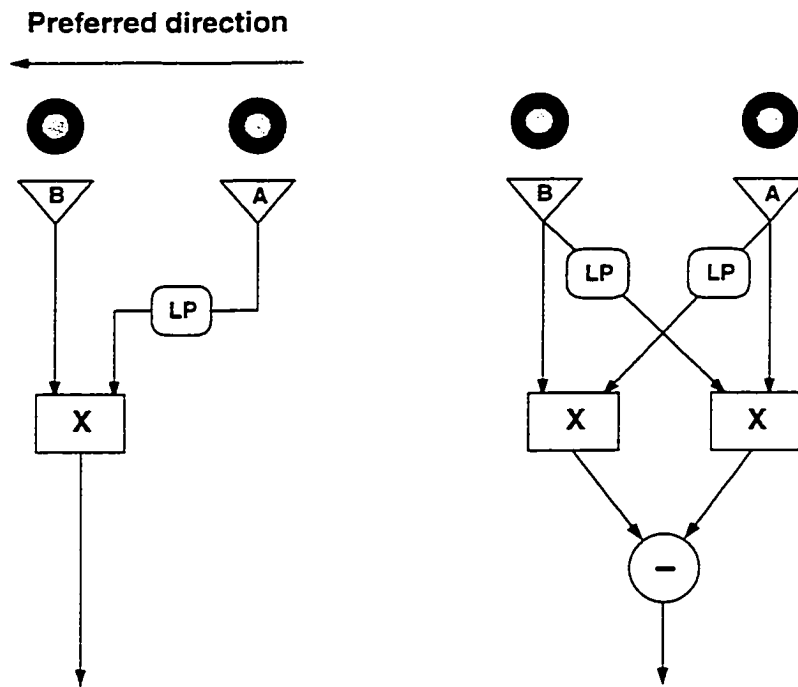
Figure 3.1: The Reichardt Model. See Chapter 1 for further details.

of the other sensor output (it should be noted in passing that this multiplication makes the Reichardt detector a non-linear system). The entire detector's output is obtained by computing the difference between the output of the subunits and averaging this difference over infinite time (which in practice, given the discussion in Chapter 2, amounts to any duration of 100 msec or more), yielding a signed signal for the direction of motion.

Because of the punctate nature of the sampling used in the Reichardt detector, this detector is subject to spatiotemporal aliasing. If a spatially periodic stimulus is displaced at a faster rate of speed than the detector's intrinsic maximum speed, then the movement direction indicated by the detector can be incorrect. Furthermore, the Reichardt detector is a direction discrimination detector, not a speed sensor—thus while one can use its output to build a speed model, it should not be considered a speed model in the same category as some of the models presented below. Still, its simple architecture makes it the most easily understood motion model, and it has

45

had broad influences on subsequent modelling efforts.

3.1.1 Adelson & Bergen's Motion Energy Model

Perhaps the most influential motion model in the last 15 years is the motion energy model by Adelson and Bergen (1985). It is a two-stage model, with the first stage consisting of a version of the linear filters described in the Introduction, while the second stage computes a non-linear combination of the outputs of these filters, so as to produce filter characteristics which are oriented in space-time, thus providing speed tuning.

First Stage: Characteristics of the Linear Filters

The linear, first stage of the Motion Energy model is, as described in the Introduction, an oriented filter with a bandpass spatial tuning and bandpass temporal tuning. Specifically, the spatial impulse functions[2] used by Adelson and Bergen (1985) are the second and third derivative of Gaussians, while the temporal functions are:

$$f(t) = (kt)^n \exp(-kt)[1/n! - (kt)^2/(n+2)!] \tag{3.1}$$

where n is either 3 or 5. Plots of the spatial and temporal functions are shown in Figure 3.2.

Since there are two spatial functions and two temporal functions, there are four possible separable spatiotemporal combinations of these functions, which are illustrated in Figure 3.3, top panel. Since these functions F are just the product of the spatial function S by the temporal function T, $F(x,t) = S(x)T(t)$, it should be

---

[2]Impulse response functions are functions which describe the response of a system to a unit pulse input. If a system is linear, then knowing the system's impulse response function allows one to derive the system's response to any stimulus, by convolution.

46
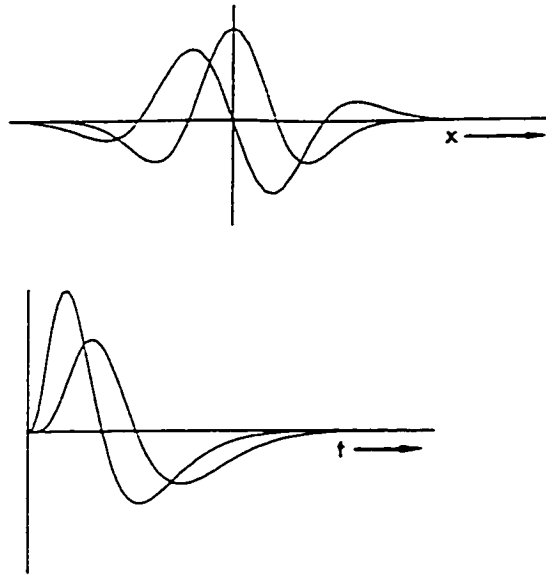
Figure 3.2: The spatial filtering functions (top panel) and temporal filtering functions (bottom panel) used in the Motion Energy model of Adelson & Bergen, (1985).
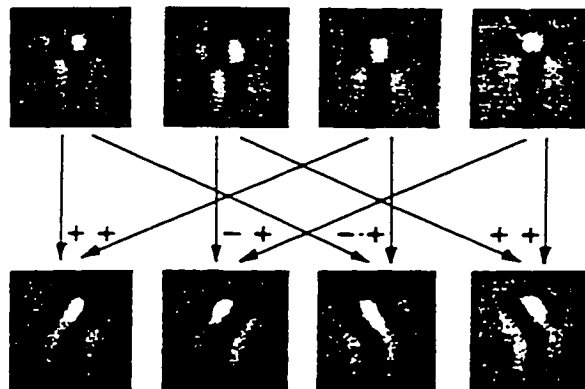
Figure 3.3: The combinations of spatial and temporal functions used by Adelson and Bergen (1985). The four spatiotemporal impulse responses show across the top are the products of two spatial and two temporal impulse responses. The ones across the bottom are sums and differences of those above. The result is a pair of leftward- and a pair of rightward-selective filters. Members of a pair are approximately in quadrature (Reproduced from Adelson & Bergen, 1985).

47

clear that these functions are *separable* in space and time. The spatial profile of the function is independent of the temporal profile. One consequence of this separability is that the filter is not oriented in space-time. An oriented spatiotemporal filter is one which responds differentially to a motion stimulus oriented in space-time. A leftward moving stimulus will have a different orientation in space-time than a rightward moving stimulus for example. Thus the separable, non-oriented mechanism just described is unable to tell leftward motion from rightward motion. This makes the first stage of the motion energy model a fairly poor motion sensor by itself.

Second stage: Nonlinear combination of quadrature pairs

The second stage of the Adelson-Bergen model is the combination of two subprocesses. The first consists of adding together the outputs of a set of the linear filters described above, thus building a spatiotemporally oriented, nonseparable filter. while the second process consists of extracting from the output of this summed linear filter a measure of motion energy.

Many combinations of separable filters can be used to obtain a non-separable filter (i.e. oriented in space-time). The combination used by Adelson and Bergen (1985) is that proposed by Watson and Ahumada (1985), which is to sum the filters according to the diagram illustrated in Figure 3.3. bottom panel. This combination yields filter pairs which are spatiotemporally oriented, and approximately in quadrature.

The second operation is the extraction of motion energy. Indeed, the output of the filters just constructed, while it will yield a good direction signal, is phase-sensitive: thus the intensity of a given filter's output will depend on the exact spatiotemporal alignment between the stimulus and the filter. A phase-independent measure of motion energy (or power) can be obtained by squaring the output of a pair of filters in quadrature. The output of this motion energy sensor is in fact a reliable, phase-insensitive indicator of motion direction, less subject to aliasing than the orginal

48
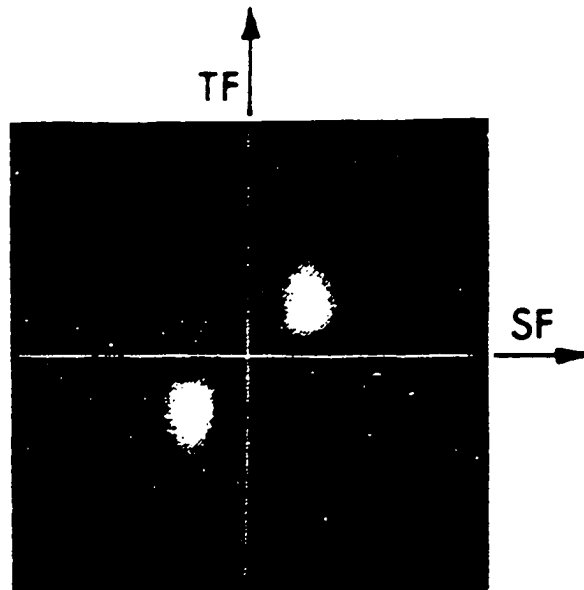
Figure 3.4: The spatiotemporal energy spectrum of a direction-selective filter built as the sum of two separable filters (Adelson & Bergen, 1985).

Reichardt model.

So far, however, no mention has been made of speed. To understand how Adelson & Bergen viewed speed (which was not a primary concern of the model), one should consider the spatiotemporal energy spectrum of the filter output, shown in Figure 3.4. This figure shows that the filter's spectrum is concentrated in two localized regions in the first and third quadrants. It is therefore sensitive to motion in one direction (say to the right—spectral sensitivity in the second and fourth quadrants would correspond to leftward motion sensitivity). Furthermore, its spectrum is localized both in SF and in TF. Thus it is sensitive only to a small range of SFs, and to a small range of TFs. This defines in turn a small range of speeds. Adelson & Bergen propose that the entire window of visibility (Watson, Ahumada, & Farrell, 1986) is tiled with filters like the one schematized in Figure 3.4. In a brief sketch, they propose that speed is extracted by an unspecified combination of the ratios of the outputs of filters tuned to different temporal frequency, but all tuned to the

49

same spatial frequency. While it is not formally specified[3], the discussion seems to assume only three temporal channels per spatial frequency—a bandpass channel for each direction, and a lowpass (static) channel. Acknowledging the weakness of ratios in conditions where the denominator can tend towards zero, as in the case of low-contrast gratings or of high speed gratings, they propose that speed estimates obtained from this ratio method be tagged with confidence values which would be low for low contrast gratings or fast gratings, and that such information would be weighed appropriately when combining (in an unspecified manner) information across spatial frequency channels.

Summary of the Adelson & Bergen model

In summary, the model proposed by Adelson and Bergen is based on a few major points, which are echoed in many subsequent models. 1) Spatiotemporally extended filters are used instead of bilocal detectors to provide spatial- and temporal-frequency tuned filters which limit the detector's vulnerability to aliasing; 2) The ouputs of these filters are combined using the motion energy design first proposed by Watson and Ahumada (1985) in a different model described below. By summing the squared output of filters in quadrature, a spatiotemporally oriented filter is constructed, which yields a phase-independent measure of the stimulus motion energy.

3.1.2   Watson & Ahumada's Scalar Motion Sensor

The Scalar Motion Sensor of Watson and Ahumada (1985) is surprisingly similar in structure to the Adelson and Bergen (1985) model but quite different in both

---

[3]These concerns would tend to imply that Adelson & Bergen viewed the ratio of interest to be the output of the e.g. leftward filter divided by the static filter, but this is not fully specified in the paper.

the components it uses and in the general perspective it presents regarding speed computation.

Like the Adelson and Bergen (1985) model, the Scalar Motion Sensor (SMS) consists of two stages. The first is a linear spatiotemporal filter, and the second is a more complex nonlinear filter derived from the linear filter but which provides spatiotemporal orientation (hence motion selectivity).

The specific functions used by Watson and Ahumada (1985) are as follows. The spatial filter used is the Gabor function. A Gabor is a sinewave modulated by a Gaussian. This function is used in many models of vision because it is both computationally convenient (being easy to evaluate, derive and integrate), fits psychophysical and physiological data relatively well, allow for the good simultaneous localization in both space and frequency (Daugman, 1985), and in some cases allows the development of useful formal proofs (e.g. in Grzywacz & Yuille, 1990). The temporal function used is the Alpha function impulse response:

$$f(t) = \xi \left[f_1(t) - \eta f_2(t) \right] \tag{3.2}$$

where

$$f_i(t) = \frac{u(t)}{\tau_i (n_i - 1)!} (t/\tau_i)^{n_i - 1} e^{-t/\tau_i} \tag{3.3}$$

and $u(t)$ is the step function, and with parameters ($\xi = 0.9, \tau_1 = 0.004, \tau_2 = 0.0053, n1 = 9, n2 = 10$) chosen to fit the same psychophysical data (Robson, 1966) used by Adelson and Bergen (1985) to motivate their temporal functions. These filters are plotted in Figure 3.5.

At this stage, this model and the Adelson and Bergen (1985) model have virtually identical filters. However, instead of adding pairs of such filters to make them non-separable and adding the sum of squares of pairs of these new filters, Watson and

51
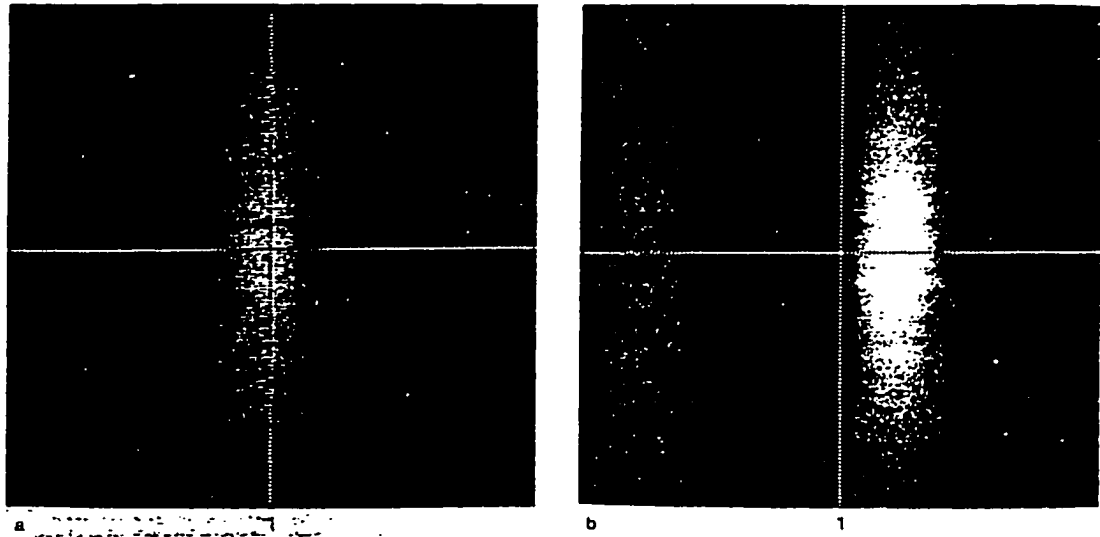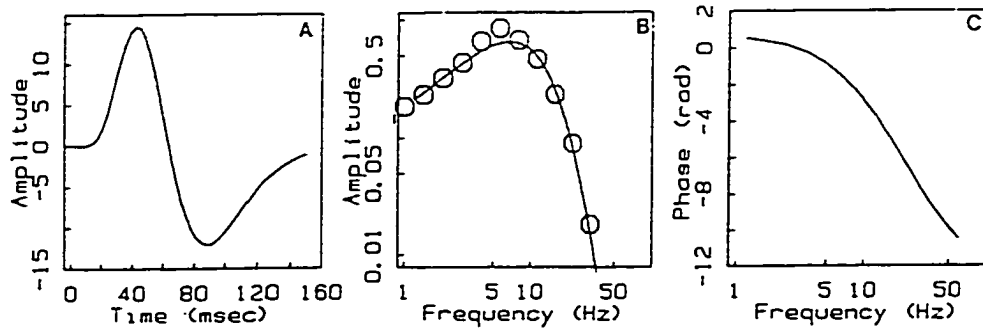
Figure 3.5: The spatial and temporal functions used in the Watson & Ahumada, (1985) model. The top panel displays the impulse response, amplitude response and phase response of the temporal filter, while the bottom panel displays the spatial impulse responses of a. main and b, quadrature paths (reproduced from Watson & Ahumada, 1985).

52
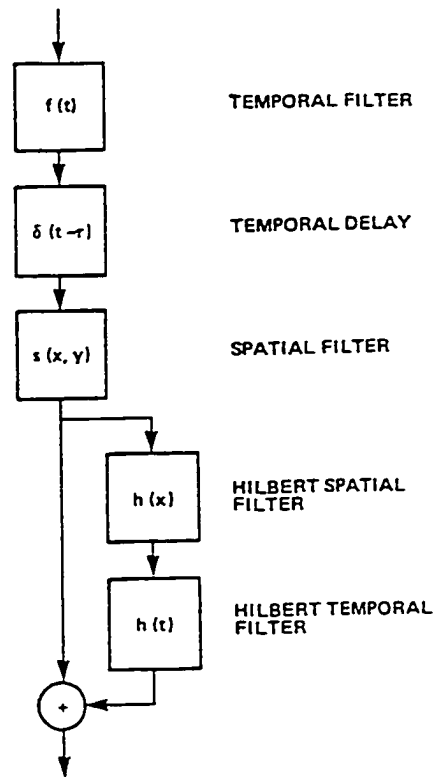
Figure 3.6: Mathematical structure of the scalar motion sensor (reproduced from Watson & Ahumada, 1985).

Ahumada (1985) propose that the output of one filter be combined with the output of the Hilbert transform in both space and time of the quadrature-matched filter. The Hilbert transform is a linear filter which has some interesting properties. It has unit gain for all frequencies (in other words it preserves the amplitude of all spatial frequencies), and converts odd functions into evens and vice versa. Thus the Hilbert transform along the spatial axis will take as input the Gabor and yield the quadrature paired Gabor, while the Hilbert transform along the temporal axis will transform the temporal filter as depicted in Figure 3.6.

At this point, there are two outputs—that of the original, Gabor/Alpha function filter. and that which went through the "quadrature path," e.g. the Hilbert transforms in both space and time. These two filters can then be linearly combined so as

to create a filter tuned for any direction.

What are the differences between the output of this combined sensor and that of the motion energy sensor? While the specific mathematical formulations of the two models are quite different, the key difference between the two which is of interest to us has to do with how speed is encoded in each model. Due to the use of the sum of squared filters in quadrature, the Adelson and Bergen (1985) model is phase-independent—thus the response of a motion energy sensor to a sinewave grating to which it is well matched (in spatial frequency and temporal frequency and direction) will be steady-state. While the signals in the Watson and Ahumada (1985) model are in spatial quadrature, their outputs are summed linearly (and not squared first). Thus the output of the SMS sensor is phase-dependent, and will in fact oscillate sinusoidally at the temporal frequency defined by the stimulus and its motion. As it is however a spatial-frequency tuned sensor, the speed information is encoded in the combination of the spatial frequency tuning of the detector (a property of the detector) and the temporal frequency of its response (a property of the dynamics of the detector given a specific stimulus). It should be noted however that the SMS detector is also an orientation-tuned, aperture-limited sensor. It thus suffers from the aperture effect, in that a given detector will respond proportionally to the projection of the stimulus velocity vector onto its preferred orientation. Thus for any stimulus motion, a set of detectors with preferred direction *around* the direction of motion will all respond. Watson and Ahumada (1985) propose that the perceived direction of motion is that of the detector with the largest response within a "direction group," a set of detectors with similar SF tuning but with varying preferred directions (see Figure 3.7). Watson and Ahumada (1985) propose that the perceived speed is chosen to be that corresponding to the unit in a direction group with the largest response for that group. Clearly, this model of speed perception assumes that the stimulus' spectrum is simple (or rather does not specify how to deal with more complex stim-

54
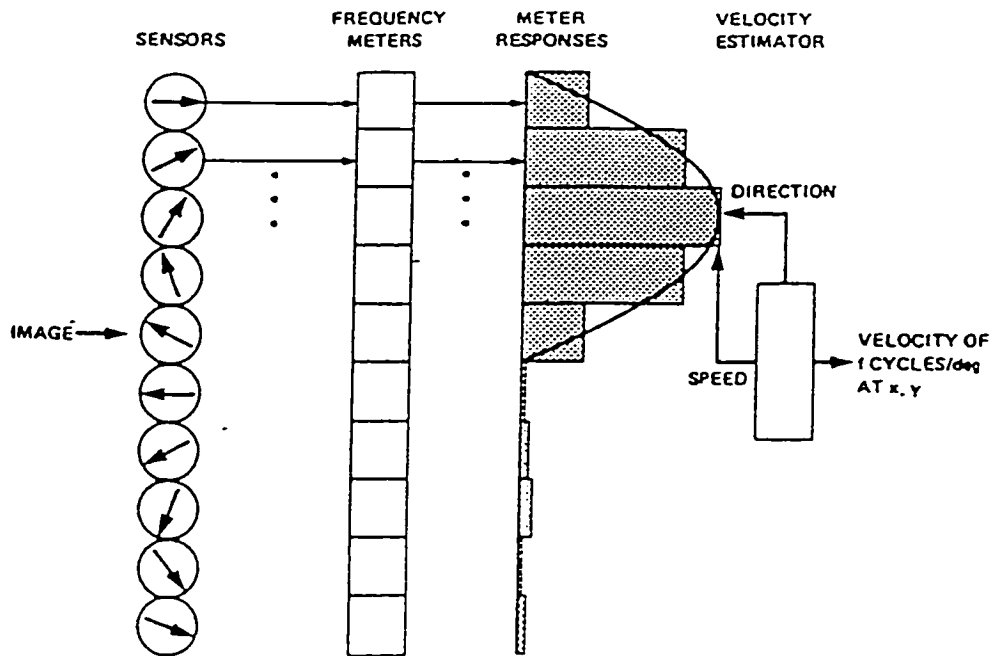
Figure 3.7: The structure of the vector motion sensor (reproduced from Watson & Ahumada, 1985). The preferred direction of the sensor with the highest response within a direction group corresponds to the perceived direction of the stimulus.

uli), and assumes that a single cell's spatial temporal frequency tuning will provide high enough resolution for speed discrimination (given the "winner-take-all" scheme just mentioned). Clearly, given the temporal frequency tuning functions described in the previous chapter and the good performance obtained on speed discrimination tasks, this claim is dubious.

The Watson and Ahumada (1985) model is thus also a two stage model. The first stage consists of linear filters of one of two types. The first type is a spatially oriented Gabor filter, with an bandpass temporal function (an Alpha function). The second consists of the Hilbert transform in both space and time of the first. The outputs of these two filters are combined *linearly* in the second stage to provide a response which is tuned in both SF and orientation. The TF of the stimulus can be extracted by examining the TF of the response of these second-stage filters, which are phase-sensitive. Analysis of the population of responses across "direction groups" of detectors tuned to the same SF but different orientations allows the determination

55

of stimulus direction and speed. While this model has the most unusual way of dealing with both temporal frequency coding and speed readout. its requirement for narrowly tuned temporal frequency mechanisms makes it quite incompatible with a fair bit of both psychophysical and physiological data. For example, its proposed speed computation does not include a role for the low-pass temporal channel, which will be shown in Chapter 7 to be an active participant in that process.

3.1.3   The Elaborated Reichardt Detector

The third model we will review. that of Van Santen and Sperling (1984, 1985) is more directly based on the Reichardt model. but includes two significant changes to that model: the introduction of spatially extended receptive fields instead of the point receptors, and different temporal filtering operations than the infinite temporal averaging proposed by Reichardt.

The van Santen and Sperling model is depicted in Figure 3.8. The spatial filters used are differences of Gaussians. similar in shape to Gabors. The temporal processing described in Van Santen and Sperling (1984) is the same temporal averaging over infinite time which was used by Reichardt. while in Van Santen and Sperling (1985), various temporal filters are considered. One of the key points of this model is that variations on the van Santen & Sperling *detector* can be shown to be computationally equivalent (on average) to both the Adelson and Bergen (1985) and Watson and Ahumada (1985) detectors. However, the model as a whole is not of particular interest in the context of speed perception, as Van Santen and Sperling (1985) point out that their detector is unable to discriminate between stimuli with speeds which vary by an order of magnitude. It will thus not be considered further.

56
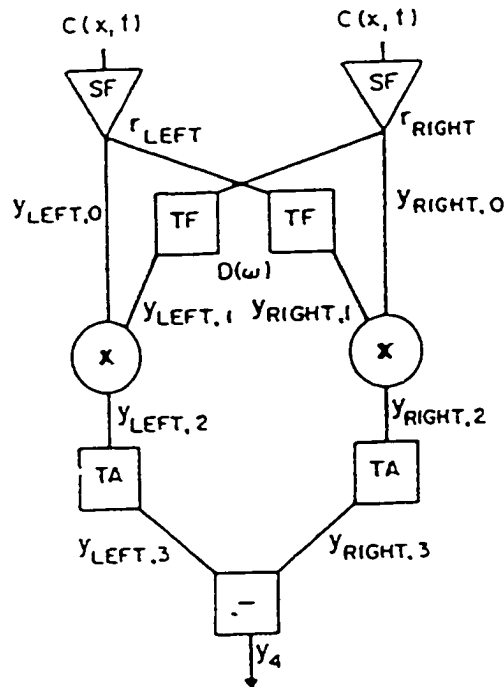
Figure 3.8: The elaborated Reichard detector of Van Santen and Sperling (1985). The input is a luminance pattern with contrast $c(x,t)$. It is sampled by linear filters (receptive fields, SF's) with spatial responses r_{left} and r_{right}; $y_{i,H}$ (H = left, right) represents the signal at the various stages i for the left and right subunits. TF indicates a linear, time-invariant filter with Fourier transform $D(\omega)$, × indicates a multiplication unit, TA indicates a temporal integration operation and − indicates a unit that subtracts its left from its right input.

3.1.4 Heeger's model

Building on the motion energy computation of Adelson and Bergen (1985), Heeger (1987) proposed a model of motion perception which is more relevant to this dissertation than the previous ones, which mostly lay the theoretical groundwork. Indeed, Heeger's model is one of the first to explicitely describe a plausible algorithm for the computation of *speed*, instead of simply direction.

The model also has two stages. The first stage consists of motion energy units, built as 3-D Gabor filters with response defined by the equation:

$$g(x, y, t) \;=\; \frac{1}{\sqrt{2}\; x^{3/2}\sigma_x\sigma_y\sigma_t} \exp\left[-\left(\frac{x^2}{2\sigma_x{}^2} + \frac{y^2}{2\sigma_y{}^2} + \frac{t^2}{2\sigma_t{}^2}\right)\right]$$

$$\times \sin(2\pi\omega_{x_0}x + 2\pi\omega_{y_0}y + 2\pi\omega_{t_0}t) \qquad (3.4)$$

which is clearly a formula with six parameters. $\sigma_x, \sigma_y, \sigma_t, \omega_{x_0}, \omega_{y_0}, \omega_{t_0}$. The first three define the spatial constant of the Gaussian component of the gabor in both spatial direction and in time, while the last three define the center frequency in both spatial frequencies (or equivalently spatial frequency and orientation) and temporal frequency.

These filters are different from the filters seen up to now in that there is no special treatment of the time axis as compared to the spatial axes. This is done for computational reasons regarding the computation in the second stage of the model. The use of a Gabor in time has two consequences regarding the ability of the model to fit experimental data. The first is that since a Gabor is symmetric around the origin, when used in the temporal dimension, it is a non-causal filter—thus the response of the filter at time t is dependent on the stimulus at times in the past (which is expected) and *in the future*, which is both unexpected and not physically realizable. A way to allow the use of non-causal filters like this is to assume that there is a fixed delay in processing Δt which is prior to the action of the filter, with Δt large enough that the value of the Gabor at times earlier than $t - \Delta t$ is insignificant. The second consequence of the use of a Gabor filter in time is that the frequency response of such a filter does not fit the psychophysical data as well as other filters such as the Alpha functions used by Watson and Ahumada (1985).

The filter just described is just the *detector* part of the Heeger (1987) model. These detectors are organized in a very specific way, as we will see, and their output

58

is processed by explicit algorithms which allow a readout of speed.

The filters are organized in spatial frequency bands, so that all filters in a family have the same spatial frequency tuning along their preferred direction (i.e. $\sqrt{\omega_{x_0}^2 + \omega_{y_0}^2}$ is constant). Within such a spatial frequency band, and for each orientation, Heeger assumes twelve filters. Four of these filters are most sensitive to stationary stimuli, four to (e.g.) leftward moving stimuli, and four to rightward moving stimuli. This architecture is repeated at a variety of scales using the Gaussian pyramid architecture (Burt, 1983), which corresponds to filters with equal bandwidths spaced an octave apart in spatial frequency, but tuned to the same temporal frequency.

To summarize this architecture, for each direction of motion (say leftward), there is a family of pairs of filters, two per octave of SF, one tuned for stationary stimuli and one for moving stimuli. Couched in psychophysical terms, this means that there are spatial frequency channels one octave apart, and that there are two temporal channels (one lowpass and one bandpass).

Up to now, the tuning properties of the elementary detectors have been specified, as well as their architectural arrangement. Of greater interest is how Heeger proposes to compute speed from the responses of these detectors. Clearly, a moving stimulus will yield responses in a variety of detectors with tuning properties which overlap the stimulus properties. While the mathematical derivation of this response is not of particular interest here, it should be noted that it relies essentially on a blurring operation. Using the Fourier properties of the Gabor filters, Heeger derives the output of each filter to a stimulus with a white (flat) power spectrum, translating in a given direction at a given speed. This result provides a system of equations (one for each filter) giving the response of the filters for any flat-spectrum stimulus as a function of its direction and speed. Given this "forward" result, Heeger's formulae allow the model to estimate (using a "backward" estimation process) the stimulus

59

speed from the filter responses.

This result is powerful, as it allows the exact computation of speed. However, it assumes that the stimuli have a white power spectrum not only along the direction of motion but across all directions—thus making it inexact for the wide majority of visual stimuli. Ackowledging this, Heeger provides a least-squares algorithm for an estimation procedure which minimizes the error between the predicted and measured motion energies, by normalizing within a direction band (but still normalizing across spatial frequencies and assuming a white spectrum within a direction band).

3.1.5 Grzywacz & Yuille, 1990

The second model we will consider which explicitely specifies a mechanism for the extraction of speed from a set of filter responses is the Grzywacz and Yuille (1990) model. The filter characteristics of this model are almost identical to those used in Heeger (1987), but the speed computation stage is quite different. Again, each stage will be considered in turn.

The filter responses used by Grzywacz and Yuille (1990) is also a 3-D (space/time) Gabor, except that unlike those used in Heeger, the Gabor is circularly symmetric in the two spatial dimensions ($\sigma_x = \sigma_y$). Furthermore, the model makes the further assumption that the spatial frequency bandwidth is narrow compared to the temporal frequency bandwidth ($\sigma_x \ll \sigma_t$). This assumption seems warranted given the 1–2 octave bandwiths estimated in the spatial frequency domain and the fact that while the specific number of temporal frequency channels is under debate, all proposed channels have very wide total bandwidth (see Chapter 2).

The heart of the Grzywacz and Yuille (1990) model is the proof (not given here) that given an infinite set of the filters just specified (more specifically including all spatial frequencies and temporal frequencies), for *any* translating stimulus, the filters

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission. www.manaraa.com

which respond most strongly to the stimulus lie on a plane in the parameter space defined by the filter's spatial and temporal frequencies—a filter F whose parameters lie away from that plane will respond less than a filter G whose location in parameter space is the normal projection of F onto the plane corresponding to the stimulus velocity. A graphical representation of this theorem is given in Figure 3.9.

In other words, the Grzywacz and Yuille (1990) model uses an inverse approach compared to Heeger (1987). Instead of estimating the velocity which yielded a pattern of activity on a limited set of filters, this model assumes (in the mathematical part of the model) an infinite set of filters and proves that the stimulus energy lies in a plane.

Given this mathematical proof, the model goes on to an implementation strategy, which assumes only a limited set of filters (as in the Heeger (1987) model). The task of the speed estimation in this implementation is then to find out what the plane of maximal (theoretical) energy is, given the pattern of responses of a set of filters. Grzywacz and Yuille (1990) propose two strategies for this estimation process in detail. The first, the *ridge strategy*, assumes a large set of velocity tuned units, which receive weighted excitatory input from the units corresponding to the spatial frequency and temporal frequency tuned filters. This is illustrated in Figure 3.10. Clearly, the weighting function which defines the amplitude of the input to each velocity-tuned unit is the crucial aspect of this strategy, as the selection of the perceived speed is performed by a winner-take-all mechanism among the velocity-tuned cells. This weighting function includes a mechanism for biasing low velocities, thus addressing the aperture problem. If the filters are based on Gabor functions, then an exact weighting function can be derived analytically.

The second strategy proposed by Grzywacz and Yuille (1990) is called the *estimation strategy*. It has the advantage over the ridge strategy that it does not require the existence of velocity tuned cells (for which there is no reported physiological

61
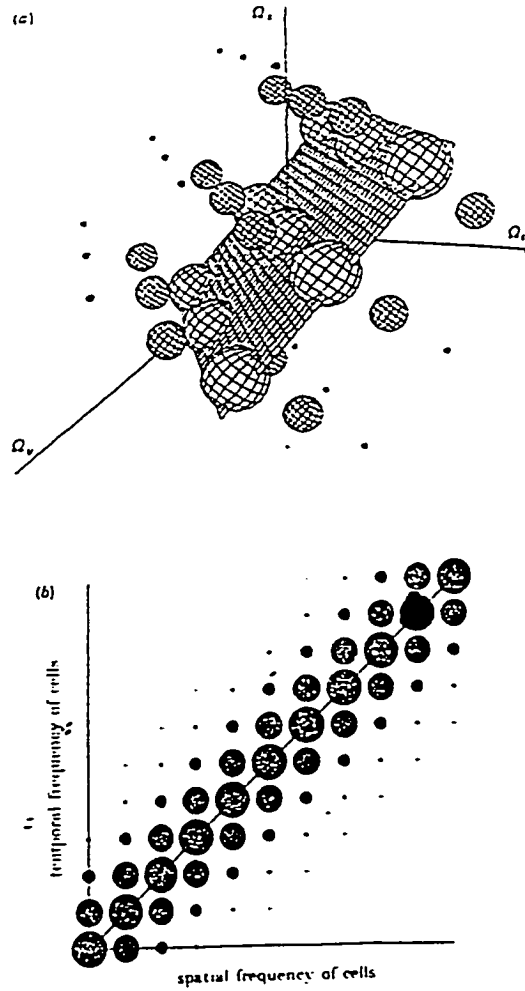
Figure 3.9: Most of the distribution of motion energies lie near the plane $\Omega n \cdot v + \Omega_t = 0$ in the space of the cells' optimal frequencies ($\Omega n = (\Omega_x, \Omega_t)$ and Ω_t). (a) This figure shows this distribution for a translating dot and indicates the plane where the motion energies (sums of responses of quadrature pairs of directionally-selective frequency tuned cells) are maximal. The motion energies rapidly decrease as the distance of the filters' optimal frequencies from the plane increases: diameter = cell response. (b) This figure shows a two-dimensional cross section of a distribution like the one in (a). (Reproduced from Grzywacz & Yuille, 1990).

62
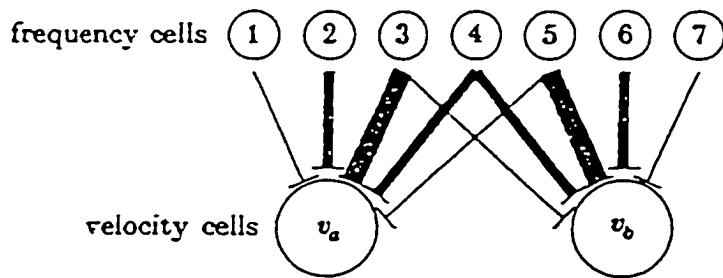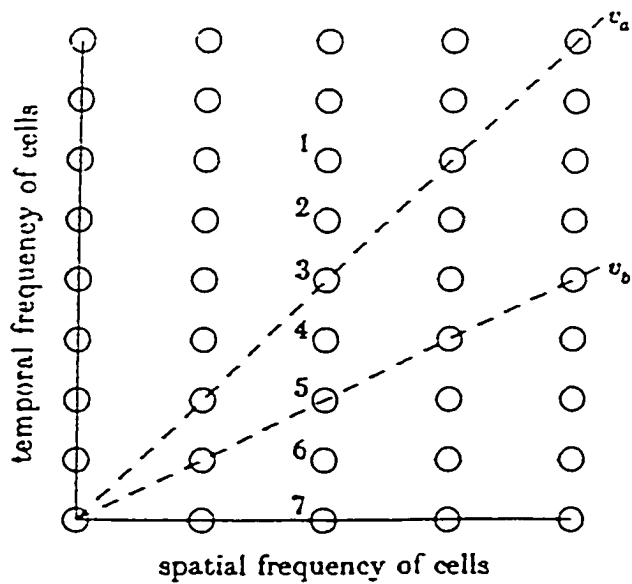
Figure 3.10: The ridge strategy. In the top part of the figure, the centre of the open circles represent some sampling locations in the space of the cells' optimal frequencies. The diameters of the circles are not bandwidth here. The cross sections of two velocity planes (corresponding to velocities v_a and v_b) are shown and seven motion-energy cells (directionally selective frequency tuned cells) are labelled. In the bottom part of the figure is shown how each of tehse seven cells make excitatory connections to cells tuned to the velocities v_a and v_b. The thickness of lines in each connection represents the connection's strength. Motion-energy cells whose parameters are close to a velocity plane make strong connections to the corrresponding velocity cells. Motion-energy cells with parameters distant from that plane make correspondingly weaker connections (Reproduced from Grzywacz & Yuille, 1990).

63

evidence, see Chapter 2). In this strategy, an estimate of the temporal frequency of the stimulus is made within each spatial frequency band. It is important for this strategy that there be detectors with identical spatial frequency tuning. Grzywacz and Yuille (1990) show that the distribution of responses of detectors with the same SF tuning will be a function of the image's spatial characteristics and of the image velocity. However, the response of the detectors is linearly decomposable into these two components. Thus regardless of the spatial characteristics of the image, the detector response can unambiguously yield the correct speed estimate.

3.1.6 Heeger & Simoncelli

A more recent model developed by Heeger & Simoncelli (Heeger, 1993; Heeger, Simoncelli, & Movshon, 1996; Simoncelli, Heeger, & Adelson, 1992; Simoncelli & Heeger, 1994, 1997) builds on the Heeger (1987) model, but with a different emphasis, that of fitting physiological data, especially that on the effects of stimulus contrast on cell responses. The model proposes three different unit architectures, modeling V1 simple cells, V1 complex cells, and MT cells respectively.

The model for V1 units in the Heeger & Simoncelli model is similar to the first-stage detector used in Heeger (1987), except that instead of being a 3-D Gabor, it is a three dimensional version of a third derivative of a Gaussian. There are 28 of these filters at each spatial location, corresponding to 8 different orientations, and which cover a sphere in the spatiotemporal frequency space.

The output of these linear filters is then half-wave rectified and squared, so that negative responses of the filter are suppressed, while positive responses are squared. The output of these rectified filters is then normalized by the summed activity of a pool of neurons with similar tuning properties. This *response normalization* is introduced in order to account for a set of data both from physiological and psychophysical

64

experiments (Robson, 1988; Bonds, 1989; Albrecht & Geisler, 1991; Heeger, 1991, 1992a, 1992b, 1993; DeAngelis, Ohzawa, & Freeman, 1992; Carandini & Heeger, 1994; Carandini, Heeger, & Movshon, 1996; Tolhurst & Heeger, in pressa, in pressb; Nestares & Heeger, in press). This normalization factor is uniform over the full range of orientation, direction and spatio-temporal frequency, thus while it effectively accounts for some contrast data, it does not affect the tuning properties of the cells relative to one another (which are of central interest to us in the context of speed models). Thus for our purposes the model's V1 simple cells are similar in many respects to the first stage of several of the previous models.

The only difference in the Simoncelli and Heeger (1997) model between V1 simple cells and V1 complex cells is the receptive field structure—V1 simple cells are sensitive to the exact location of stimulus contrast within the subunit which make up its receptive field, while V1 complex cells integrate over their receptive field in a location-independent manner. The Simoncelli and Heeger (1997) model of a V1 complex cell is the result of adding together the output of several V1 cells which are distributed over a local spatial region, but with the same spatiotemporal frequency tuning, orientation and phase profile.

Both types of V1 cells are tuned in spatiotemporal frequency and orientation, which corresponds to power spectra which are localized in two locations diametrically opposed to one another in spatiotemporal frequency space (see Figure 3.11). This cell, therefore, will only fire given stimuli which not only have a specific speed and orientation, but also a specific spatial frequency content. A stimulus of the "right" speed and direction but with a single spatial frequency which does not match the cell's tuning function well will yield no response. Thus these cells are not truly "speed-tuned."

The MT cells in this model are constructed so as to provide just this kind of speed tuning. Their architecture is (not coincidentally) very similar to that of the V1 cells:

65
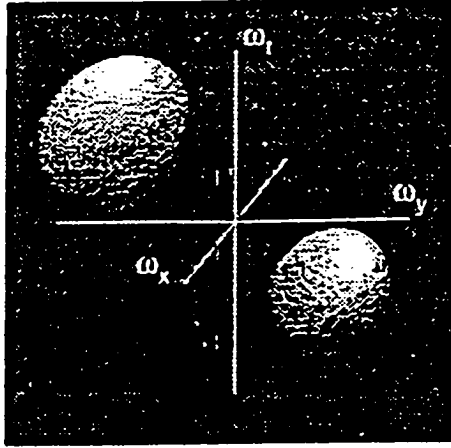
Figure 3.11: Selectivity of a V1 neuron corresponding to a pair of localized spatio-temporal frequency bands, symmetrically arranged about the origin. (Reproduced from Simoncelli & Heeger, 1997).

the MT cells pool the responses of sets of V1 cells tuned to a given speed but whose power spectra tile the spatiotemporal frequency plane corresponding to that speed. To be more precise, the MT cell receives input from all V1 cells, weighting this input by a measure of how related a given V1 tuning is with the speed tuning of the MT cell. This weighting includes inhibitory connections, so that V1 cells which are *incompatible* with the speed tuning of a given MT cell will, when active, decrease the response of the MT cell. This generalized excitatory/inhibitory framework is similar to the shape of the linear receptive field used in the V1 model. It should be clear that the frequency selectivity of the MT cells will be that illustrated in Figure 3.11, as that tuning defines the wiring diagram for these cells.

The output of these MT cells undergoes half-squared rectification and divisive normalization, just as the V1 cells did. This provides a second stage of contrast normalization, and avoids negative responses from MT cells incompatible with the image motion.

As Simoncelli and Heeger (1997) describe, even with this carefully designed speed tuning, the output of a single MT cell cannot by itself yield the speed of the stimulus, since the response of a cell is dependent not only on its speed tuning, but on the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission. www.manaraa.com

contrast of the stimulus (more precisely the twice-normalized contrast response of the unit), and on the spatial spectrum of the stimulus (a broad spectrum stimulus with a given contrast will result in activity of more V1 cells which are connected to the MT cell in question than a single sinewave stimulus). Thus stimulus velocity must be decoded from the population of speed-tuned MT cells.

Simoncelli and Heeger (1997) present displays of the model output for a set of test cases, arguing that the distribution of MT responses "implicitly" codes the perceived velocity. It is unfortunate that the model does not make explicit how one is to go from a distribution of activity on the MT responses to a perceptual judgment. To be fair, the model's main aim was to model physiological data, which it does quite well. It remains to be seen whether a reasonable "decision rule" will yield good match with the psychophysics. Of special concern in this regard is the disambiguation of one-dimensional patterns (the classic "aperture problem"), which would seem to require a somewhat *a posteriori* bias for low velocities, and the requirement of integration over spatial extent in building complex cells, a point highlighted by Grzywacz and Yuille (1990) which is likely to pose challenges regarding the perception of mutliple non-transparently moving objects.

3.2 Gradient Models–Early Models

In a parallel track to the correlation models just reviewed exists an entirely different set of models, based on a mathematical insight regarding the temporal statistics of image intensities during rigid motion. This insight, first reported by Limb and Murphy (1975) in a study of motion in television signals, is that the change in time of the local intensity of a visual stimulus is proportional to the velocity of the stimulus and of the spatial gradient at that point. This is true instantaneously, and only

67

for a restricted class of stimuli,[4] but it captures mathematically the relationship pointed out earlier on grating motion between spatial frequency, temporal frequency and speed of gratings—a low spatial frequency grating has a lower spatial gradient, and at a given speed, the temporal frequency will be proportional to the spatial frequency. To state the result more precisely, the component of image motion along the direction defined by the spatial intensity gradient[5] $\vec{v}_\perp$ is given by the equation:

$$\vec{v}_\perp = -\frac{E_t}{E_\perp}$$

where E_t is the temporal gradient of the image intensity, $E_\perp$ is the magnitude of the spatial gradient of the image intensity in the direction of the gradient (the direction along which the change is fastest). The negative sign is present because the velocity being computed is that of the surface moving with respect to the observer (hence the negative of the sampling point motion). The component of the velocity vector *along* the isointensity contour lines (perpendicular to the gradient), $\vec{v}_\parallel$, can not be determined with this method. Other methods such as feature-tracking methods can compute $\vec{v}$ directly, but assume that feature can be identified easily and matched between frames, a problem which is often computationally intractable.

Limb and Murphy (1975) originally resolved this underspecification by assuming that object segmentation had already been performed, and by assuming a random gradient distribution. Under those conditions, the *average* velocity can be estimated. Fennema and Thompson (1979), on the other hand, did not assume either presegmen-

---

[4]Specifically, it is exactly true for lambertian surfaces with no shadows, as specularities (highlights) and shadows tend not to travel with the stimulus.

[5]The spatial intensity gradient is a vector quantity indicating both the *direction* of greatest change in image intensity (in a grating, the direction perpendicular to the grating), and the *rate of change* of the image intensity (in a sinewave grating, this value depends on the location, with maximal value halfway between the peaks and the troughs).

68

tation or random gradient distribution, but instead computed the estimated velocity vector at many points in the image, and used clustering techniques to estimate object groupings. This method assumes frontoparallel translation (without which the velocity estimates across objects would be too different from each other to cluster effectively). Horn and Schunck (1981) developed a related model which traded the assumption of frontoparallel translations for one of smoothness. Indeed, Horn and Schunck (1981) resolved the underspecification from the gradient identity discussed above by requiring the minimization (using a least-squares method) of the integral over the image of the magnitude of the *velocity* gradient. This allows the Horn and Schunck (1981) model to deal correctly with rotating objects.

The next model to make use of the gradient identity was Marr and Ullman (1981), a model which estimated the image velocity at edges (defined as the zero-crossings of the convolution of the Laplacian of a Gaussian over the image: $\nabla^2 G(x, y) * I(x, y)$). The model does not attempt to deal with the underspecification of the velocity vector, and simply proposes that the cells compute the velocity to within a 180° range, and that several cells are used (in an unspecified manner) to restrict this orientation ambiguity. The model does not attempt to deal explicitly with the issue of speed, much like the early direction models of the correlational type.

In addition to the need for a regularization assumption (such as the smoothness constraint used by Horn and Schunck (1981)), the gradient models suffer from an instability due to the use of a ratio. Indeed, if the speed of the stimulus is given by:

$$v = -\frac{\partial I / \partial t}{\partial I / \partial x} \qquad (3.5)$$

then it should be clear that when $\partial I / \partial x$ gets close to zero, the computed speed gets very large, regardless of the temporal derivative. This is the case for all areas in the image which have little contrast.

3.2.1 Koch, Wang & Mathur

A proposal for dealing with this ambiguity was put forth by Koch, Wang, and Mathur (1989) in a model which is a derivative of the Marr and Ullman (1981) model. As in that model, the first stage of processing is performed by two types of units, S and T units, which are presented as homologs of the lowpass and and bandpass temporal channels discussed in Chapter 2. The mathematical definition of these units is given by:

$$S(i,j) = \nabla^2 G * I(i,j) \tag{3.6}$$

$$T(i,j) = \frac{\partial \left[\nabla^2 G * I(i,j) \right]}{\partial t} \tag{3.7}$$

where ∇^2 is the Laplacian operator (i.e. $\partial^2/\partial x^2 + \partial^2/\partial y^2$) and G is a two-dimensional Gaussian, an operator first proposed by Marr and Hildreth (1980) as being useful in finding edges in images.

The novel aspect of the Koch et al. (1989) model is that instead of just using the ratio of T and S, which suffers from instability, they propose a neurally inspired architecture which implements the more complex formula:

$$U(i,j,k) = \frac{-T(i,j)\nabla_k S(i,j)}{|\nabla_k S(i,j)|^2 + \epsilon} \tag{3.8}$$

where ϵ is a constant and ∇_k is the spatial derivative along a direction indexed by k. Since we are only concerned with speed, the directional derivative is not important to us. What *is* important, however, is the behavior of U as a function of T and S. Unlike in the standard gradient model, for regions with low image contrasts, as $\nabla_k S$ goes to 0, the parameter ϵ prevents the output U from diverging. In such conditions (defined mathematically by $\nabla_k S \ll \epsilon$), $U \approx -T\nabla_k S$, in other words the product of a temporal derivative operator and a second-order spatial filter. Thus the Koch

et al. (1989) model is the standard gradient model at high contrasts. and a second-order correlation model at low contrast. Importantly. it can only accommodate two temporal channels. unlike several of the previous models.

3.2.2 The Multi-Channel Gradient Model

The final gradient model which will be considered here is the Multi-Channel Gradient Model (McGM) (Johnston, McOwan, & Buxton, 1992; Johnston & Clifford, 1995. 1995). This model is very much like the standard gradient model. except for its solution to the underdetermination of the denominator of the gradient. According to this model, "when the first partial derivatives of the brightness function in the xt-plane are close to zero, the second partial derivatives are unlikely to have low values." This assertion is based on results in Catastrophe theory (Poston & Stewart, 1978), and its validity in the case of visual stimuli may need some empirical verification. For example, in regions of low stimulus contrast, where the first derivative of the brightness function is zero. one would assume that the second partial derivative would either be underdetermined. or vary wildly due to the fact that the deviations from zero are likely to be due to noise in the detectors rather than image structure. If one assumes for now that this claim is borne out anyway, then Johnston *et al.* argue that the motion signal can be extracted from any nth-order derivative, not just the first derivative as described in the original Horn and Schunck model. This extension to higher-order derivatives allows them to model the detection of second order motion patterns (see e.g. in Derrington & Badcock, 1985; Turano & Pantle, 1989; Zanker, 1990, 1993). Thus in some ways the McGM model can be viewed as the regular gradient model, except that the spatial and temporal operators can be higher order than in the standard model.

71

3.3 Smith & Edgar

The final model we will consider is somewhat of an outsider to the models previously presented, as it is billed as a temporal frequency model rather than a speed model. It is however highly intuitive, and one can trivially consider speed perception to be based on such a temporal frequency estimate and that from a spatial frequency model. The Smith and Edgar (1990) model simply assumes two temporal channels, one lowpass and one bandpass, with overlapping sensitivity functions over their entire range. These sensitivity functions are such that at no two temporal frequencies is their ratio identical. Thus, this ratio of two channel outputs can serve as a unique measure of temporal frequency, as it will be monotonically related to temporal frequency. The model is tested with adaptation studies, which yield a good fit only if two conditions are met: first, adaptation must be subtractive and not divisive, so that after adaptation the sensitivity of the channel is reduced by a fixed amount throughout its sensitivity range, as opposed to by a fixed proportion. Second, for the model to fit their data, Smith and Edgar had to posit that the bandpass mechanism be much more adaptable than the lowpass mechanism, a distinction for which there is no physiological data (or model-independent psychophysical data).

3.4 Summary

While there are important differences between the models described in this chapter, the similarities among them are perhaps more important. All of these models propose two stages of computation. The first stage consists of the computations which are well established as being performed by V1 cells, such as spatial-frequency and temporal-frequency specific processing. The second stages of computation differ more significantly among models, as it is this stage which truly extracts the

speed information. The experiments in the next chapters are aimed, not so much as model-testing experiments, but as experiments which further specify the models. What combination rules should these models use for spectrally complex stimuli if they are to match human data? What are their spatial and temporal properties? Chapter 9 will then propose a new model which provides a framework within which the psychophysical results can be incorporated.

Chapter 4

Speed Discrimination is Contrast Dependent

All of the models of speed perception which were considered in Chapter 3 assume that the units which carry the speed signal are multiplexing information about several aspects of the stimulus, scaled by a contrast-dependent factor, and produce a single-valued output. Thus a change in the absolute response level of such a unit cannot be attributed to any one of the many possible sources of response variation. These sources of variation include changes in the matches between the properties of the stimulus and the unit's tuning functions (e.g. SF, TF, orientation and position). Importantly, a change in response can also be due to changes in the stimulus contrast, with no other stimulus change. This type of model is quite common in psychophysics, and many models of discrimination in various domains include this type of design (Wilson, 1986; Watt & Morgan, 1984; Klein & Levi, 1985; Regan & Beverley, 1985). The standard mechanism which discrimination models employ to "demultiplex" the response and attribute changes in response to one or the other sources of variation is to use a population code across banks of differently tuned units. Take as an example orientation discrimination. If a stimulus changes orientation, the units which are

74

tuned to the old orientation will respond less after the change. while units which are tuned to the new orientation will respond more. Thus by examining the output of sets of units, the change in the responses of the units can be attributed (with high likelihood of being correct) to a change in orientation and not to two changes in the contrasts of two different stimuli.

The fundamental consequence of such a design is that all types of discriminations for which the units show tuning (i.e., SF. TF, orientation), as well as contrast discrimination, will be based on the output of a common, unique set of filter responses. Figure 4.1 shows a schematic of a subset of the filter sets, displayed along two dimensions. The bandwidth of the spatial frequency tuning of the units increases down the vertical dimension, and the preferred orientation of the units varies along the horizontal dimension. The spatial frequency discrimination system is concerned with demultiplexing changes between the units which lie in the vertical axis, while the orientation discrimination system is designed to examine the outputs of units across the horizontal axis. Contrast discrimination depends on the sum of the responses of all the units (the dimensions corresponding to TF and spatial position are not depicted for clarity).

Discrimination performance is never perfect. however. Indeed, all models of detection and discrimination have at their core a definition of the source of noise which is responsible for subject errors. These sources of errors can be due to neural limitations. or to faulty assumptions about the world. The most common theory used for analyzing subject errors is signal detection theory (Green & Swets, 1966), which assumes that the response is based on a decision variable which varies monotonically with filter response, itself a monotonic function of stimulus contrast. This decision variable is then processed by a decision mechanism which is noise-free. The source of noise in this model is therefore always present in the decision variable (indeed the result of signal detection theoretic analysis is a measure d' of the signal to noise
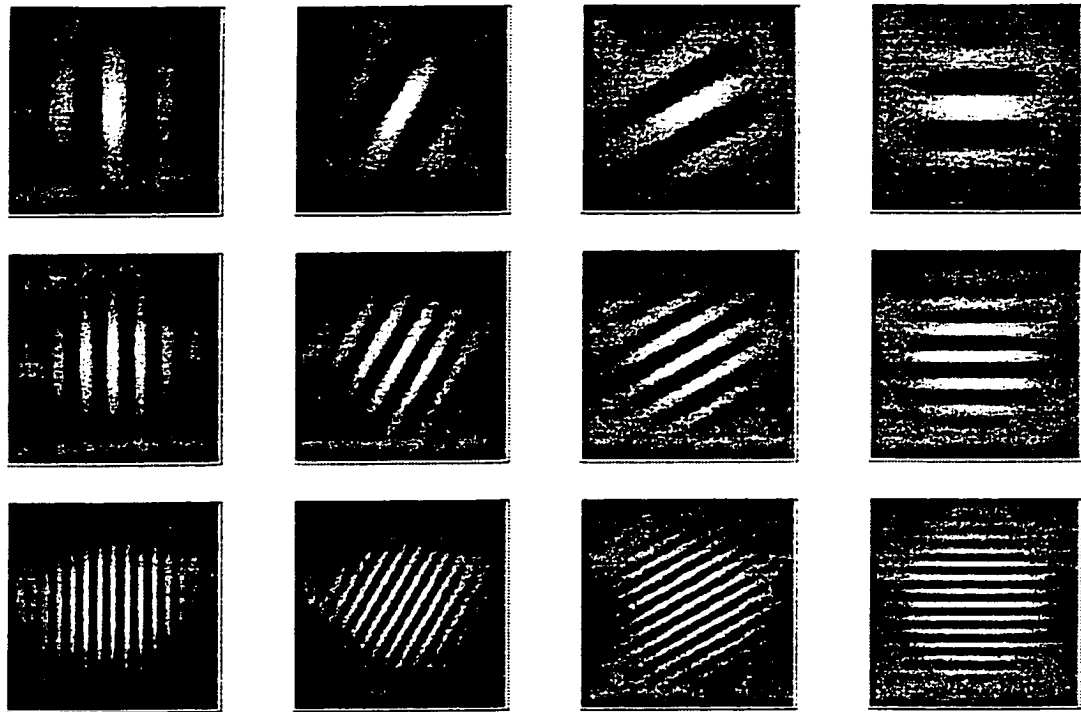
Figure 4.1: Figure of banks of filters tuned to orientation and spatial frequency. The filter orientation varies along the horizontal axis, and the spatial frequency varies along the vertical axis. See text for details.

ratio within that variable[1]). This assumption is often carried over into discrimination models, which assume that the decision variable is also obtained by a noise-free process from the output of the filters described above. Thus performance in a given discrimination task depends only on the signal to noise ratio at the level of the filters, weighted according to the weights assigned to whatever filters are most appropriate for that task (the *decision rule* for that task). This weighting of filter outputs, however, is assumed to be noiseless, along with all further processing. Thus performance in such a model is *necessarily* limited by the signal to noise in the filters. This is a powerful assumption, since it allows one to use filter characteristics obtained by,

---

[1] The d' statistic is only valid for models which assume a Gaussian noise distribution or for binary variables with large numbers of trials, both of which are commonly assumed in psychophysics.

76
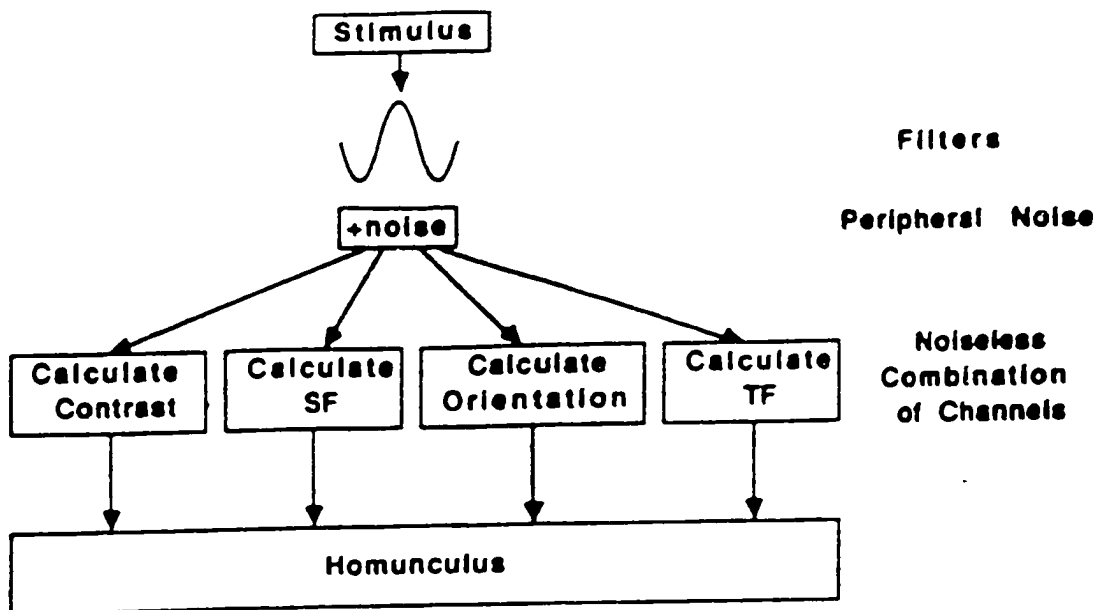
Figure 4.2: A schematic diagram of the error-propagation class of models, which include the models of Wilson (1986), Klein and Levi (1985), Regan and Beverley (1985). A number of filters respond to the stimulus, of which one is shown. The filters are bandpass or lowpass in SF, TF, and orientation, and their responses are monotonically increasing functions of contrast. Gaussian noise is added to the output of each filter. A noiseless combination process is used to compute estimates of SF, TF, orientation and contrast from the filter responses. Since all discrimination tasks are limited by the noise added to the filter outputs, they will all have the same contrast dependence. (Reproduced from Bowne, 1990).

e.g., contrast discrimination experiments, and apply them to. e.g., TF discrimination tasks.

Is this assumption valid, and if not, what are the consequences for speed discrimination models? The only report testing this assumption is the study by Bowne (1990). Bowne presented an argument similar to the one above. and proceeded to test it by measuring the effect of contrast modulation on the discrimination thresholds for observers in four tasks: contrast discrimination. spatial frequency discrimination, orientation discrimination, and temporal frequency discrimination. According to the

model presented above (which Bowne refers to as the error-propagation model, since the only source of noise is at the first stage of processing, and it is propagated without addition throughout the process—see Figure 4.2), *all* discrimination thresholds should increase linearly with d', which in turn should be approximately proportional to the relative change in contrast, SF, orientation or TF. Thus the error-propagation model predicts a similar shape for the threshold expressed in Weber fractions[2], regardless of the stimulus attribute which is being discriminated. Bowne's data (see Figure 4.3) show that contrast discrimination decreased approximately linearly with increasing contrast, for contrasts between 2% and 50%, while SF, orientation and TF discrimination performance are independent of the contrast level used.

Bowne's conclusion from these results is simply that discrimination performance (except for contrast discrimination) is not limited by noise at the level of the filters (which he terms *peripheral noise*), but instead by *central noise*, which occurs at the level of (or just prior to) the decision rule. This is best illustrated by referring to his schematic of the discrimination process, redrawn in Figure 4.4. The key addition to the model presented earlier is that of a separate, independent noise source at the stage of each (task-specific) decision rule. Performance in one task can therefore not necessarily be directly compared to performance in another task.

Do Bowne's results mean that speed discrimination is never limited by noise in the filters described above? If that is true, then all of the models described in Chapter 3 need to be reconsidered in light of this "late-noise" model. As discussed in Chapter 2, correctly measuring TF discrimination performance is hard, as observers generally base their responses on perceived speed rather than perceived TF (McKee

---

[2]The Weber fraction is the ratio between the change in the stimulus attribute required for a given percentage of correct answers divided by the value of that attribute. For example, a Weber fraction for contrast discrimination in Bowne's experiment is the change in contrast required for 92% correct performance and the contrast at which the threshold was estimated.

78
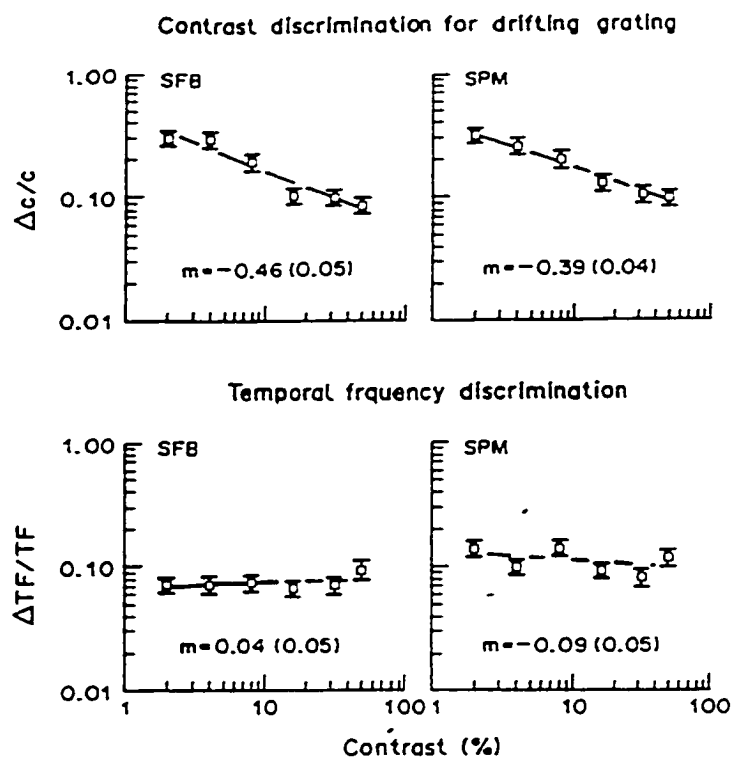
Figure 4.3: Contrast discrimination and TF discrimination thresholds are shown as a function of pedestal contrast, using 1 cpd gratings drifting with a base TF of 5 Hz. Contrast discrimination improves at high contrasts, while TF discrimination does not. (Reproduced from Bowne, 1990).

& Welch, 1985; McKee et al., 1986). This would seem to indicate that Bowne's TF results are really speed discrimination rather than TF discrimination results.

Let us assume for the sake of argument that Bowne's data are accurate measures of speed discrimination performance *in general* (several concerns with the generality of his data will be discussed below). This would indicate that the speed discrimination results obtained in all of the reports described in Chapters 2 and 3 are due to noise in central processes, and not to peripheral noise. Furthermore, it would make it very difficult to apply the data obtained on the properties of the spatial and temporal channels (Chapter 2) and the models of speed perception (Chapter 3), let alone new data. It is thus important to determine whether speed perception is always limited by central noise, or whether under some conditions at least peripheral
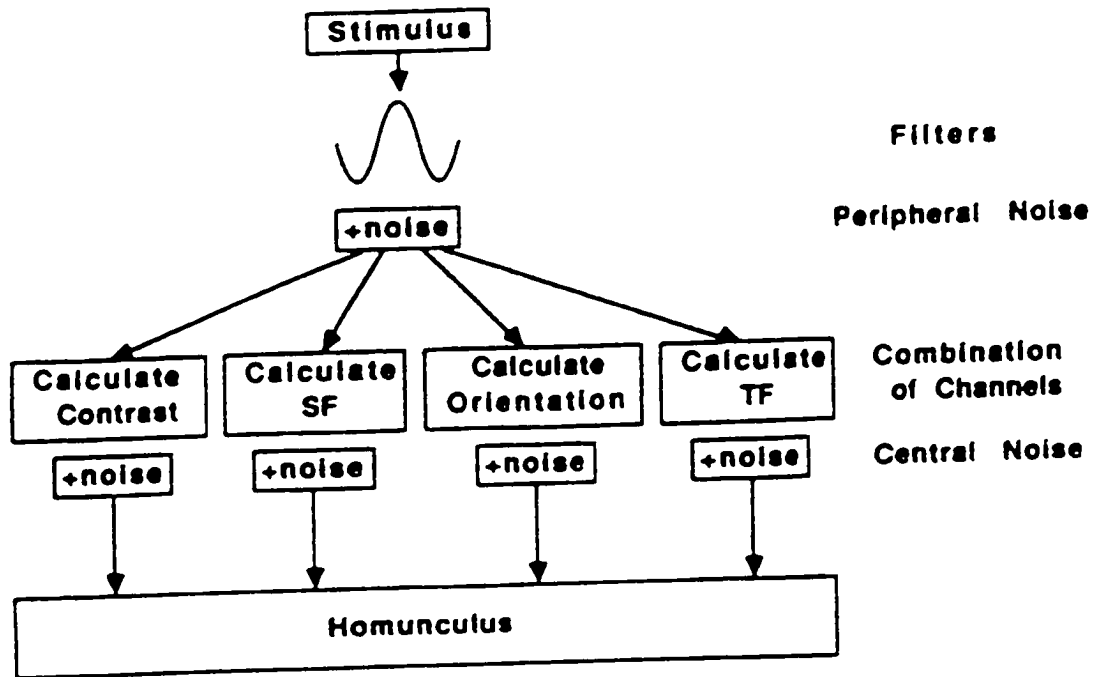
Figure 4.4: A realistic model of discrimination, including both peripheral and central noise sources. The stimulus is first passed through a number of filters with outputs which depend on the SF, TF, orientation and contrast of the stimulus. One filter's center-surround spatial sensitivity profile is shown. "Peripheral" noise is added to each filter's output. The outputs of these filters enter combination processes, which extract estimates of stimulus parameters such as SF, TF, orientation and contrast. Central noise is added to each of these estimates to form the decision variables for discrimination tasks. Because the central noise may be different for each task, measurement of contrast discrimination does not provide enough information to predict other discrimination thresholds. The box at the bottom labeled "homunculus" represents all higher mental processes which use the decision variable to select a response. Any noise added by the homunculus is assumed to be negligible compared to the explicit noise sources shown. (Reproduced from Bowne, 1990).

80

noise can be the limiting noise.

There are some methodological concerns with Bowne's experiments which need to be considered before his data taken to mean that peripheral noise is not the limiting noise for speed discrimination. The subjects in Bowne's task are Bowne himself and subject SPM, who is an extremely highly trained observer used to speed discrimination tasks. Furthermore, their thresholds for TF discrimination correspond to Weber fractions of less than 10% at $d' = 2$, which is quite good performance (see e.g. McKee et al., 1986; McKee & Welch, 1989; Turano & Pantle, 1985). It is thus likely that as in all psychophysical tasks with highly trained observers, the subjects used all available cues to perform the task as well as possible. Two characteristics of his stimuli should be noted. First, Bowne's displays had fixed durations of 500 msec, with abrupt onset and offset. One consequence of this fact is that the distance traveled by a grating in the displays was perfectly correlated with the grating's TF, thus providing a spatial cue which could at least in theory be more available to the subject than the TF of the stimulus[3]. In addition to having fixed durations, the stimuli used by Bowne had fixed contrasts within a condition. As described in Chapter 2, perceived temporal frequency and perceived contrast can affect one another. The *apparent contrast* of the gratings may thus have correlated with their TF, thus providing another correlate of TF which might have had been less noisy than the TF signal[4]. The experiments presented in this chapter are designed to address these methodological concerns with Bowne's data.

---

[3]The spatial phase of the gratings was randomized, however, so the absolute positions of gratings were not useful cues.

[4]While this last statement seems to imply the central noise limits of Bowne, what is meant by that is that while the signal within a unit is equally noisy regardless of the task, discrimination depends not only on the signal to noise within a mechanism but on the specificity of tuning of sets (at least pairs) of mechanisms.

4.1 Experiment 4.1—Methods

This experiment is a replication of Bowne's TF experiment except that the two methodological concerns discussed above were taken care of: the duration of the stimuli is jittered around the mean duration in a manner uncorrelated with the speed of the stimulus, so that the distance traveled by a grating cannot be used as a cue to speed. Also, as discussed in Chapter 2, perceived speed is influenced by contrast and vice versa (Thompson, 1982; Stone & Thompson, 1992), so contrast was also jittered around the mean value within a block (and between the two intervals of each trial), so that the effect of contrast on perceived speed could not be a good cue to speed. Both of these jitters were of $\pm$ 20%, so any speed discrimination threshold less than 40% cannot be due to either of these two known covariates of perceived speed. Both duration and contrast jittering was used in all experiments described in this dissertation, except where otherwise noted. The effect in contrast is thus one exhibited as the mean contrast changes—while the contrast in each condition was jittered across trials, it was jittered around a mean which was changed depending on the condition.

4.1.1 Stimulus Generation

All stimuli used in this dissertation were sinewave gratings displayed on a γ-corrected Tektronix oscillocope model 608 equipped with a P^{31} phosphor. Mean screen luminance was measured with a SpectraScan photometer at about 20 cd/m^2 [5]. The gratings were generated by an Innisfree, Ltd. Picasso image generator under computer control via a set of digital to analog and digital I/O interface cards. The

---

[5]The precise mean luminance of the display depends on the orientation of the scope relative to the dim light in the room and the subject, as well as such minor effects as the color of the clothing worn by the people present in the room.

Picasso outputs one or two gratings on a given oscilloscope frame. The frame rate was set at 100 Hz for all of the experiments reported here. The contrast, spatial frequency, temporal frequency and orientation of each grating is set at each frame, thus allowing precise control over all relevant stimulus characteristics.

The oscilloscope screen has a rectangular shape, with dimensions 10 cm × 13 cm, but in this experiment, a circular aperture made of black matte cardboard and with a diameter of 9 cm was placed in front, thus limiting the effective screen size. A small fixation dot was affixed to the center of the oscilloscope in some of the trials, and had no effect on performance.

Observers viewed the display with their heads stabilized in a chin rest, so that their eyes were at the level of the center of the oscilloscope. The room was dimly lit throughout the experiments. Two observers were used for this experiment: subject DA, the author, a myope wearing corrective optics, and subject LW, an emmetrope. In all experiments, observers were allowed to set their own pace by taking breaks between blocks.

4.1.2   Task

The method used to obtain speed discrimination thresholds in this experiment is the method of single stimuli. In each block, a set of 150 trials was presented to the observer. A trial consisted of two intervals in which moving gratings were displayed. The observer's task was to indicate by means of a mouse button press which of the two intervals contained the stimulus which moved faster. Incorrect responses were automatically indicated to the observer by a computer beep. In both intervals of all trials of this experiment, the spatial frequency content of the stimuli was identical. Speed was manipulated by varying the temporal frequency of the gratings. The relative changes in speeds of the two intervals varied across trials, so that 30 trials

83

had a speed difference of $N\%$, 30 trials had a speed difference of $2N\%$, etc. up to $5N\%$, where N was chosen so as to yield good fits with the psychometric function (described below). In all cases the mean speed of the two stimuli across intervals was the same, so that for the trials with a speed difference of 20%, one interval contained a stimulus which moved at 90% of the mean speed, and the other a stimulus which moved at 110% of the mean speed. The order of the trials was randomized within a block, and which of the two intervals in a given trial contained the fastest stimulus was also randomized across trials, so that on a given trial, the observer had no way of predicting which interval contained the fastest speed or what the speed difference was.

Responses were tabulated and the percentage of correct discriminations was computed for each of the speed differences. These percentages were then fit with a Weibull function[6], yielding the estimated change in speed needed for a 75% correct discrimination performance, corresponding to a $d' = 1$.

In all conditions, stimuli lasted 120 msec on average, with the $\pm 20\%$ duration jittering mentioned above. The gratings were 1 cpd gratings moving at 5 Hz. Contrast was chosen to be one of 2.5%, 5%, 10%, 20%, or 40% depending on the condition. One or two blocks of 150 trials were taken for each data point.

4.2 Results

The mean thresholds for each subjects as a function of contrast are plotted in Figure 4.5. In this and all subsequent plots, except if noted otherwise, the error bars indicate the standard error of the mean. The results are virtually identical between the two subjects, with subject LW's thresholds being consistently lower than DA's.

---

[6]The Weibull fitting algorithm was provided by Sam Bowne.

Speed discrimination performance is quite poor at very low contrast. reaches its best performance (lowest threshold) around 10% contrast. and appears to worsen slightly for very high contrasts.
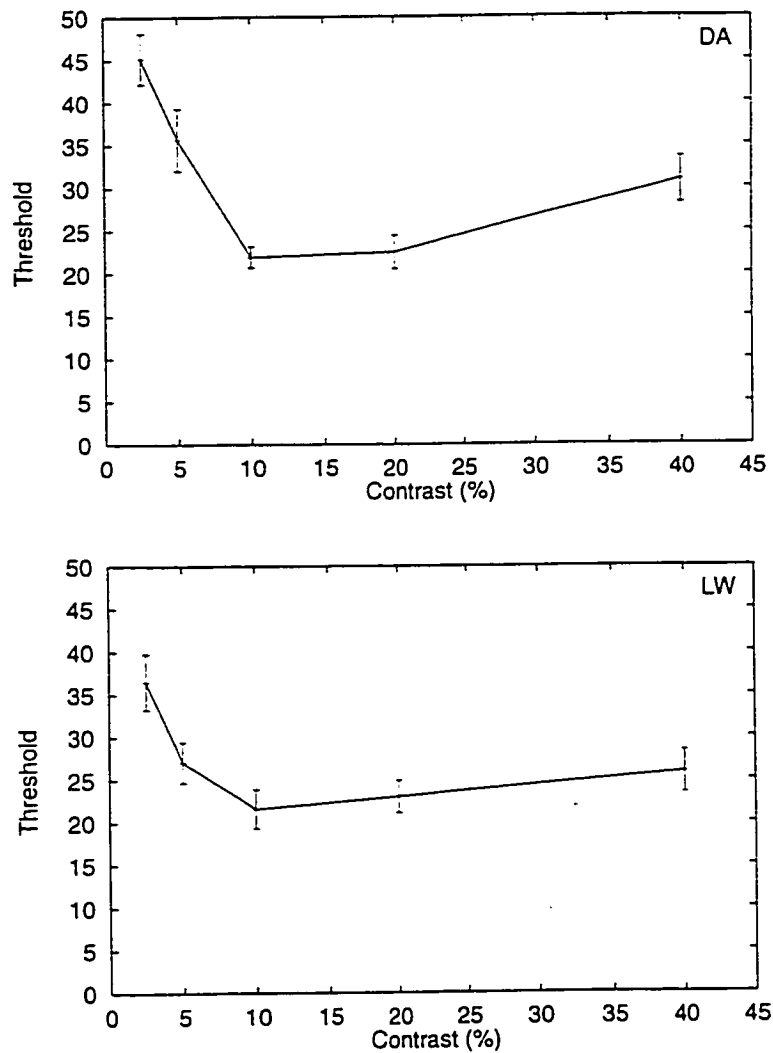
Figure 4.5: Speed discrimination threshold (in percent) plotted as a function of stimulus contrast for 1 cpd, 5 Hz gratings presented for 120 msec. In this and all subsequent plots, error bars indicate ± one standard error of the mean.

4.3 Discussion

Two minor aspects of the data will be considered first. First, the absolute level of the thresholds corresponds to much worse performance than that reported in many studies of speed discrimination (McKee et al., 1986; McKee & Welch, 1989; Turano & Pantle, 1985). This is no doubt because of the extreme conditions employed here. The gratings are visible for on average 120 msec, which is quite brief. Furthermore, the duration is jittered, thus affecting perceived speed (Giaschi & Anstis, 1989). Finally, at low contrasts, the gratings are not very visible. One concern one might have is that the gratings are in fact at threshold for the lowest contrasts used in this experiment. To test this hypothesis, a simple QUEST procedure (Watson & Pelli, 1983) was used to measure the detection contrast for gratings as were used in this experiment, under the same conditions. For observer DA, that contrast was 1.15%. For subject LW, that contrast was 0.97%. Thus even the lowest contrast used was at least an octave superthreshold, arguing that the contrast effect (found even for 5% gratings) is a true contrast effect and not a detection artifact.

The second point concerns the increases in threshold for 40% contrast. While this has been seen before (Welch, personal communication), this effect is hard to explain, and would require further study to make sure that it is reliably reproduceable. To account for such a result requires a mechanism whose response as a function of contrast saturates at high contrasts. With such a mechanism, differential responses to, e.g., slow and fast stimuli would yield progressively more similar responses in speed-tuned mechanisms as the contrast response is compressed in the saturating mechanism.

As can be readily seen in Figure 4.5, with increasing contrast, thresholds decreased by approximately 35%–50%, showing that in our experimental conditions, speed discrimination performance is in all likelihood limited by what Bowne terms

peripheral noise.

How can this contrast data be reconciled with Bowne's data, which show quite clearly that performance on a TF discrimination task being limited by central noise instead of peripheral noise? Two of the differences between his experiment and ours are most likely not the source of the differences in the results. The fact that we used a different threshold estimation technique is unlikely to affect the thresholds (in pilot experiments, various techniques were used yielding similar data—the procedure described above was chosen because it yielded stable data). Also, Bowne used a 4.5° aperture at 114 cm from the observer, while the observers in this study viewed the stimuli through a 9° aperture at 57 cm. This change is also unlikely to be the source of the effect, since increasing stimulus area should if anything have improved performance, while our data show worse performance than Bowne's. While the absolute Weber fractions are quite comparable, the two tasks yield thresholds for very different d' values. Bowne's use of the QUEST procedure means that his thresholds correspond to a $d' = 2$, while the method of constant stimuli and Weibull fitting used here yields threshold estimates corresponding to $d' = 1$. The thresholds obtained in this experiment thus need to be multiplied by two to be compared with Bowne's. They are noticeably higher than Bowne's, most likely because the jittering of contrast and duration introduces noise in the decision process – furthermore, the short duration used in this experiment reduces the amount of signal available for the discrimination.

There are however three candidate explanations for the differences between the results of the two studies.

4.3.1 Corrections in the methodology

The first explanation is that the methodological issues which motivated this study are responsible for the discrepancy in the presults. According to this explanation, Bowne's subjects might have been doing a spatial displacement task in lieu of the TF task they felt they were doing. The gratings Bowne used were 1 cpd gratings moving at 5°/sec. Thus in 500 msec, they traveled 2.5° of visual angle, about half of the size of the aperture through which they were viewed. If this is the case, then it is the spatial displacement task which was shown to be limited by central noise, a result which would be interesting in its own right but outside the domain of this dissertation. Experiments testing spatial displacement discrimination ability would have to be run to test whether this is indeed the case. Similarly, the apparent contrast of the gratings could have been a cue to speed. It is however unlikely that Bowne's subjects were using it to perform this task. Indeed, examination of his Figure 5 (reprinted above as Figure 4.3) show that Bowne's subjects' ability to discriminate contrast does show a contrast dependence. The contrast Weber fractions are always larger than the TF Weber fractions in Bowne's data—thus for the TF discrimination results to be due to contrast discrimination, the change in perceived contrast due to the indirect effect of TF on contrast must be greater (in proportion) than the effect on perceived contrast of a direct change in physical contrast. Numerically, this would require that for low contrast gratings, an 11% change in TF (which correspond to performance higher than 92% correct for both observers in Bowne's study (see Figure 4.3) must result in a change in perceived contrast *greater* than an 11% change in stimulus contrast (which is not enough to yield the same level of performance in the contrast task for the same two observers). While this is possible, there are no published data regarding such an effect at these TFs. In my hands, the effect of TF on contrast is greatest at very high TFs (indeed, around the critical flicker fusion

88

frequency, small changes in TF can result in massive changes in apparent contrast). Bowne used quite moderate TFs of 5 Hz, thus it is unlikely that his TF discrimination results are due to perceived contrast cues.

4.3.2 Learning Effects

An alternative explanation of the difference between Bowne's data and the data presented here lies with the observers. While all the subjects in both Bowne's study and this study are well trained, it is possible that Bowne's subjects logged more hours than DA and LW *at the specific stimulus configurations used* (the experiments reported in Chapter 8 indicate that indeed some stimulus-specific learning can occur even for well-trained observers). It is possible that Bowne's observers had undergone perceptual learning at the peripheral level, thus pushing the level of the peripheral noise down far enough that the central noise became the limiting noise. While in theory the inverse process could have occured—that Bowne's subjects for some reason increased the level in central noise until it became the limiting noise—this is highly unlikely not only because of their high level of training but because their performance is higher than that obtained here.

Finally, the displays used in this experiment had a mean duration of 120 msec, while Bowne used 500 msec displays. Just as performance saturates for high contrasts. presumably because of signal-to-noise ratio reduction due to high signal strength, it is quite likely that performance can be enhanced by longer durations, until it saturates. This is the topic of the next chapter on temporal integration.

All of these explanations can be summarized by saying that due to either methodological or learning differences, the performance of Bowne's subjects was limited by central noise, while ours appears to be limited by peripheral noise. While the two results appear contradictory, the evidence *for* peripheral-noise limits in discrimina-

89

tion processing is a significant result, while Bowne's results consist of a failure to find an effect. In other words, these results mean that the models based on contrast-modulated signals *can* be used, provided that conditions of high peripheral noise are used. Furthermore, our results indicate that for contrast effects to be found in speed discrimination tasks, it is important to choose stimulus parameters at which the discrimination process is indeed limited by peripheral noise and not central noise (e.g. by using short jittered durations and low contrasts). Almost all subsequent experiments presented in this dissertation use such stimulus parameters, precisely so that the manipulations can yield information on the local speed mechanisms, rather than on later processing. It should be kept in mind that when an effect is not found, it is quite difficult to know whether the lack of effect is due to this central noise problem, or to a true absence of effect of the manipulation on the visual system.

Chapter 5

Temporal Integration in Speed Discrimination

In the previous chapter, evidence for integration of information in a speed discrimination task was found when using short duration, low-contrast gratings (i.e. under experimental manipulations which minimize the signal strength, thus minimizing the signal-to-noise ratio within the putative motion detectors).

While the dependent variable studied in that experiment was stimulus contrast, it was noted that the duration of the stimuli which was used was much shorter than used in Bowne (1990). The experiments presented in this chapter test what the influence of stimulus duration is on speed discrimination thresholds.

5.1   Methods

Subjects LW and DV participated in this experiment. Both are emmetropes.

The same general setup was used as in the experiments reported in Chapter 4. The oscilloscope was covered with a 9° cardboard aperture within which the stimuli were displayed.

The gratings were all 1 cpd gratings with 10% contrasts (jittered by ±20% as discussed earlier). Durations were chosen for each block to be either 120 msec. 200 msec, 300 msec, 400 msec or 500 msec. Duration was jittered by ±20% for subject DV, but not for subject LW. The durations were randomized across blocks, and the two subjects were presented with the duration conditions in reverse order. Speed discrimination thresholds were estimated with the Probit method. In this method, subjects are presented with 150 trials, and are asked to judge whether each stimulus is faster or slower than the mean speed of the stimuli. As this speed can only be estimated by experience with the stimulus set, observers were given some practice runs to allow them to set the criterion for this mean speed. Grating speeds where chosen in a block-randomized fashion to have one of five speed centered around a mean speed, and separated from the next by a constant step in speed. For example, if the mean speed was 2°/sec and the step size was 0.2°/sec, then speeds were chosen from 1.6, 1.8, 2.0, 2.2, and 2.4°/sec (20% of the trials each). The step size was selected so as to yield a good psychometric fit. In the experiments reported in this chapter, the mean speed was always 5°/sec, corresponding to a TF of 5 Hz. The observer indicated whether a stimulus was slower or faster than the mean by using a computer mouse button. Incorrect answers resulted in computer beeps. No answer was considered incorrect for the stimuli which had the mean speed. The results were tabulated and fit with a Probit distribution, which yielded several measures. A χ^2 measure was reported, and was used to throw out blocks which were not well fit by a psychometric function. A point of subjective equality (PSE) indicated the estimate of physical stimulus speed which corresponded to 50% performance given the psychometric function, and was used to check against large shifts in the PSE, of which there were none. Finally, the speed increment corresponding to 75% "fast" responses was also estimated, yielding the threshold change in speed needed for $d' = 0.67$ performance, accompanied by a measure of the standard error of that

estimate. These thresholds and standard errors were tabulated and combined across blocks of 150 trials. All data points correspond to at least 300 trials. Error bars indicate ± the standard error of the threshold.

5.2 Results

The estimates of speed discrimination threshold as a function of stimulus duration are plotted in Figure 5.1. They show that as stimulus duration increase. performance improves. The most significant improvement in performance is between 120 msec and 200 msec for both subjects, but it is possible that thresholds continue to drop until 500 msec, although further statistical analyses would be required to test this rigorously. Subject DV's thresholds are consistently higher than subject LW's. One possible reason for this difference is that LW has logged more hours doing speed discrimination experiments than subject DV. Another probable cause for the difference is that the stimuli presented to DV had jittered durations, which made the task harder.

5.3 Discussion

These results argue that just as when signal strength is modified by controlling stimulus contrast, as in Chapter 4, signal to noise can be increased by using a longer stimulus duration. This is most likely the reason why our contrast results showed discrimination performance limited by peripheral noise while Bowne (1990) failed to. The lack of large change in performance between 200 msec and 500 msec is consistent with the report by (McKee et al., 1986) that Weber fractions did not change significantly between 230 msec and 500 msec stimuli. It is quite possible that integration could be found with lower contrasts, as this would decrease the signal-to-

93

Figure 5.1: Speed discrimination threshold (in percent) is plotted as a function of stimulus duration for 1 cpd, 5 Hz gratings at 10% contrast. The data is for subject DV in the upper panel, and for subject LW in the lower panel.

94

noise ratio, thus possibly allowing more "room" for a temporal integration process to occur.

This duration of between 120 msec and 200 msec is quite consistent with what one would expect from the temporal filters which have been proposed, such as the Alpha functions of Watson (1986). Finally, it is interesting that our data appear to indicate a slight trend towards better performance with longer displays, up to the longest duration tested. This would be consistent with the notion of probability summation in time (Watson & Ahumada, 1985). Further testing would be required to make this determination more precisely.

Chapter 6

Size Characterization of the Speed Mechanism

The last chapter presented results which argue that the temporal integration window of the mechanism responsible for speed discrimination is at least between 120 msec and 200 msec. The experiments presented in this chapter investigate the same question in the spatial domain. Specifically, what is the spatial area within which the speed discrimination system integrates?

6.1   Methods

Subjects LW and CJ participated in this experiment. LW is an emmetrope, and CJ wore his normal optical correction. The methods are as in Chapter 5, with one exception. Since the stimulus parameter under study is its spatial extent, the electronic aperture capability of the Picasso frame generator was used. While observers viewed the full oscilloscope display without a cardboard aperture, the gratings were only displayed within an electronically defined aperture. For subject LW, the aperture was always 1 cm tall and either 1, 2, 4, 8 or 10 cm wide. For subject CJ, the

same conditions were used, as well as a condition where the *width* was kept constant at 1 cm, and the *height* was varied (the maximum height was 9 cm, the height of the oscilloscope. The part of the scope not within the electronic aperture was at mean luminance. In all conditions, the gratings were vertical gratings moving left or right. Thus in the case of subject CJ, half of the gratings were elongated along their direction of motion, while the other half were elongated in the orthogonal direction (see Figure 6.1).

The gratings were all 1 cpd gratings with contrasts chosen randomly from the 2.5%–5% range. Their speed were chosen around a mean speed of 5°/sec (corresponding to a TF of 5 Hz), using the Probit method described in Chapter 5, with a step size of 15%. The display duration was jittered ±20% around 120 msec. All data points correspond to at least 300 trials, except for the data for subject CJ in the vertically elongated condition, which correspond to 150 trials. The error bars indicate the standard error of the threshold.
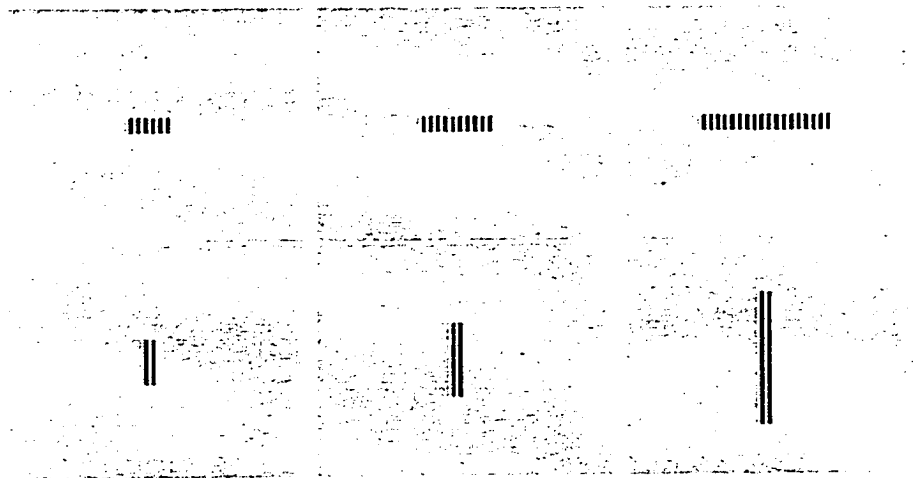
Figure 6.1: Schematic representation of single frames of some of the stimuli used in this experiment. The top panel depicts some of the gratings used by both subjects, while only subject CJ saw the gratings depicted in the bottom panel.

6.2 Results

The estimates of speed discrimination threshold as a function of stimulus size are plotted in Figure 6.2. They show that as stimulus size increases, performance improves. No asymptote is visible in LW's data, for stimuli up to 10 cm wide and 1 cm tall. Subject CJ's data is similar to LW's. The thresholds for tall displays are consitently higher than for long displays.

6.3 Discussion

The only report on the size and shape of the mechanisms responsible for speed perception is an abstract by (Bowne, 1990), who compared speed discrimination threshold for vertically elongated horizontally moving gratings vs. horizontally elongated horizontally moving gratings. His target gratings had a SF of 1 cpd, variable TFs around a reference TF of 5 Hz, and were displayed for 500 msec at very low contrast (0.1–1%). The mask gratings had randomly chosen SFs between 0.7 and 0.9 cpd, a contrast of 1%, and a fixed TF of 5 Hz, and were on for the entire 500 msec. Two comparisons should be made between Bowne's data and the data just reported. Firstly, our results replicate his finding that elongation in the horizontal direction results in higher gains in performance than elongation in the perpendicular direction. Secondly, one may wonder why we found an effect without a mask, while Bowne states in his abstract(Bowne, 1990) that without a mask he found no effect. One likely reason for this difference is that, as discussed in Chapter 4, 500 msec displays may be too long to allow the detection of peripheral noise effects.

What can we conclude from these results? The conclusion which Bowne drew from his results is worth repeating here. Bowne felt that the elongated nature of the speed discrimination mechanism was designed to allow for efficient integration
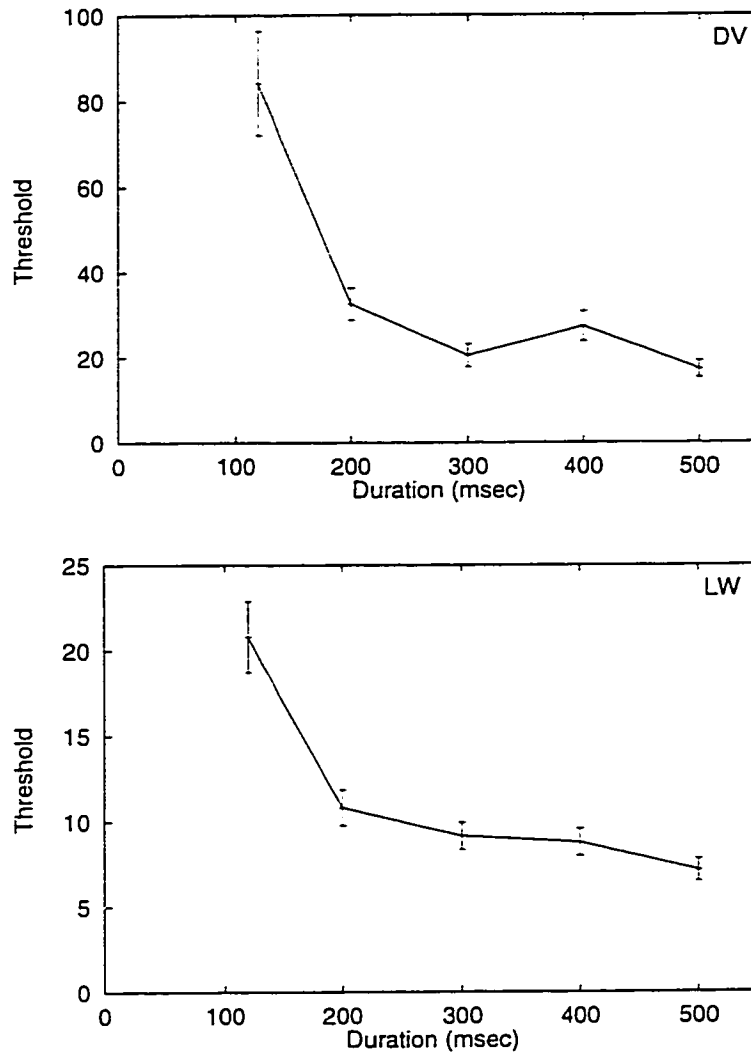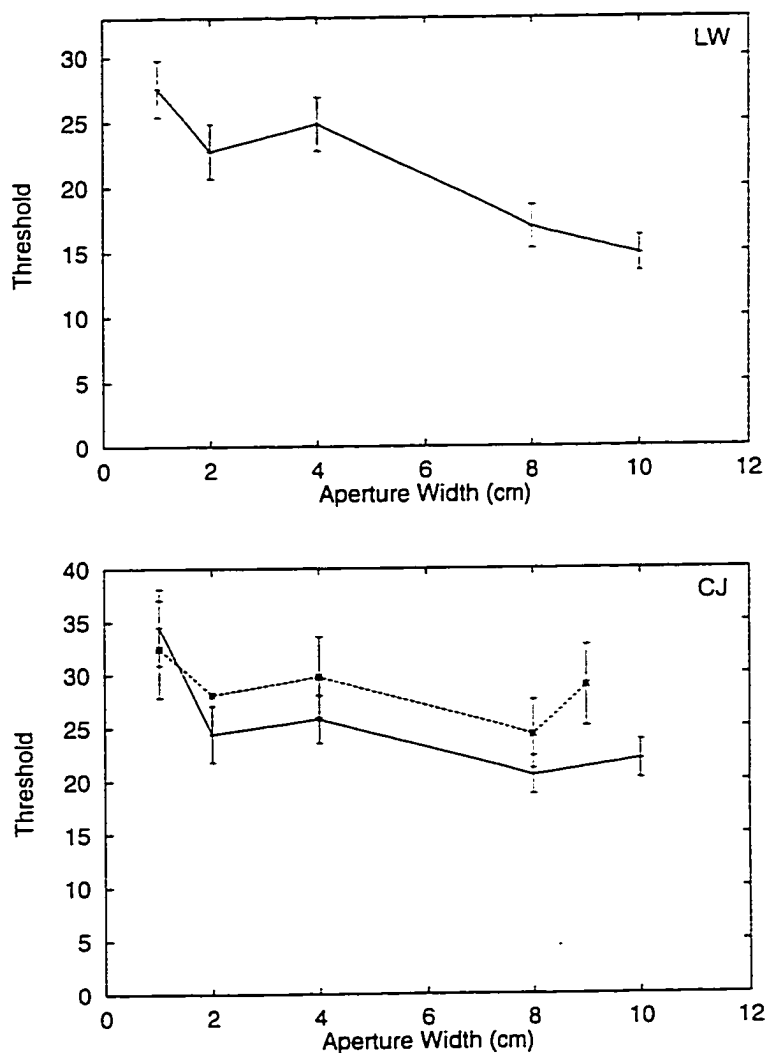
Figure 6.2: Speed discrimination threshold (in percent) is plotted as a function of stimulus size for 1 cpd, 5 Hz gratings at 10% contrast presented for 500 msec. The data is for subject LW in the upper panel, and for subject CJ in the lower panel. The dashed line indicates the vertically elongated conditions, the solid line the horizontally elongated conditions.

99

of the motion signal over the trajectory of the grating, so that speed information could in theory be made more reliable as the object was "tracked." This idea can be related to a set of recent psychophysical results which argue for similar notions (e.g. Burr, Ross, & Morrone, 1986; Watamaniuk & Sekuler, 1992; McKee & Welch, 1985; Turano & Pantle, 1989; Werkhoven, Snippe, & Toet, 1992; Watamaniuk, McKee, & Grzywacz, 1994; Grzywacz, Watamaniuk, & McKee, 1995). Regardless of whether speed information is integrated especially efficiently along trajectories, one might wonder what the exact shape of the speed sensor is. The fact that the larger the stimulus, the lower the threshold should surprise no one. This is true for spatial frequency discrimination tasks as well (e.g. Anderson & Burr, 1987), yet does not mean that the units one would wish to call SF units are infinitely large. The experiment just described used the same general approach as that used by (Anderson & Burr, 1987). Their data allowed them to identify the receptive field size at which the *nature* of the integration mechanism shifted from including within-receptor integration to just being cross-receptor integration. The data presented above is not finely sampled enough to allow for such a "knee-finding" operation, although the method above could clearly be used to perform such an experiment, as a large effect size is seen.

One interesting topic for further research would be to determine whether the size and/or shape of speed-sensitive units varies with spatial frequency, temporal frequency or speed. According to the psychophysical and computational results mentioned above, which argue that speed computation is especially sensitive to trajectories, one might expect that fast speeds should be measured most efficiently by very elongated units, while slow speeds could be measured adequately by units with little elongation in the direction of motion.

Chapter 7

Effects of Adaptation on Perceived Speed

7.1 Introduction

At this point, the ability of the speed discrimination system to integrate over signal, time and spatial extent has been demonstrated. While the data in Chapter 6 argues for integration across cells with distinct receptive fields, and the data in Chapter 5 suggests integration in time, no evidence has been presented so far arguing for integration across the spatial and temporal filters described in the earlier chapters. The experiments presented in this and the following chapter were designed to investigate this question, along the temporal frequency and spatial frequency axes respectively. Chapter 2 presented the first stages of motion filtering as being performed by spatiotemporal filters which are assimilated to a set of channels with two key characteristics, the channel's spatial and temporal tuning. The models presented in Chapter 3 are of special interest in this context because most of them propose that the computation in speed consists of the comparison of outputs of several temporal channels. For example, the Smith and Edgar (1990) model proposes that speed is

computed by the ratio of a low-pass and a band-pass channel. The Grzywacz and Yuille (1990) model views speed as the result of a computation based on the outputs of large numbers of filters. The "estimation" strategy described in that model can be viewed as the computation of speed over a potentially large set of temporal channels.

In a trivial way, finding evidence for integration across temporal channels in speed perception is a *fait accompli*, as the very studies which identify the temporal channels make use of speed discrimination experiments. The somewhat more specific question which is addressed here is whether information from all temporal channels is used in speed discrimination, and if so how this integrative process occurs. While this type of endeavor is fairly easy to do in the domain of spatial channels, due to their fairly narrow tuning, temporal channels are so broad that isolating any small number of them is made much more difficult. The proposed temporal channels have been characterized as lowpass or bandpass. Interestingly, while there is evidence for multiple bandpass channels, no one has proposed multiple lowpass channels. The lowpass channel is generally attributed the role of stationary, "pattern" vision (Kulikowski & Tolhurst, 1973), while the bandpass channels are viewed as the core of the motion system. Indeed, since their sensitivity does not drop to zero for 0 Hz, it is possible that speed computation is based solely on the output of the bandpass channels. Such a finding would have significant implications for the models of speed which currently propose a role for the lowpass channel.

How can one test for the involvement of the lowpass channel in speed discrimination? In the Grzywacz and Yuille (1990) model for example, the lowpass channel can be viewed as just another temporal channel, with no special status. While involvement of the low-pass mechanism appears innocuous from the computational point of view, it leads to an interesting experimental prediction—if the low-pass mechanism is truly low-pass, and if it is an active participant in the speed perception process, then it might be adaptable by stationary stimuli, and such adaptation should affect

102

the perceived speed of subsequently presented stimuli. Thus looking at a stationary display could affect perception of motion, a somewhat unintuitive prediction. This hypothesis is the main focus of the experiments presented in this chapter.

7.2   Predictions and Previous Evidence

To examine what one might expect the consequences of adaptation to stationary stimuli to be, let us consider first a temporal frequency estimation model similar to that presented by Smith and Edgar. For the sake of simplicity, we will assume two mechanisms, one low-pass and one band-pass. As discussed in Chapter 2, there is disagreement in the literature on the number of band-pass mechanisms (see e.g. Smith & Edgar, 1991; Mandler & Makous, 1984), but the argument presented here can easily be adjusted to a model with more than one band-pass mechanism. Given two broadly tuned filters as described in Figure 7.1, the temporal frequency of the stimulus can be obtained by comparing the relative outputs of the two filters. Smith and Edgar used a ratio to perform this comparison, and the model presented by Grzywacz and Yuille uses a more complex plane-fitting process, but the basic result is the same. If one considers "fast" temporal frequencies (as indicated by F in Figure 7.1a), then the response of the band-pass channel is large and that of the low-pass channel is small. On the other hand, for "slow" temporal frequencies (indicated by S in the Figure 7.1), the low-pass response is large and the band-pass response is small. In general, the ratio of the band-pass response over the low-pass response is monotonically related to the temporal frequency, and if one assumes a fixed spatial frequency, to speed. This is the intuition behind the Smith and Edgar model. Now, let us consider what happens if one can selectively adapt the band-pass channel. After such an adaptation, the relative output of the band-pass channel for any stimulus in its bandwidth would be less than without such an adaptation (as indicated by

the dashed sensitivity curve in Figure 7.1b). Thus. the ratio described above would be smaller than it would be without the adaptation, corresponding to a decrease in perceived speed. If one could selectively adapt the low-pass channel. then one should reduce the relative output of the low-pass channel (see Figure 7.1c), which in turn should result in an increase of perceived speed.

A review of the literature on the effects of adaptation on speed perception shows that effects such as those described above have indeed been found, but that some effects which one would expect given the theory have not been reported.

The earliest investigations of the effect of adaptation to moving stimuli on perceived speed established that the perceived speed of a moving stimulus decreased with adaptation duration (Wohlgemuth, 1911; Gibson, 1937). Carlson (1962) investigated a broader range of stimulus parameters, and found that when the adapting stimulus was faster than the test stimulus and in the same direction, the test stimulus appeared slower. Scott, Jordan, and Powell (1963) extended this result to motion in opposite directions, and found that when the test stimulus was in the opposite direction from the adapting stimulus, it appeared to move faster than without adaptation. In an unpublished dissertation cited in Thompson (1981), Clymer (1973) reports that after adapting to moving gratings, subsequent gratings appeared to move slower, except in the case of slowly moving gratings, which caused test gratings to appear to move faster.

Thompson (1976, 1981) studied speed aftereffects under a broad range of viewing conditions. In most cases, adaptation to a moving grating caused subsequently presented gratings to appear to move slower than they would otherwise. This is consistent with the adaptation of the band-pass channel, as described above. Only when the test grating moved in the opposite direction from the adapting grating, and when the test grating moved at a temporal frequency under 1 Hz did the test grating appear to move faster than it would without the adaptation. Thompson reported in

104
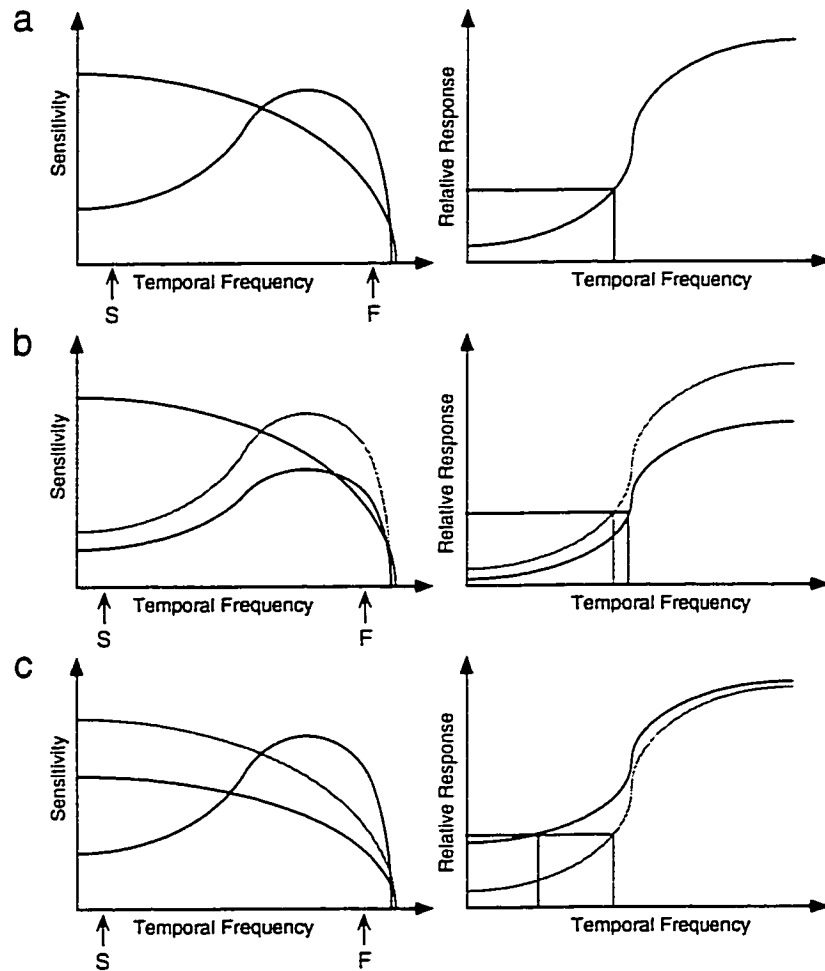
Figure 7.1: In each of these three panels, the left graph displays the relative sensitivity of a lowpass and bandpass mechanism, while the right graph schematizes the relationship between the relative response and the temporal frequency. S and F indicate slow and fast temporal frequencies respectively (see text). (a) Schematic representation of the relative sensitivities of the low-pass and band-pass channels before adaptation. The readout procedure from a given relative response to the corresponding temporal frequency is indicated. (b) Sensitivity curves of the low-pass and band-pass channels after selective adaptation of the band-pass channel (the unadapted functions are shown in grey). The effect of adaptation is to decrease perceived speed, as indicated in the right panel. (c) Sensitivity curves of the low-pass and band-pass channels after selective adaptation of the low-pass channel. The effect of adaptation is to decrease perceived speed, as indicated in the right panel.

105

a personal communication (1996) that he had tried to elicit increases in perceived speed using stationary gratings. but failed to obtain significant results.

The experiments presented by Smith and Edgar (1994) showed a similar set of results: after adaptation to fast gratings, most gratings appeared to move slower than they would otherwise, especially slow gratings. After adaptation to slower gratings (down to 2 Hz), slow gratings still appeared to move slower than they would otherwise, but fast gratings appeared to move faster than they would without adaptation. Smith and Edgar fit their ratio model using both multiplicative and subtractive adaptation, and found that the former always fit the data poorly. while the latter could result in a good fit, but only if the adaptability of the band-pass filter was assumed to be about an order of magnitude higher than that of the low-pass filter.

Given the relative simplicity of the general model, it is worth considering why one prediction (that adapting to high TF stimuli should yield decreases in perceived speed) is readily obtainable, while the other (that adapting to very low TF stimuli should yield increases in perceived speeds) is not easily elicited, or is only elicited for high test speeds. Given the number of studies of the motion aftereffect, there is likely an experimental reason why this effect has not been reported before. If we assume that the low-pass channel does exist, then the reason for this gap in the literature has likely to do with the requirement that the low-pass channel be adapted *relatively more* than the band-pass channel. To determine under which conditions this differential adaptation might be obtained, the data of Anderson and Burr (1985) are illuminating—their summary graphs are reprinted in Figure 7.2. The sensitivity functions of the two channels are depicted for a range of temporal frequency, and for three different spatial frequencies, from relatively low (0.1 cpd) to relatively high (10.0 cpd) spatial frequencies. Two facts are worth noting. The first is that while the band-pass channel clearly has a preferred range of temporal frequencies between

106
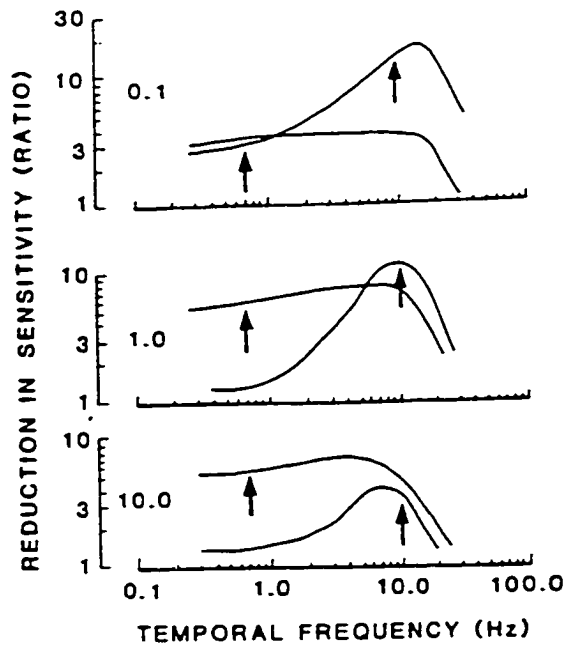
Figure 7.2: Schematic functions depicting the sensitivities obtained via masking experiments of the lowpass and bandpass systems as a function of temporal frequency, for three different spatial frequencies (0.1, 1.0 and 10.0 cpd). From Anderson and Burr (1985)

2 and 20 Hz, it has a non-zero sensitivity down to quite low temporal frequencies. Thus in order to minimize the adaptation of the band-pass channel, it is important to use as low a temporal frequency as possible. The second fact is that the relative sensitivities of the low-pass and band-pass channels at low temporal frequencies is highly dependent on the spatial-frequency being used. Specifically, the higher the spatial frequency, the greater the ratio between the low-pass channel sensitivity and the band-pass channel sensitivity.

These two facts in effect dictated the stimulus parameters which are most likely to yield the desired effect, and which were used in the experiments reported here. To maximize the difference in sensitivity of the two channels (more precisely to maxi-

107

mize the low-pass sensitivity while minimizing the band-pass sensitivity). stationary gratings and a relatively high spatial frequency (5 cpd) were used[1].

It should be noted that one cannot necessarily predict from the sensitivity curves in Anderson and Burr (1985) the results of an adaptation experiment. Indeed, the different channels could have different adaptability, meaning that even equating for sensitivity, one channel might be more susceptible to adaptation than another. This is the argument which Smith and Edgar appealed to in explaining their model fit.

The experiments presented in this chapter aim to test whether one can selectively adapt a low-pass mechanism, and thus test whether this mechanism is an active participants in the computation of speed. What types of results would provide effective data for such a hypothesis?

It is not sufficient to show that after adaptation, gratings appear to move faster than without adaptation, since this could be due to the selective adaptation of a band-pass mechanism which had a lower "peak TF" than another band-pass mechanism[2]. Thus such a result in isolation would not be evidence for the selective adaptation of the low-pass mechanism, or of its contribution to the perception of speed. In general, the ratio of perceived speed after adaptation over perceived speed before adaptation is hard to interpret in isolation because of the lack of precise knowledge regarding the number of temporal mechanisms and their tuning functions.

Similarly, it is not sufficient to show that slow gratings viewed after adaptation to

---

[1]While our stimuli were stationary, no effort was made to have complete retinal stabilization. Such a stabilization would have yielded extreme reduction in apparent contrast as a result of photoreceptor adaptation.

[2]Clearly if this adaptation occurred with stationary gratings, then the mechanism could be said to have low-pass characteristics—however if this mechanism is more sensitive to high TFs than to stationary gratings, such as the band-pass mechanisms described in Figure 7.1, it would not correspond to the "low-pass" mechanism, which has a monotonically falling sensitivity function as a function of TF.

stationary gratings appear to move faster than slow gratings after adaptation to fast gratings. While this would be evidence for adaptation of the low-pass mechanism, this adaptation could occur over the entire temporal frequency range. In this case, one would expect slow gratings to speed up after adaptation more (proportionally) than faster gratings, regardless of the temporal frequency of the adapting stimulus.

To demonstrate selective adaptation of the low-pass system due to low temporal frequency adaptation, one needs to show that as one raises the adapting temporal frequency from 0 Hz (stationary gratings) to higher values, the speeding-up effect of adaptation on the perceived speed of slow gratings decreases, while the slowing-down effect of adaptation on the perceived speed of fast gratings increases. This is more easily explained by referring to a schematic data figure. Figure 7.3 displays the required trend in the data. The y-axis corresponds to the ratio of perceived speed after adaptation to the perceived speed for the same stimulus before adaptation. Values greater than 1 indicate increases in apparent speed, while values less than 1 indicate decreases in apparent speed. While the relative position of the data points relative to the $y = 1$ line is not critical to the argument, the change in slope as one increases adapting temporal frequency from stationary to a higher temporal frequency is crucial.

7.3 General Methods

Subjects DA and LW were the subjects used in this experiment.

The adapting and measurement procedures were the same for all adaptation experiments. A trial consisted of an adaptation period, followed by alternating matching and top-up adaptation periods, as described below. The top-up adaptation periods are required since the measurement phase lasts long enough that without them the adaptation state would decrease while the measurements are being made.

109
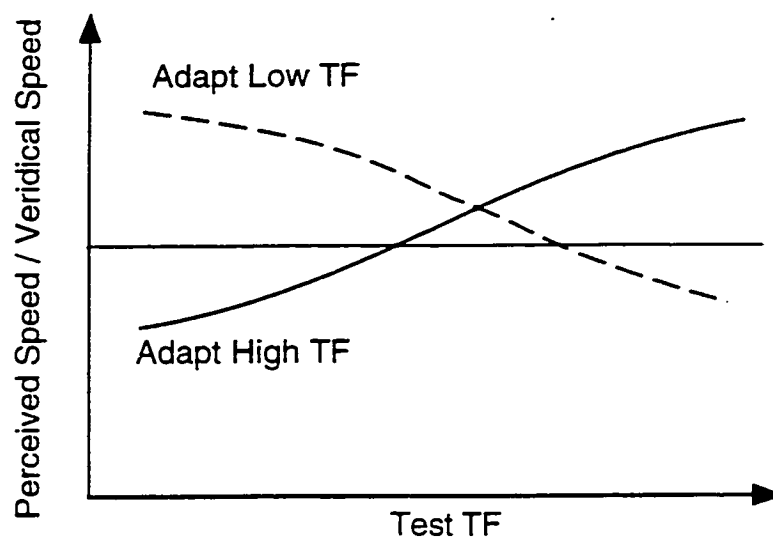
Figure 7.3: Ideal data showing selective low-pass filter adaptation. The solid curve, labeled "Adapt High TF", shows a decrease in perceived speed after adaptation to high temporal frequencies, especially for low test temporal frequencies. After adaptation to low TF stimuli (dashed curve), a perceived speedup is expected. The relative position of the lines relative to the veridical speed percept (indicated by the gray line) is not crucial—the change in slope between the dashed and solid lines is.

The adaptation period lasted two minutes, during which time the subject was instructed to keep fixation on a narrow strip of dark cardboard which bisected the oscilloscope vertically. To minimize the effect of retinal adaptation, the subjects were told to keep fixation on the cardboard strip, but were allowed to move their gaze up and down within the area of the strip. During this adaptation period, the adapting stimulus covered one half of the display, while the other half of the display consisted of a mean-luminance blank field. When the adapting stimulus was moving, it was always moving away from the center of the display.

In the match period, the display consisted of two vertical gratings covering one half of the display each, and moving away from the central dividing line. The subject's task was to decide which half of the display was moving faster than the other. Six one-second displays were used, after each of which the subject indicated which

110

side moved faster using one of two buttons on the computer mouse. The user's response was used by the computer to guide a staircase procedure detailed below. After six such one-second displays, the computer switched to a 15-second topup period in which the adapting stimulus was presented again, in order to maintain the adaptation level as constant as possible (the total duration of the adjustment period is somewhat longer than six seconds due to the delay in user response). The topup stimulus was identical to the adapting stimulus except that it was displayed for 15 seconds instead of 2 minutes. Distinctive tones were played through the computer's speaker to indicate to the subject which phase the experiment was in.
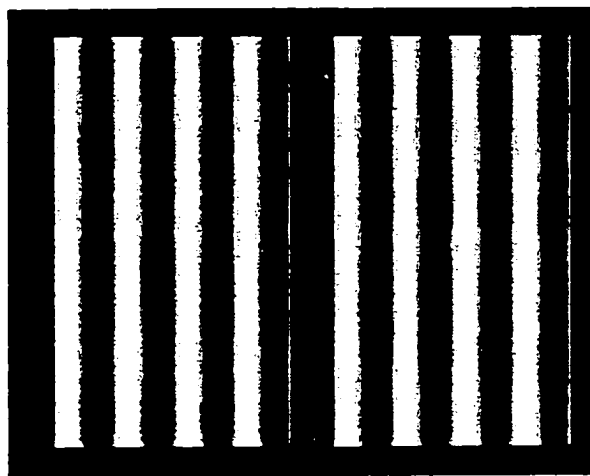
Figure 7.4: Schematic representation of the stimulus.

The alternation between test and topup periods lasted until the staircase procedure finished, which in most trials happened after one or two adjustment periods.

The staircase procedure consisted of a double staircase. One staircase started at a value 50% higher than the match, while the other started at a value 50% lower than the match. The step size for each staircase started at 30%, and halved at each staircase reversal, but was not allowed to go below a 5% change (in other words step sizes for each staircase followed the sequence 30%, 15%, 7.5%, 5%). The entire procedure was considered terminated once both staircases had undergone at least 5

111

Subject	Test Freq. (Hz)	Matched Freq. (Hz)	Standard Error
DA	2	2.00	0.02
	4	4.27	0.10
	8	7.95	0.16
LW	2	2.07	0.02
	4	3.75	0.07
	8	7.58	0.07

Table 7.1: Pretest: Matched temporal frequencies with no adaptation, and associated standard errors, for both subjects.

reversals each at the smallest step size.

7.4 Pretesting: Experiment 7.1

In order to measure changes in perceived speed, a pretest condition was run, which consisted of the same matching procedure as outlined above, but without the preceding adaptation period. Three matches were obtained for each of the three test temporal frequencies, for each observer. In this and all experiments in this chapter, the order of the blocks within an experiment was randomized within a subject, and the other subject ran the conditions in reverse order, to both minimize the size of, and help detect, any learning effects or other ordering effects. No such effect was found. In addition, the grating under user control in the matching task was randomly the left or right grating, to detect any asymmetries. No such asymmetry was found in either subject. The results for the pretest are indicated in Table 7.1.

The results of the pretest indicate that on average, the subjects are accurate in their matches of TF, especially for the low TF, which is of specific interest in this chapter. All of the matches reported in this chapter will be reported after normalization by the corresponding pretest match listed in Table 7.1. Thus a perceived speed of 1.0 would indicate that the match was exactly that obtained in the pretest, a perceived speed of 2.0 would mean that the manipulation had resulted in a twofold

112

speedup. It is important to remember that these are normalized values. thus are unitless.

7.5 Effect of Adaptation to Stationary Gratings: Experiment 7.2

This experiment was designed to establish the effect on perceived speed of adaptation to gratings with the same spatial frequency as the test gratings. as a function of adapting grating temporal frequency. The adapting gratings were at 0, 1, 2, 4 and 8 Hz. The test gratings were at 2, 4, and 8 Hz. All gratings in this experiment had a spatial frequency of 5 cpd and were vertically oriented. The data are depicted in Figure 7.5.

7.5.1 Results

The perceived speedups and slowdowns in Figure 7.5 are rich in effects. Let us consider first the effect of adaptation to 4 Hz gratings (black bars). After adaptation to such fairly slowly moving gratings, slow gratings appear even slower, while fast gratings are not affected. This is consistent with and a replication of some of the previously published results (e.g. Mandler & Makous, 1984), and is consistent with adaptation of a bandpass channel. Contrast this data with the perceived relative speeds after adaptation to stationary (0 Hz) gratings (white bars). After adaptation to stationary gratings. all gratings appear to have sped up. This could be due to two factors. Either the low-pass channel was adapted (as was our aim), or the "most lowpass" of the bandpass channels has been adapted. However, if the latter were the case, then the size of the effect should increase with test TF (as for 4 Hz adaptation). That the effect does not increase is a sign that the bandpass channel is

113
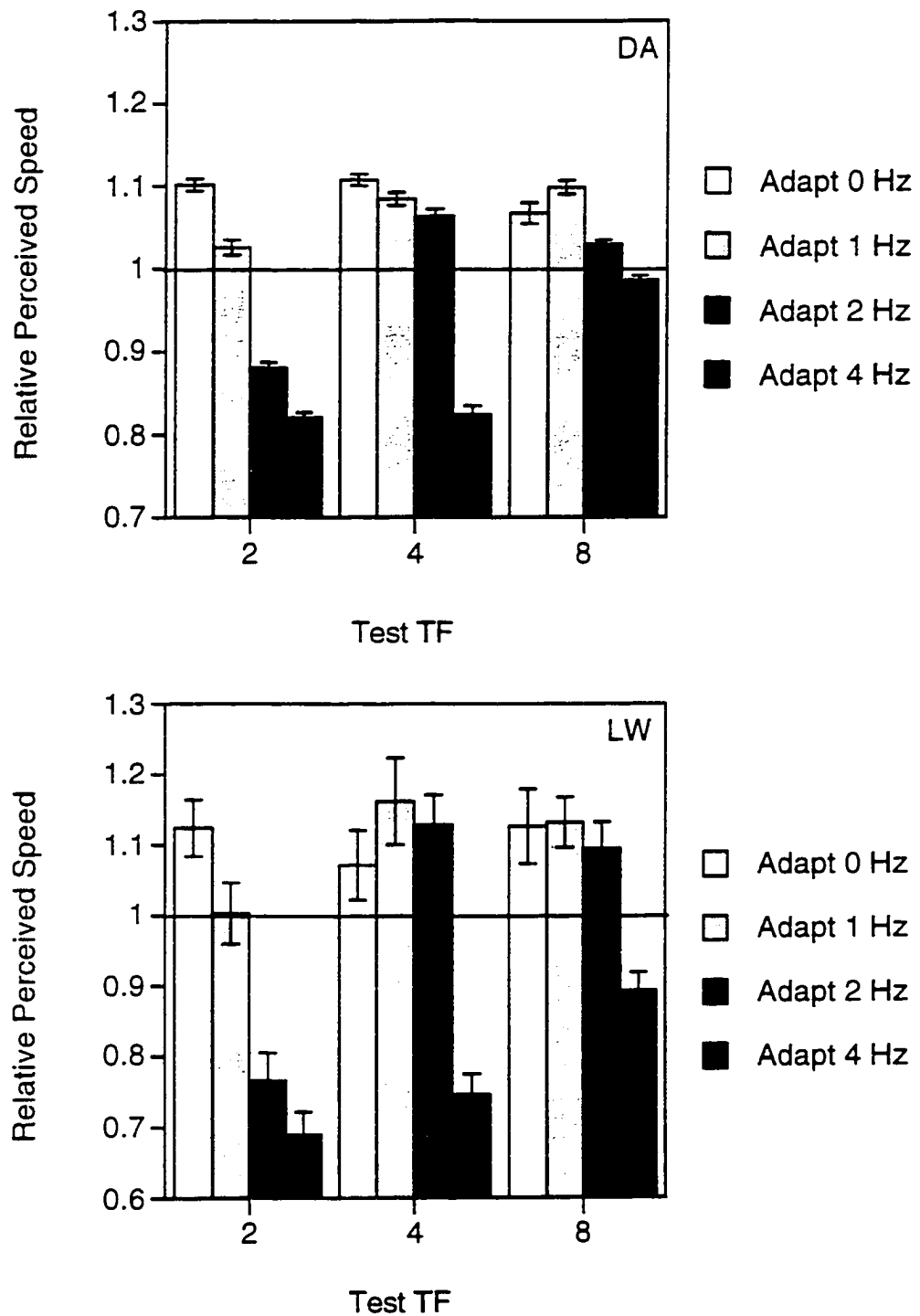
Figure 7.5: Perceived speed ratios for observers DA and LW.

114

not responsible for that effect. One should expect the size of the effect to decrease as the lowpass channel sensitivity (and hence adaptability) falls off. Our data indicate a slight trend in the case of subject DA, and no significant change for subject LW. This would indicate that the adaptability of the lowpass channel is fairly constant up to at least 8 Hz. This is consistent with other results from Welch & Jackson (personal communication) showing sensitivity of the lowpass channel up to 20 Hz in masking studies. The results for adapting TFs between 0 and 4 Hz are, appropriately enough, intermediate between the two patterns just described.

7.5.2 Discussion

The data in Figure 7.5 show some of the required effects.

- Adaptation to stationary gratings results in apparent speedups.

- Adaptation to faster gratings results in apparent slowdowns

However, two predicted effects were not shown. The effect of adaptation to stationary gratings is not significantly largest on slow gratings, and the effect of adaptation to faster (4 Hz) gratings is not larger for fast test gratings than for slow test gratings. The lack of the first effect has already been discussed, as is most likely due to the fact that the lowpass channel's sensitivity does not decrease until very high TF's. The lack of the second effect is most likely due to the fact that the fastest gratings used were still relatively slow, at 4 Hz. This is consistent with the data from Thompson (1981), which show that the effect of adaptation changes from being attributed to low-pass adaptation to being attributed to band-pass adaptation only for adapting TFs of 16 Hz or greater.

An interesting additional conclusion can be drawn from comparing the results of adaptation to the 2 Hz gratings compared to 1 Hz and 4 Hz gratings. Indeed, after

115

adaptation to 2 Hz gratings, test gratings at 2 Hz are significantly slowed down, while 4 Hz and 8 Hz gratings are not slowed down. This effect is not present either for the lower adpating TFs or for the faster adapting TF of 4 Hz. This suggests that a separate channel with sensitivity around 2 Hz has been adapted by the 2 Hz grating, significantly more than a channel with lower preferred TF (as evidenced by the 0 Hz and 1 Hz data) and than another channel with higher preferred TF (as evidenced by the 4 Hz data). This is thus evidence for the presence of at least three channels—one lowpass, affected by 0 Hz grating adaptation, one sensitive to the 4 Hz gratings, and an intermediate one, with peak relative sensitivity nearer 2 Hz than either 1 or 4 Hz.

In summary, however, the data clearly indicate that adaptation to stationary gratings affected speed perception, and the difference between the pattern of data for stationary and 4 Hz grating argues that the low-pass channel is the mechanism responsible for that effect.

7.6 Spatial Frequency Control: Experiment 7.3

In experiment 7.3, the spatial frequency of the gratings was kept constant at 5 cpd. One possible reason why the apparent speed of gratings presented after adapting to stationary gratings was greater than without adaptation could be that the adaptation resulted not in a change in the measurement of temporal frequency, but instead in a change of the measurement of spatial frequency. If one assumes that the measurement of speed is the ratio of temporal frequency and spatial frequency, then an increase in apparent speed could be due to a decrease in apparent spatial frequency, with no change in apparent temporal frequency.

To test this hypothesis, the effect of adaptation to stationary gratings on the perception of spatial frequency was measured, using a very similar paradigm, in Experiment 7.3. The adapting stimulus was the stationary vertical grating, and the test

116

Subject	Condition	Test TF (Hz)	Matched Frequency (cpd)	Standard Error
DA	Pre	2	5.145	0.034
		8	4.755	0.134
	Post	2	5.390	0.042
		8	5.250	0.045
LW	Pre	2	4.950	0.152
		8	4.760	0.173
	Post	2	5.064	0.124
		8	5.424	0.127

Table 7.2: Spatial Frequency Control: Matched spatial frequencies before and after adaptation to 0 Hz gratings.

stimuli were moving vertical gratings as in the previous experiment. The subject's task was to match spatial frequency, and the staircase procedure was adjusted to modify the spatial frequency of the stimulus in the adapted field, as opposed to the temporal frequency adjustments made in Experiment 7.2. These matches were made both for a 120 sec adaptation, and without the adaptation, to allow again for the possible existence of biases in the matching procedure.

7.6.1 Results

The matched spatial frequencies and standard errors for those matches are presented in Table 7.2.

7.6.2 Discussion

As can be seen from the data in Table 7.2, the effect of adaptation on perceived spatial frequency was to increase perceived SF a little in the case of 2 Hz gratings, and more so in the case of 8 Hz gratings. Thus any model of speed perception which relied on perceived spatial frequency as a step in the computation of speed would have predicted that the perceived speed would have *decreased* after adaptation

to stationary gratings. As our results show an *increase*, such models are made improbable. The effect on 2 Hz gratings does not in fact appear to be robust. given LW's data. The effect on 8 Hz gratings is more interesting, but would require further data before too much should be made of it. Regardless, it appears irrelevant to the topic of discussion here.

7.7 Perceived Contrast Control

7.7.1 Experiment 7.4

Another potential confound in the Experiment 7.2 is that the result of the adaptation could be due not to a shift in the computed temporal frequency, but instead to a change in the perceived contrast, which in turn affects the perceived speed (Thompson, 1982; Stone & Thompson, 1992). It is unlikely that contrast adaptation is responsible for our results with low temporal frequency adaptation since the effect of adaptation is to *decrease* apparent contrast, which in turn has been linked with *decreases* in apparent speed (Thompson, 1981; Smith & Edgar, 1994), not *increases* as reported in Experiment 7.2, except at very high temporal frequencies in some experiments (but see Stone & Thompson, 1992 for a failure to replicate Thompson, 1982). Nevertheless, to rule out the influence of apparent contrast on our results in Experiment 7.2, we performed a pair of control experiments. In Experiment 7.4, the apparent contrast post-adaptation was measured in a contrast-matching task. In Experiment 7.5, the effect of contrast on speed was measured. In Experiment 7.4, observers adapted to vertical stationary gratings, and then matched the contrast of moving test gratings.

118

Subject	Condition	Test TF (Hz)	Matched Contrast (%)	Standard Error
DA	Pre	2	25.475	0.209
		8	25.540	0.346
	Post	2	20.815	0.396
		8	22.030	0.371
LW	Pre	2	23.855	0.240
		8	25.175	0.332
	Post	2	19.450	0.230
		8	19.750	0.196

Table 7.3: Contrast Control: Matched Contrast before and after adaptation to 0 Hz gratings.

Methods

The matching procedure used the same staircase as described for Experiment 7.1, except that the stimulus parameter which the user controlled was the grating contrast instead of its temporal frequency.

Results

The results for both observer's contrast matches both with and without adaptation to a 0 Hz grating for 120 sec are presented in Table 7.3.

Discussion

The results for both subjects (see Table 7.3) indicate that after adaptation, perceived contrast dropped by an amount equivalent to a change of 20% or so in physical contrast. The next experiment tested what consequences such a physical change in contrast would have on perceived speed.

Subject	Test TF (Hz)	Relative Perceived speed
DA	2	0.975
	8	1.17
LW	2	0.955
	8	0.64

Table 7.4: Contrast Control: Relative perceived speed between 25% contrast gratings and lower contrast gratings (16% gratings for subject LW and 18% for subject DA) after adaptation to 0 Hz gratings. Values lower than 1 indicate that the low-contrast grating appeared to move slower than the higher contrast grating, and values greater than 1 indicate that the low-contrast grating appeared to move faster than the higher contrast grating.

7.7.2 Experiment 7.5

In Experiment 7.5, a standard matching paradigm without adaptation was used to measure the effect of contrast on perceived speed using the contrast values obtained in Experiment 7.4. The values chosen for the contrast of the match grating was the lowest contrast match each observer gave. These were of 18% for subject DA and 16% for subject LW.

Results

The relative perceived speed (normalized by the pretest values obtained in Experiment 7.1) for both observers are displayed in Table 7.4.

Discussion

The data in Table 7.4, indicate that for subject DA, the reduction in contrast led to an apparent increase in speed of fast gratings and no change in the apparenty speed of the slow gratings. This is the opposite effect from that which this control was designed to test for (slow gratings moved faster and fast gratings did not). Subject LW showed a different effect, so that fast gratings appeared to move slower, while there was no effect for slow gratings. Again, these results do not support

the hypothesis that the effects found in Experiment 7.2 are due to an effect on the perceived contrast and a subsequent effect on perceived speed.

7.8 Afterimage control: Experiment 7.6

One final control experiment is needed. After adaptation to stationary gratings, one could imagine that an afterimage is formed. While no subject reported seeing such an afterimage[3], probably due to the presence of small eye movements and the high spatial frequency of the adapting gratings, it is possible that such an afterimage could have been present but at a low enough levels that it was not spontaneously reported. The effect of such an afterimage on speed perception is not fully specified in any of the models discussed above, but one can imagine that the superposition of the afterimage with the test gratings could yield beats in the image, which could in turn be used to match speed. How these beats could yield the patterns of results described for Experiment 7.2 is unclear. To test against this possibility, Experiment 7.6 was performed, in which an "artificial afterimage" was simulated. Given that the purpose of this experiment was to find out if such an afterimage due to adaptation could bias speed perception, no adaptation was performed. Instead, a simple speed matching task was performed, where the speed of a single grating was matched to a display consisting of a moving grating and a stationary grating, superimposed.

7.8.1 Methods

The same matching procedure as in Experiment 7.1, except that one of the fields had in addition to the moving gratings, a fixed grating of the same spatial frequency

---

[3]The afterimage referred to here is that created by vertical adaptation gratings, not that created by the oscilloscope display itself, which was quite bright relative to the rest of the room, and created a very noticeable contrast and color afterimage.

121

Subject	Test TF (Hz)	Matched Speed	Standard Error
DA	2	1.12	0.02
	8	1.18	0.01
LW	2	1.16	0.01
	8	1.00	0.03

Table 7.5: Afterimage: Relative perceived speeds with 5% afterimage superimposed.

as the test gratings, with a fixed contrast of 5%. This contrast results in much higher perceived contrast than the eventual afterimage, since it was quite visible, while the afterimage was never seen by observers. Thus this experiment is likely to overestimate the effect of afterimages if any is found.

7.8.2 Results

The relative perceived speed (normalized by the pretest values obtained in Experiment 7.1) of gratings matched to the gratings superimposed with the artificial afterimage are presented in Table 7.5.

7.8.3 Discussion

The data in Table 7.5 show the first effect which may provide an alternative explanation to the effects of Experiment 7.2. Both observers show that the addition of a stationary afterimage sped up the perceived speed of slow gratings. This effect was also found for one observer for fast gratings, but not for the other observer. The fact that the effect of test TF on this apparent speedup does not match the trends observed in Experiment 7.2 seems to indicate that while an interesting effect, it does not neatly account for the results in that experiment. Furthermore, the process by which addition of a stationary grating results in apparent speeding up of a moving gratings is far from obvious. One possibility is that a motion repulsion effect is responsible. According to such an account, the presence of the very low grating results

122

in segregation of the two gratings by the motion analysis system, which results in enhanced separation between the two estimates of speed, thus in an apparent speed up.

To test the motion repulsion effect, a final control experiment was run on subject LW, using 1 Hz gratings instead of 2 Hz gratings. Thus the perceived speed of both 2 Hz and 1 Hz 10% gratings was measured in the presence of a 5% afterimage. The relative perceived speed of the 2 Hz grating was 1.10, replicating the speed up observed in the last experiment, but the relative perceived speed of the 1 Hz grating was 0.96, thus showing that the afterimage effect disappeared with 1 Hz gratings. As a final control, the basic adaptation effect was run again on the same subject for 1 Hz gratings, and yielded an apparent speed up of 15%, showing that while the basic effect of adaptation to stationary gratings applies to 1 Hz gratings, the afterimage effect does not.

In conclusion, all of the control experiments above either showed no effect, showed effects which would tend to go counter the main effect under study, or showed effects which depended on stimulus parameters in ways that the main effect did not.

7.9   General Discussion

The study of motion perception is generally viewed as the study of the perception of moving stimuli. In contrast, spatial vision is generally viewed as the dealing with aspects of vision best studied with stationary stimuli. The results presented here indicate that these two domains are neighbors rather than opposites. Our results indicate that exposure to stationary displays can affect the perception of moving displays; more precisely, exposure to stationary displays can affect the perceived speed of moving displays. These results should be viewed as fitting quite coherently with the rest of the data on adaptation and speed perception. For example, the

123

results of Thompson (1976, 1982), Schrater and Simoncelli (1995) are all compatible with these data. Even the Smith and Edgar (1994) data are consistent with the results on stationary adaptation—the bandpass channel could very well be more adaptable than the lowpass channel. The fundamental novel result is that the lowpass channel *can* be adapted, provided the stimulus configuration is chosen carefully.

Chapter 8

Spatial Frequency Integration

8.1 Introduction

As should be clear by now, observers can readily estimate the retinal speed of moving visual stimuli. In any of the models described in Chapter 3, a translating visual display will results in activation of a set of spatial and temporal frequency tuned mechanisms. The mechanisms which will be most strongly activated will be those with spatial frequency tunings which correspond to the spatial frequencies present within the spatial Fourier spectrum of the display, and with temporal frequency ranges which correspond to the rate at which the spatial structure of the display passes by a given retinal locus. Thus a given stimulus is processed in parallel by a range of mechanisms, each of which analyze a restricted region of the stimulus' spatiotemporal frequency profile. Outside of the laboratory, however, all visual objects have rich spatial frequency structure, and therefore activate broad distributions of such mechanisms. Furthermore, a translating object moving at a given speed will cause activity across several mechanisms with SF and TF frequency tunings corresponding to that speed. Given this bias in the distribution of activity across these mechanisms (along lines of same speed), it would seem efficient for the visual system,

125
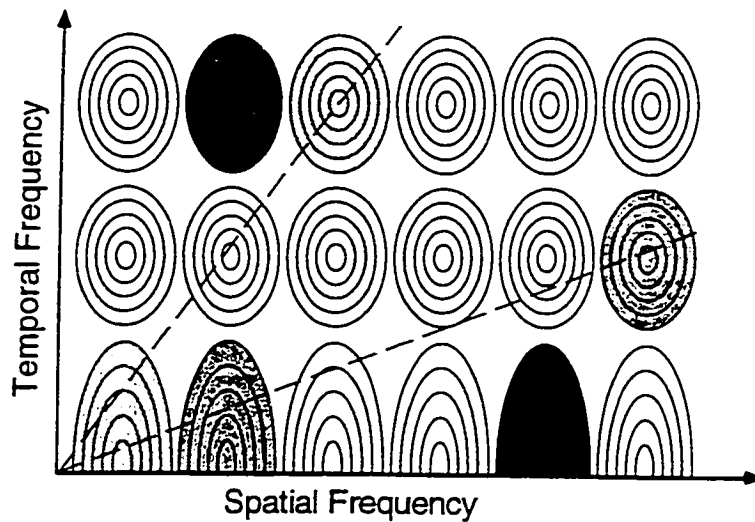
Figure 8.1: Schematic representation of the peaks of the filters relatively most sensitive as a function of the spatial frequency (horizontal axis) and temporal frequency (vertical axis). Only the peaks of the tuning functions are represented by the concentric ovals, and the filter sensitivity is non-null over a large range of the SF/TF map (especially along the TF dimension). Three different stimuli are represented by shaded pairs of mechanisms. The angle between the best fitting line and the abscissa corresponds to the speed of the stimuli. The dark gray mechanisms will not be activated by the same stimulus, since they correspond to very different speeds.

when computing speed, to pool signals which correspond to a given speed, but not signals which correspond to different speeds (see Figure 8.1). We refer to such a pooling process as integration. This chapter reports the results of psychophysical experiments testing whether when computing speed, the human visual system integrates information corresponding to a given speed but distributed across spatial frequencies, temporal frequencies and orientations.

As mentioned in earlier chapters, the first stage of processing of visual information is assumed to be filtering by mechanisms tuned in spatial and temporal frequencies. These mechanisms are modeled as linear integrators of information over their spatial and temporal frequency range. Given that by definition these mechanisms

126

integrate information within their frequency tuning ranges, experiments testing for speed-specific integration must take into account the possibility that any observed integration is due to within-mechanism integration. This can be done by using component stimuli which are known to be well outside the tuning regions of the elementary mechanisms. Estimates for the spatial-frequency bandwidth of the mechanisms (defined as the full width at half-height of the sensitivity function) depend on the estimation method used, but typically are found to be between 1 and 2.5 octaves (for a review, see Olzak & Thomas, 1986). Thus in order to find evidence of speed-specific integration across mechanisms, gratings of very different spatial frequencies must be used. In the experiments reported below, one grating had a spatial frequency of 0.5 cpd and the other had a spatial frequency of either 3 or 6 cpd, corresponding to SF separations of approximately 2.5 and 3.5 octaves, which places them in well separated spatial frequency channels (especially for the 0.5/6 cpd combinations).

The predominant model of analysis of spatial patterns is one of narrow-band mechanisms which can be thought of as separate neural pathways. Indeed, there is strong evidence for the notion of separate pathways for each spatial frequency band, both psychophysically (Campbell & Robson, 1968) and physiologically (De-Valois et al., 1986). If one takes this model as a working hypothesis, then to look for evidence for integration across spatial frequencies, one can measure speed discrimination thresholds for a single low spatial frequency grating, for a high spatial frequency grating, and then for a compound grating with both frequency components—the relative performance in the last condition relative to the other two could then be informative regarding both whether such an integration occurs, and if so, what type of integration it consists of.

127

8.2 Methods

In all the experiments described in this chapter, the experimental setup was the same. The stimuli were sinusoidal gratings generated by the Picasso function generator. The stimuli consisted either of a single low spatial frequency grating at 0.5 cpd, a single high spatial frequency grating at either 3 or 6 cpd, or both gratings superimposed. The speed of the gratings was controlled by setting the phase of each grating at each frame. The average duration of the stimuli was set at 120 ms, but any given stimulus had a duration which was chosen randomly at each trial from a distribution of durations ranging from 100 ms to 140 ms, so as to reduce the usefulness of the distance traveled by the gratings as a cue to speed. Furthermore, since apparent speed can be influenced by contrast, the contrast of the low spatial frequency grating was chosen at each trial from a distribution ranging, depending on the observers, from 2.5% to 5%, 3% to 6% or 4% to 8% contrast. The contrast of the high spatial frequency was a variable manipulated in each experiment as will be described below, but it was constant throughout a block, so that the contrast of the high spatial frequency grating was also chosen from an octave-wide range within a given block. The data from six observers are reported below. Subjects DA, EKF and LW were aware of the aims of the experiments, and subjects MI, DV and AD were students naive as to the aims of the experiment. LW, DV, MI and AD are emmetropes, and EKF and DA are myopes but wore their normal optical corrections throughout the experiments. All observers had extensive practice in making speed discrimination judgments under similar conditions prior to data collection.

Speed discrimination thresholds were obtained using the Probit method described in earlier chapters, using a mean speed of 2°/sec for most conditions and 1°/sec in some conditions. Each data point in all discrimination experiments correspond to the means of thresholds estimated with between 150 and 600 trials.

128

8.3   Experiment 8.1: Integration across spatial frequencies

Experiment 8.1 examined whether observers were able to use information from both spatial frequency ranges when making speed discrimination judgments. In this experiment, both gratings were vertically oriented and moved horizontally—the direction of motion (leftward or rightward) was chosen randomly on each trial.

Any given block was randomly chosen from one of three conditions, counterbalanced across blocks. In one condition (Low SF), only the low spatial frequency grating was displayed, and a speed discrimination threshold was obtained for that grating alone. In the second condition (High SF), only the high spatial frequency grating was displayed, and a second speed discrimination threshold was obtained, for it alone. In the third condition (Compound), both gratings were displayed and a speed discrimination threshold for the compound gratings was obtained. In this experiment, when both gratings were displayed they both moved at the same speed and in the same direction, and the ratio between the contrast of the high spatial frequency grating and that of the low spatial frequency grating was kept constant (for some observers both gratings always had equal contrast, and for the remaining observers the high SF grating was always at twice the contrast of the low SF grating).

Speed discrimination thresholds were averaged within a condition for each observer. The standard error for the estimate of the threshold was computed from the standard errors provided by the Probit fit algorithm. The data for Experiment 8.1 are plotted in Figure 8.2. In this and all subsequent plots, the thresholds indicated are the thresholds in speed (°/sec). Since the reference speed was 2°/sec, Weber fractions are half of the plotted results. In some cases, the High SF condition was too difficult, and observers were unable to make speed discrimination judgments, or

129

the Probit procedure was unable to estimate their thresholds. Those conditions are indicated with the bars which reach ceiling and have no error bars. indicating that the threshold was too high to be measured using the technique employed here.

The basic result from these data is that under these low contrast. short duration conditions, 1) speed discrimination for 0.5 cpd gratings is quite poor (corresponding to Weber fractions of up to 50% depending on the observers). 2) speed discrimination for the 3 and 6 cpd gratings is even worse (often not measurable). and 3) speed discrimination for the compound gratings is significantly better than either of the single grating conditions, yielding lower thresholds.

Given the assumption described above of independent processing by mechanisms with relatively narrow spatial frequency tuning, these data suggest that there is indeed pooling of information across spatiotemporal filtering mechanisms when doing speed perception.

8.4 Experiment 8.2: Integration across orientation differences

Integration across spatial frequency ranges can be related to the fact that real objects have a broad spatial frequency spectrum. and that speed information about an object is available throughout their spectrum. If one considers objects which are not just translating in the frontoparallel plane but rotating in the frontoparallel plane or looming towards the observer, it should be clear that speed information about the object's motion is available at many orientations, and under some circumstances (like the rotation in the picture plane and the looming described above), information about the speed of an object (or, equivalently, self-motion) can span many orientations. Experiment 8.2 examined whether the apparent ability of the visual system to pool
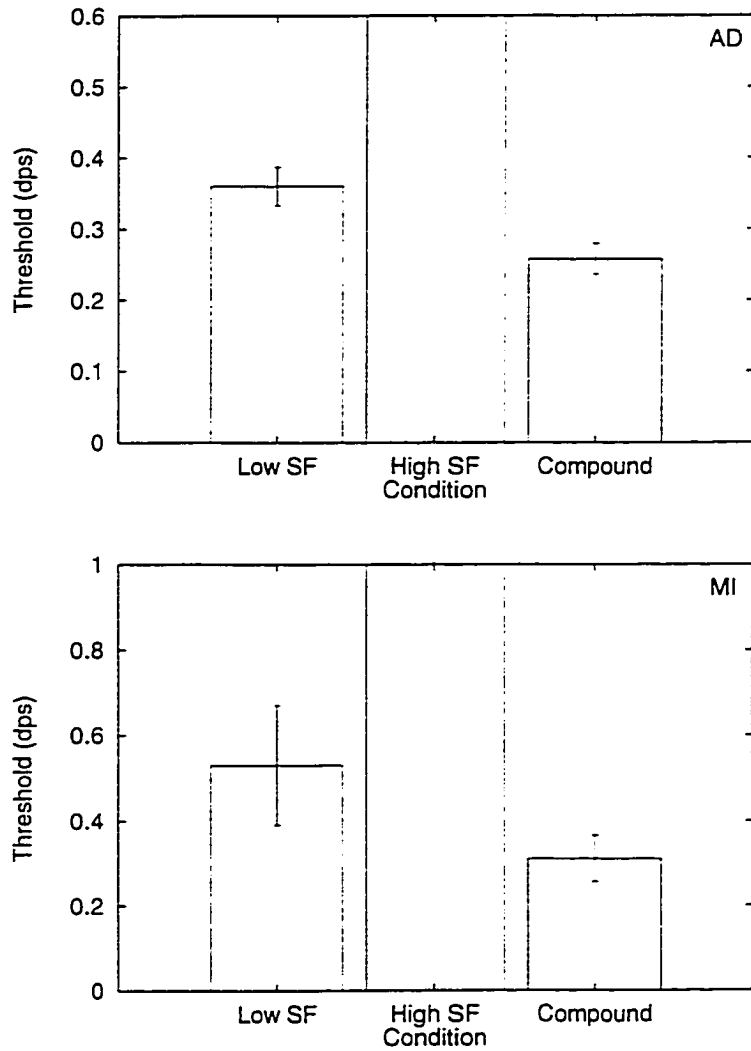
Figure 8.2: Results from Experiment 8.1. Discrimination thresholds are displayed for each of three contrast conditions. (Continued in next figure)
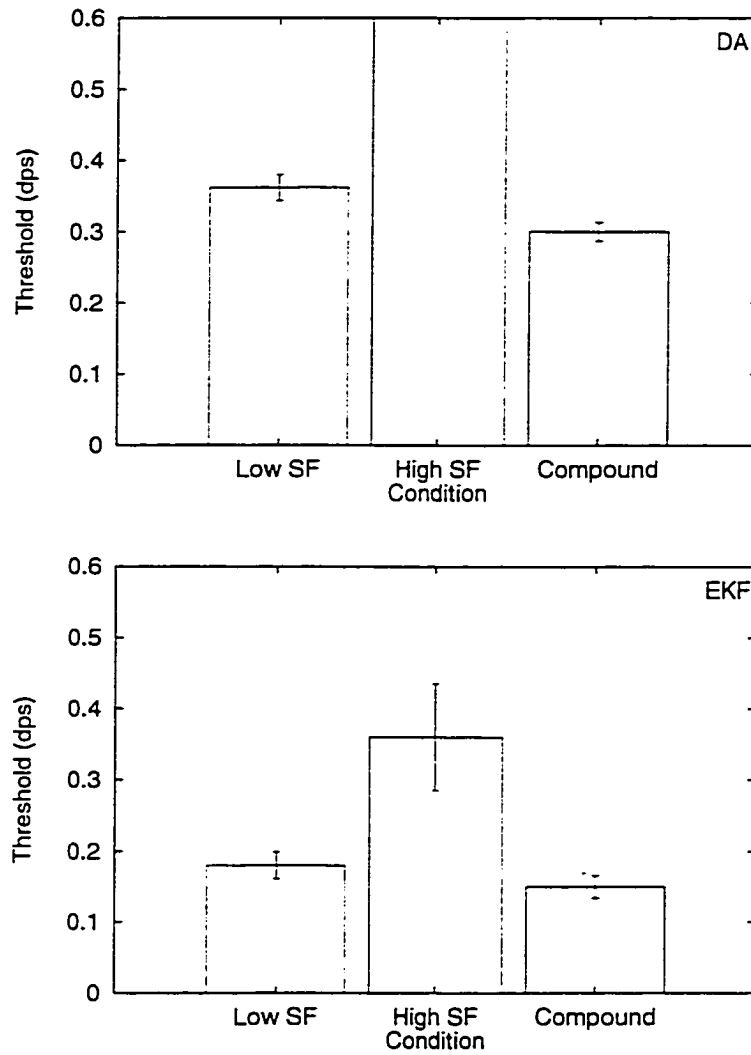
Figure 8.3: Results from Experiment 8.1. (Continued from previous figure)

signals over wide ranges of spatial frequencies could be extended to deal with differing orientations. The experimental conditions were the same as in Experiment 8.1. except that in the Compound condition, the orientation of the high spatial frequency grating was chosen to be one of 0°, 90° or 180°. When the angle was set to be 0°. the conditions were identical to those used in Experiment 8.1. When the angle was set to be 90°, the two gratings were in a plaid configuration (it should be noted that because of the large difference in spatial frequencies, these gratings did not cohere into a rigid plaid, and were instead perceived as sliding gratings). When the angle was 180°, the two gratings had identical orientation (vertical). but moved in opposite directions. Their speeds were always identical in a given trial. and varied across trials as in Experiment 8.1.

Means and standard deviations for the speed discrimination expressed in Weber fractions were computed as in Experiment 8.1 and are plotted in Figure 8.4.

This experiment yielded somewhat bizarre results. Subject LW showed basically no effect of the orientation of the high spatial frequency grating, with improved performance (over the single grating condition) even for gratings going in the opposite direction. Subject S1 showed the same lack of dependence on the angle of the high SF component. Subjects DV and VA, on the other hand, could not perform the speed task when the orthogonal high spatial frequency grating was added (data not shown). Thus for some observers, the angle of the high SF grating is irrelevant and its presence is all that is required for the improvement in performance, while for others, the presence of a 90° grating is highly disruptive and results in worse performance. Discussion of these results is deferred until later.
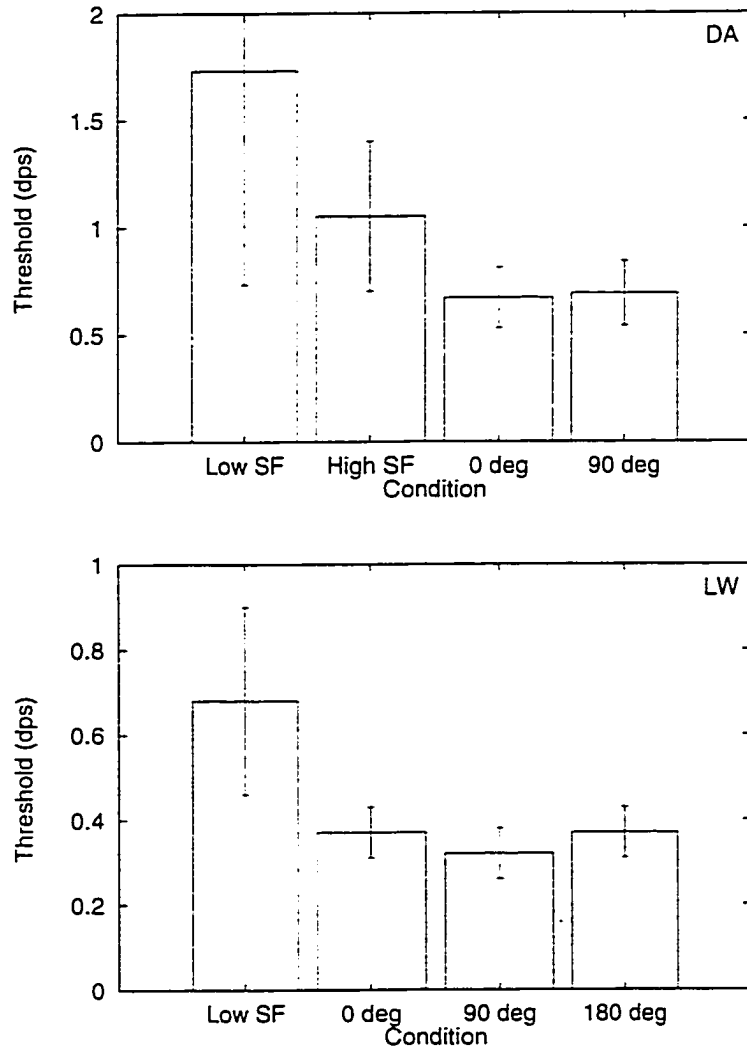
Figure 8.4: Results of Experiment 8.2. Discrimination thresholds are displayed for the low SF grating alone, the high SF grating alone (not depicted for subject LW, as it was not measureable, and for various compound gratings, with angles between the two gratings either 0, 90 or 180°.

134

8.5 Experiment 8.3: Correlational variations

The results from Experiment 8.1 are easily accomodated by the plane-fitting model of Grzywacz and Yuille (1990), as described in Chapter 3. Indeed. the stimulus parameters were chosen to yield informational structure in the displays which should enhance the system's ability to perform the plane-fitting mechanism described by Grzywacz and Yuille. One natural assumption to make about the plane-fitting process is that the best distribution of activities (that likely to yield the most accurate speed estimate) should in fact correspond to a plane which passes through the origin in the spatiotemporal frequency plane. Experiment 8.3 tested this assumption of the model, by varying the correlational structure of the stimuli across spatial frequency bands. Specifically, the relationship between the information available in the high SF band and that available in the low SF band was varied across conditions. Two new conditions were used. the 'anticorrelated' and 'uncorrelated' conditions. In the 'anticorrelated' condition, the speed of the high SF grating was chosen so that the correlation between the high SF speed and the low SF speed was -1: on a trial in which the speed of the low SF grating was greater than the mean speed, that of the high SF grating was lower than the mean speed, and vice versa (see Figure 8.5 below). In the 'uncorrelated' condition, the speed of the high SF grating was chosen at random (but from the same distribution as in Experiment 8.1), so that it could not provide any useful information regarding the correct answer on a given trial. More precisely, on half of the trials it provided useful information, and on the other half of the trials it provided misleading information. In this experiment, feedback was always determined by the actual speed of the low SF grating, which was the most salient part of the stimulus. and which observers reported "using" to do the task.

Means and standard deviations for the speed discrimination expressed in Weber fractions were computed as in Experiment 8.1, and are plotted in Figure 8.6, along
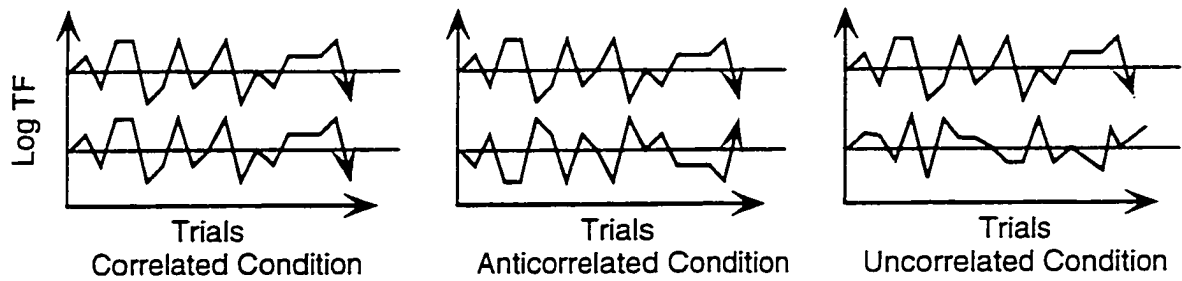
Figure 8.5: Schematic of the speeds of the two gratings in a given block for each condition. In the correlated condition both speeds covary with a correlation of 1—they both track the same path. In the anticorrelated case, they have a correlation of −1, and in the uncorrelated case, the two paths are independent.

with the means obtained for the low SF grating alone (to provide a reference point showing no integration), and the Compound condition of Experiment 8.1 (to provide the reference point corresponding to theoretically maximal integration).

The data from Experiment 8.3 appears to argue that observers can sometimes use either the anticorrelated signals (Subject VA) or the uncorrelated signals (Subject DV) as well as the correlated signals.

These results appear to challenge the assumption described above that the best-fitting plane for speed estimation must pass through the origin of the spatiotemporal frequency space, and instead argue for a model in which speed estimates are performed within a spatial frequency band, and then integrated across spatial frequency bands in a task-specific manner—indeed, the data show (for most observers) very little if any decrement in performance when the correlation was negative (which corresponds to highly unnatural correlational structure) relative to the more ecological positive correlation among signals.

136

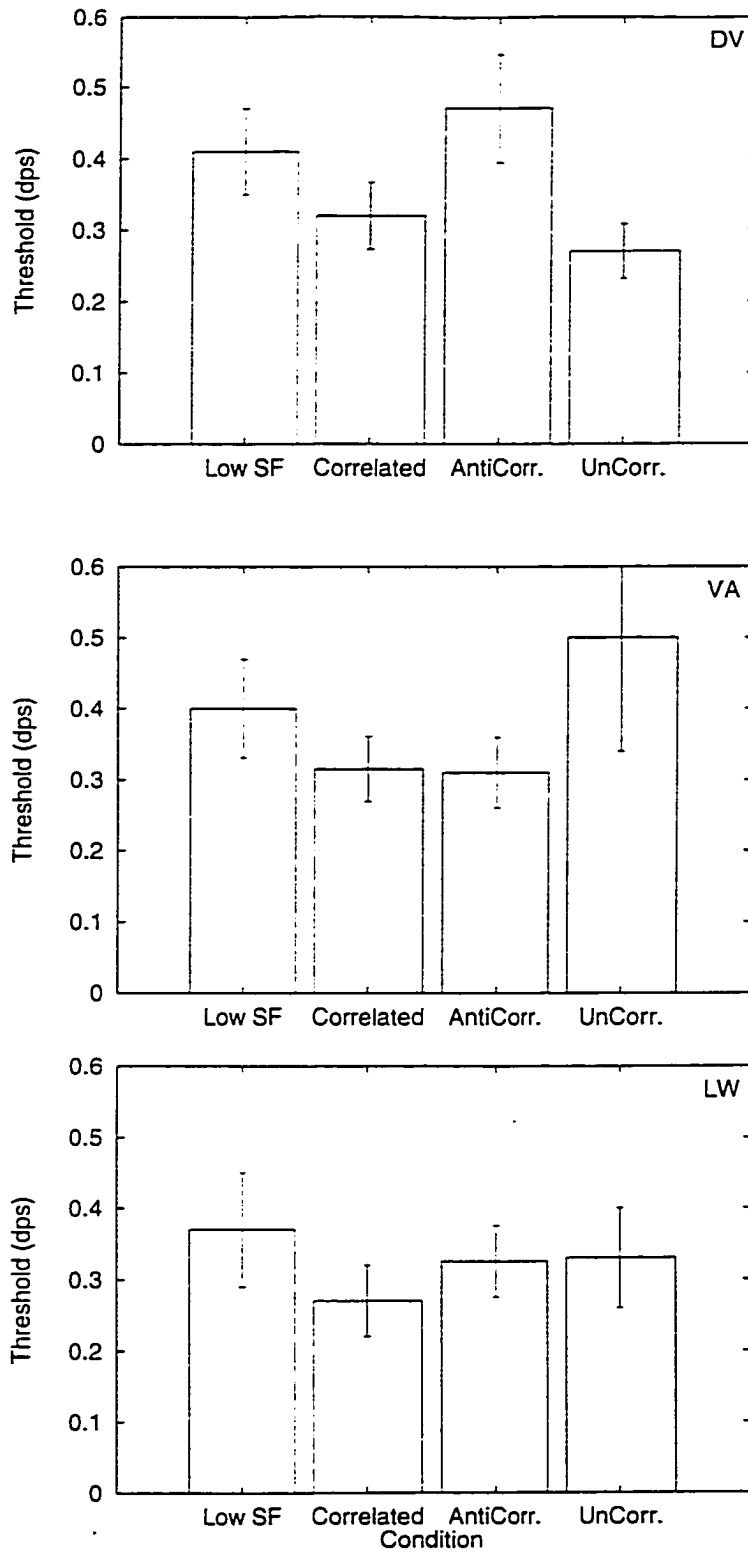Figure 8.6: Results from Experiment 8.3. Discrimination thresholds for various stimulus correlations are depicted. All gratings have a speed of 2°/sec.

137

8.6 Experiment 8.4: High Contrast inhibition

All of the experiments presented thus far argue for a highly efficient speed computation system which is somehow able to integrate information across large spatial frequency ranges, orientation ranges of up to 180° even pick up sometimes on negative correlations among signals in separate spatial frequency bands. Experiment 8.4 examined whether the mechanism under study is equally able to integrate information across contrast ranges. It examined the effect of varying the contrast of either grating on the ability of observers to perform speed discrimination in the Compound condition described in Experiment 1. Experimental conditions are the same as those used for Experiment 1, except that the contrast of the low or high spatial frequency grating was varied across blocks. When the low spatial frequency grating's contrast was varied, the contrast of the high spatial frequency grating was randomly chosen from the range of 2.5%–5% or 5%–10% as in Experiment 8.1, and when the high spatial frequency grating's contrast was varied, the contrast of the low spatial frequency grating was chosen from the range of 2.5%–5% contrast, again as in Experiment 8.1. Because the gratings were not far above detection threshold, only conditions corresponding to increases in contrast were used (making the gratings invisible in some trials would change the task to a detection task, and thus make the results hard to analyze). In the first set of conditions, the contrast of the low SF grating was multiplied by a factor of 2, 4 or 8 across blocks. In the second condition, the contrast of the high SF grating was multiplied by a factor of 0.5, 1, 2, 4, 8, 16 and 32 across blocks.

Results are plotted in Figures 8.7 and 8.8 for 3 observers. Let us consider first the results when the contrast of the *low SF* grating is increased (Figure 8.7). As can be readily seen, performance increases (thresholds drop) and appears to plateau with very high contrasts. This is hardly surprising, as all models would predict that
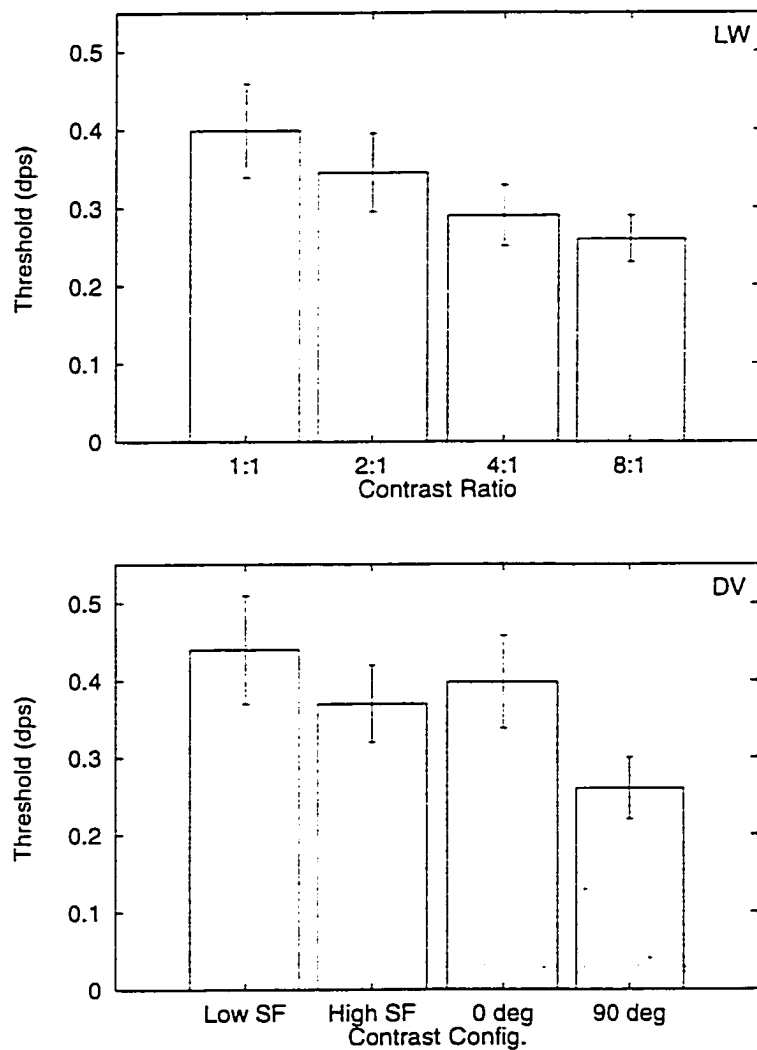
Figure 8.7: Speed discrimination thresholds for compound gratings as the contrast of the low SF grating is varied.

139
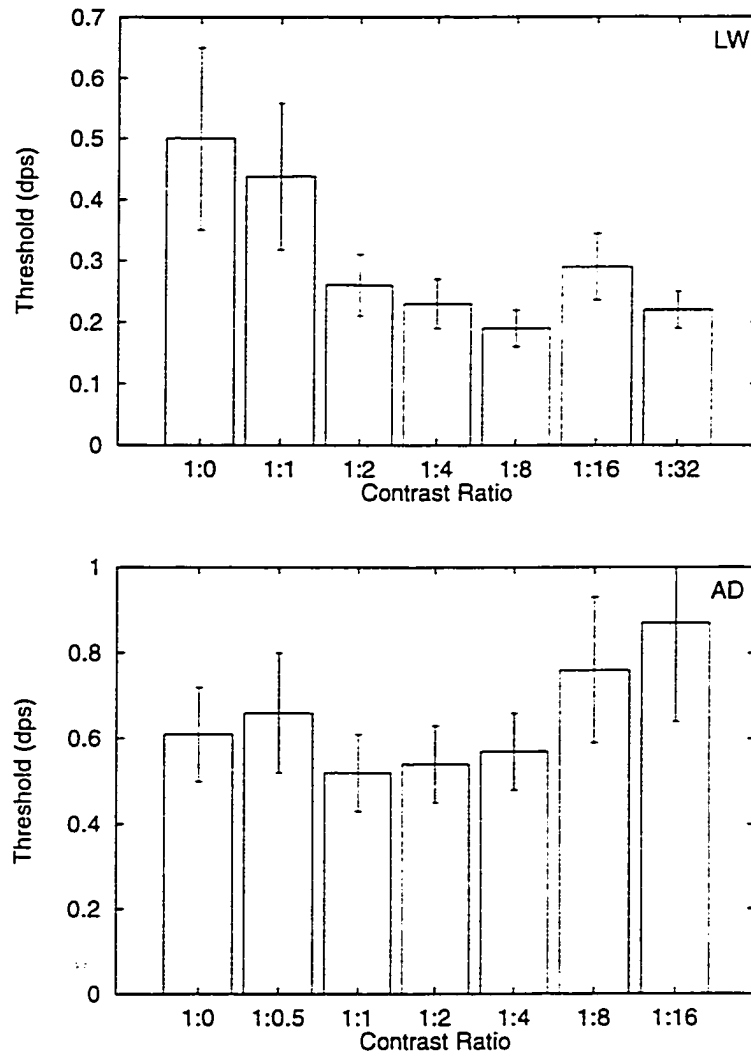
Figure 8.8: Speed discrimination thresholds for compound gratings as the contrast of the high SF grating is varied (continued in next figure).

140
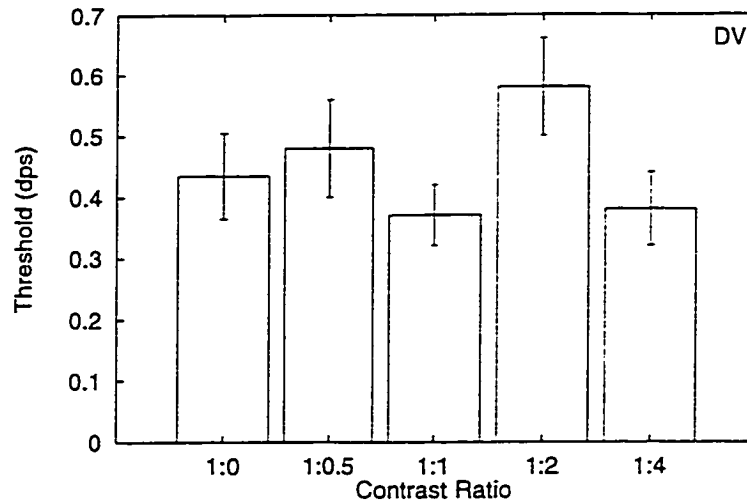
Figure 8.9: Speed discrimination thresholds for compound gratings as the contrast of the high SF grating is varied (continuation from previous figure).

the information available in the display should be monotonically related to contrast energy. Consider now the results when the contrast of the *high SF* grating is increased (Figure 8.8). For all observers, while there is a improvement in performance for low contrast high SF gratings (as discussed in Experiment 8.1), increasing the contrast of the high SF grating beyond a certain point results in a decrement in performance (For subject DV this decrement appears not to hold for even larger contrasts). This decrement in performance for high contrast signals is paradoxical—it is important to remember that for all the conditions depicted in Figure 8.7 and 8.8 the low spatial frequency grating is always available, at a fixed contrast range. Thus any logical integration rule should at worst do as poorly in the Compound condition as in the Low SF alone condition. It appears that the presence of a high contrast, high spatial frequency stimulus can mask, or inhibit, the signal due to the low contrast, low spatial frequency grating, within the context of a speed discrimination task.

141

8.7 Experiment 8.5: Masking revealed by detection thresholds

One obvious question which arises given the results in Experiment 8.4 is whether the masking effect of high SF gratings on low SF grating is measurable in a detection task. As described earlier, the standard model of spatial frequency analysis holds that stimuli with such widely separated frequency content should be processed by independent mechanisms, but adaptation studies have suggested that rather than being independent, spatial frequency channels in fact interact via a tonic mutual inhibition (DeValois, 1977). To establish whether this inhibitory effect is the source of the loss in performance for high contrast high SF gratings, a detection experiment was run. Detection threshold for the low SF grating was measured with a variety of contrasts for the masking high SF grating using a one interval, two alternative forced choice paradigm. The stimulus duration and temporal frequencies were the mean values of those used in Experiment 8.4. In this experiment, observers were presented with a set of trials in each repetition of each condition. In each trial, a 120 msec stimulus was presented after a 500 msec blank period. A trial could contain either both the low spatial frequency grating and the high spatial frequency grating (the test trials), or just the high spatial frequency grating (the catch trials). A given trial had a 50% chance of being a test trial, and a block terminated when 35 test trials had occurred. The responses to the test trials were used to control a QUEST adaptive threshold estimation procedure (Watson & Pelli, 1983) so that the contrast of the low SF grating in the test trials was that at which observers were expected to be 92% correct. Observers indicated whether they had detected the low SF grating in a given trial by using one of the two mouse buttons. Incorrect responses (false alarms, i.e. "Yes" responses on catch trials, and misses, i.e. "No" responses on test

trials) were indicated using computer-generated beeps. This procedure yielded the contrast of the low SF grating required for a detectability corresponding to a d' value of 2. Threshold detection contrasts were measured for each observer twice at each of 5 values of the high SF contrast (0, 5%, 10%, 20%, 40%). Results are plotted in Figure 8.10.

As can be seen from examination of Figure 8.10, detection thresholds for the low SF grating was influenced by the presence of the masking high SF grating, with threshold rising from 1.6%-1.8% contrast when no mask is present to 2.1%-2.8% contrast in the presence of a 40% masking grating.

These data are consistent with the DeValois results. It is noteworthy that the separation in octaves between the gratings used in this experiment (3.5 octaves approximately) happens to be quite similar to the separation in frequencies at which DeValois found maximal effects of adaptation on detection ability. It should be noted however that our spatial frequencies were quite different from hers, since she adapted to gratings at 1.19 cpd, while the corresponding gratings in our experiments were those at 0.5 cpd.

8.8   Experiment 8.6: High SF Grating as Mask

Can a single interaction account both for the results of experiments 8.1 through 8.3 on one hand, which seem to argue for a cooperative, integrative interaction, and 8.4 and 8.5, which seem to argue for a competitive, inhibitory interaction? As mentioned earlier, DeValois (1977) has argued that there is a tonic inhibitory interaction among channels, and explains her data showing increases in SF discrimination 3 octaves away from the frequency at which an adapting grating was presented by suggesting that the adaptation of mechanisms tuned to the adapting grating dampens their inhibitory action on the other mechanisms. The high contrast results of
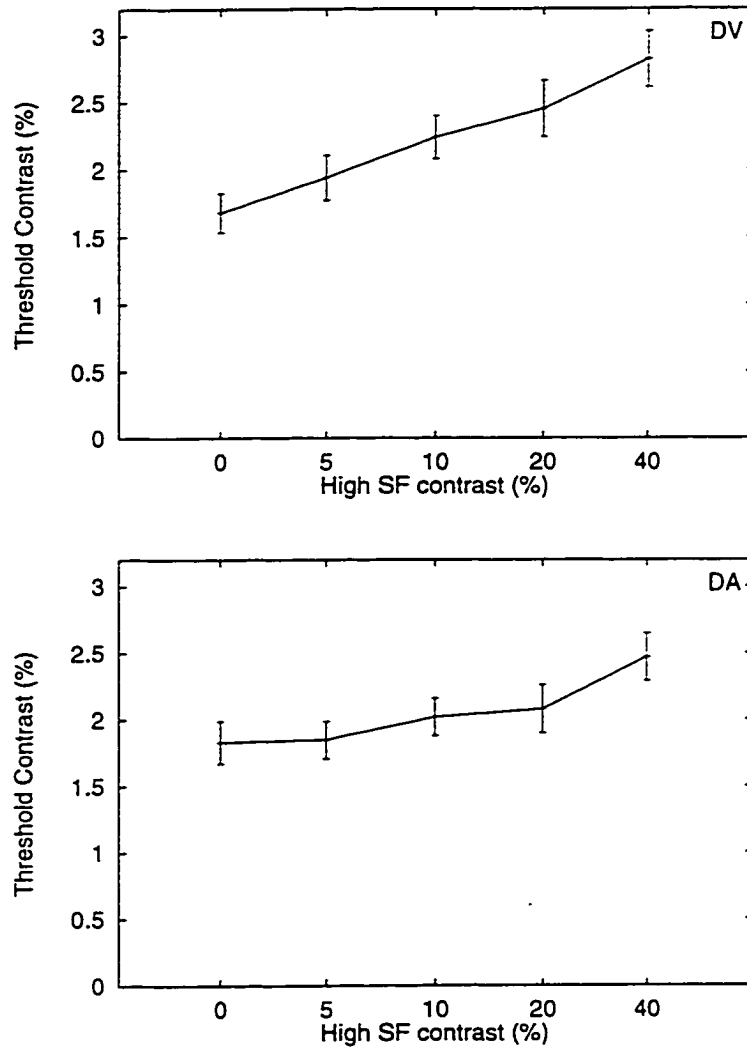
143

Figure 8.10: Detection thresholds for three observers for the low spatial frequency grating, as a function of the contrast of the masking, high spatial frequency grating. Thresholds estimated for $d' = 2$, using QUEST. (Continued in next figure)

144
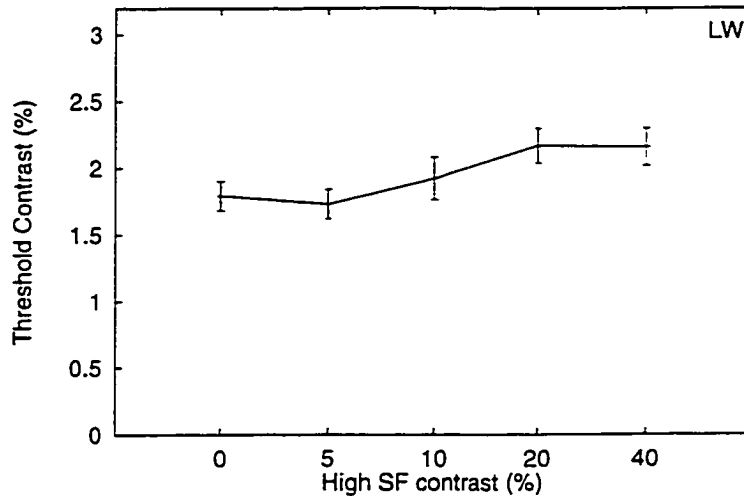
Figure 8.11: Detection Thresholds for three observers for the low spatial frequency grating, as a function of the contrast of the masking, high spatial frequency grating. Thresholds estimated for $d' = 2$, using QUEST. (Continuation of previous figure)

Experiment 8.4 can be viewed in a similar light—if one assumes that the inhibitory interaction suggested by DeValois is proportional to mechanism activity, then one can consider an explanation for the data not in terms of various kinds of integration across mechanisms, as originally conceived and described in the introduction, but instead as a nonlinear effect resulting from a single inhibitory mechanism. Were this to be the case, the information about speed which is present in the high SF band should be irrelevant. The "uncorrelated" condition in Experiment 8.3 seems to argue against this explanation—however, it should be noted that in that experiment, the high SF band provides variation across trials, which very well might be construed by the visual system as information, albeit information which is decidedly unhelpful on average in solving the task.

To test the hypothesis that the main effect of the high SF grating is that mediated by this cross-mechanism inhibition, a final experiment was run which was a slight modification of Experiment 8.3. A single new condition was added, in which the

145

speed of the high SF grating was held constant at the mean speed of 2°/sec through-out all trials. Thus it provided no information regarding the speed discrimination task, helpful or otherwise, but presumably resulted in the same amount of inhibition as the grating used in Experiment 8.1. Speed discrimination thresholds were obtained for the single low SF grating condition, the compound/correlated condition (both replications of Experiment 8.1), and the new condition, in which the speed of the high SF grating was always 2°/sec. Results are plotted for two observers in Figure 8.12.

The data indicate that the addition of the constant-speed high SF grating resulted in an improvement in performance which was in this case better than that yielded by adding "useful" speed signals in the high SF band[1]. Thus, integration of information is most likely not the source of the improvement seen in Experiments 1 through 3, and instead a non-specific interaction between SF channels is responsible, which yields increases in performance for low levels of contrast, and decreases in performance for high levels of contrast. It should be noted that in many of the experiments with low contrast high SF gratings, but in this experiment especially, observers, when debriefed, reported not knowing that there were differences between the various blocks they had just run. It is striking to see that such a subtle change in the perceptual quality of a stimulus can have measureable changes in performance, especially given the lack of "information" in the mask.

---

[1]One might wonder why the replication of Experiment 8.1 did not work—the correlated condition did not improve discrimination performance, as it did in that experiment. This could be due to the fact that the contrast of the high SF grating was fairly low (equal to the low SF grating, as opposed to twice as large in most of the conditions of Experiment 8.1).
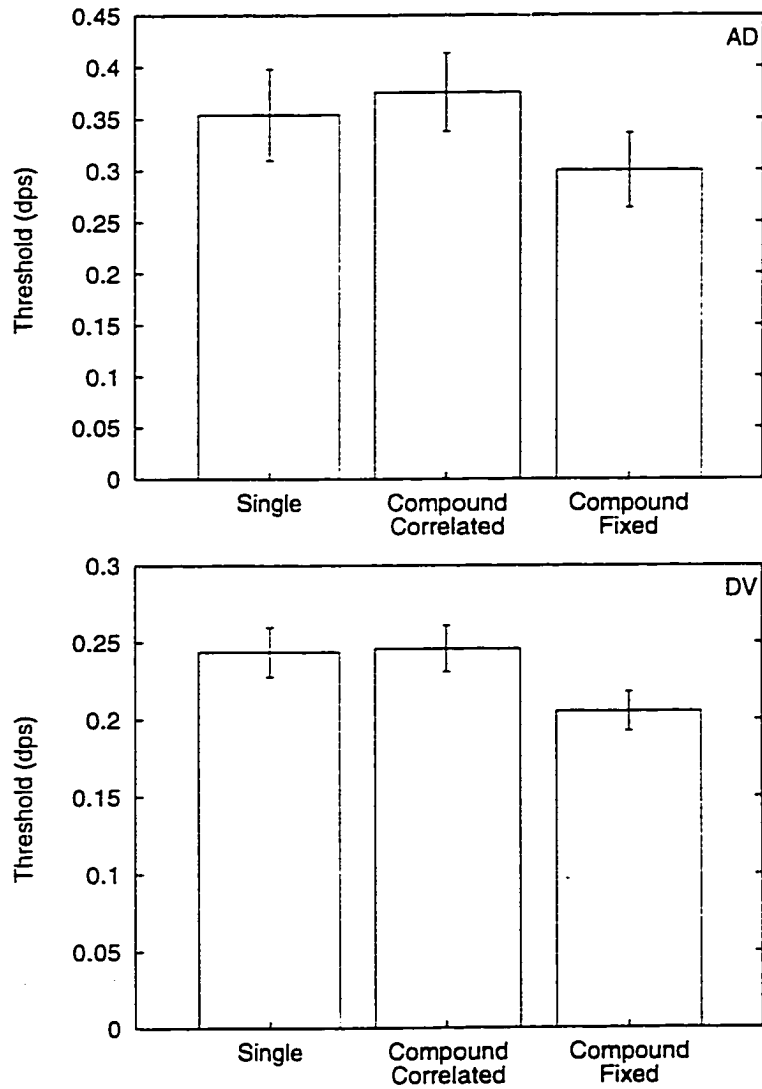
146

Figure 8.12: Speed discrimination threshold is plotted as a function of the spectral content of the stimuli.

147

8.9 General Discussion

The results described of all six experiments reported above can be broadly categorized into two general effects: increases in performance in speed discrimination due to the addition of a low-contrast stimulus component (as in Experiments 8.1, 8.2, 8.3 and 8.6), and decreases in performance due to the addition of a high-contrast but otherwise identical stimulus component (as in Experiments 8.4 and 8.5). Can both of these results be explained adequately within the framework described in the introduction? As mentioned above, the standard model of processing of spatial patterns assumes a bank of relatively narrowly tuned (about 1–2 octave bandwidth), independent mechanisms (e.g. Pantle & Sekuler, 1968; Blakemore & Campbell, 1969). There are generally two sub-assumptions contained within the assumption of independence—independence of noise, and independence of signal. The first states that the noise in one mechanism is not statistically correlated to the noise in any other channel—indeed, when the noise between two channels is correlated, performance of an observer which assumes independent noise sources decreases, as the correlated noise tends to be mistaken for signal. The assumption of independence of signals is that the signal present in one channel will not affect the response of another channel. This assumption has been shown to be not entirely valid, and several reports indicate the presence of a mutual, tonic inhibition between mechanisms tuned to frequencies up to 3 octaves apart (e.g. DeValois, 1977, and see a review in Olzak and Thomas, 1986). Can this inhibitory interaction among SF channels account for the data?

Consider first the results of Experiment 8.5 on detection thresholds, which shows the most straightforward effect of inhibition: a decrement in performance. If one assumes that there is a weak inhibition between mechanisms tuned to widely separated spatial frequencies, then this result is easily explained, as described in Figure 8.13.

148
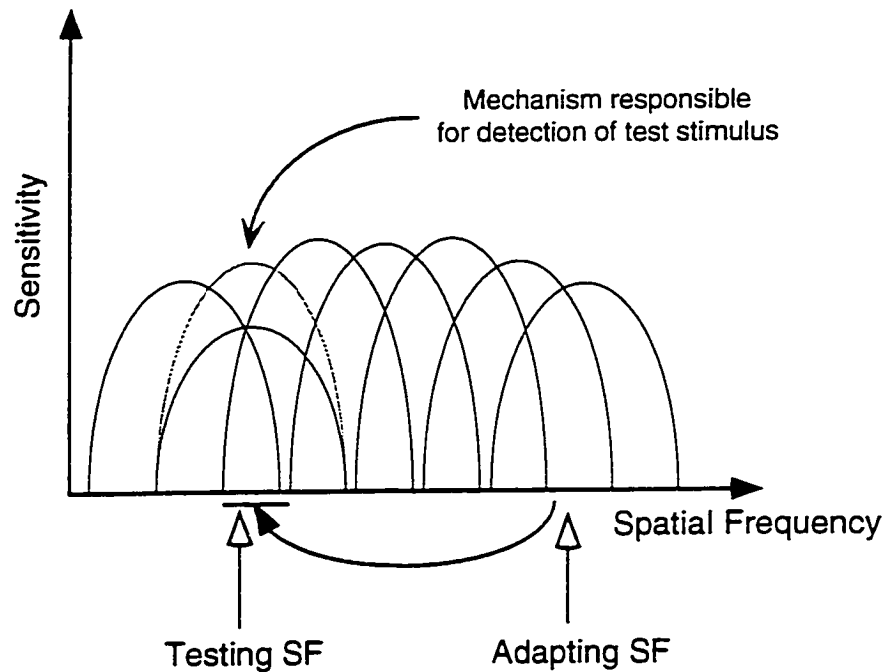
Figure 8.13: Effect of high SF inhibition on the tuning function of the mechanism responsible for the detection of the low SF grating.

The detection threshold for a 0.5 cpd can be directly related to the sensitivity of the mechanism whose sensitivity peak is at 0.5 cpd. Inhibition can be viewed as either a subtractive or a divisive influence on the tuning function of the mechanism—the tuning function in Figure 8.13 uses a divisive inhibition scheme for illustrative purposes, but a subtractive mechanism yields the same effect—a lower sensitivity peak at 0.5 cpd means that a larger signal is required for reliable detection—thus the increase in detection threshold.

Consider now the case of speed discrimination. We will assume for this discussion that discrimination is performed using a set of mechanisms tuned in both spatial and temporal frequencies, but otherwise like the line-element model of Wilson (see e.g. Wilson & Gelb, 1984). In this model, the stimulus is analyzed by a bank of processing pathways. Each pathway consists of a mechanism tuned in spatial frequency and orientation. The output of each mechanism which undergoes a contrast nonlinearity,

149

has Gaussian noise added to it, and then is used to perform the discrimination. This model is able to account for a variety of spatial discrimination data, hence it appears to be a good starting point for the explanation.

Discrimination ability can be intuitively related to the amplitude of the difference in output of a set of mechanisms to two slightly different stimuli. For example, consider two gratings under the conditions used in Experiment 8.1, Low SF condition, which are just discriminable (i.e. so that their speeds are sufficiently different that their discriminability corresponds to a $d' = 1$). Consider the responses of differently tuned mechanisms to these two stimuli. Clearly, if they are above threshold (which they are in the experimental conditions in question), the response of the mechanism which has peak tuning at 0.5 cpd and 1 Hz will not differ much whether one stimulus is presented or the other. Maximal differences in responses will be in by the 'off-frequency' mechanisms, which are tuned to either lower or higher values of spatial and temporal frequency (Regan & Beverley, 1983). Thus we can consider as the most important mechanism for discrimination at a given speed the response of a mechanism which is tuned slightly "off" the mean stimulus properties.

The contrast nonlinearity which occurs before the discrimination stage is such that differences in small input signal get amplified (the nonlinearity starts out accelerating), while the same differences in large input signals get minimized (the nonlinearity ends up compressive). The specific shapes which have been proposed for this nonlinearity vary slightly among authors (see e.g. Legge & Foley, 1980; Foley & Legge, 1981; Carandini & Heeger, 1994; Carandini et al., 1996), but for example Legge and Foley have proposed a function which has the "bend", or "knee" at around 1.5% contrast. In other words, for contrasts lower than 1.5%, the function is expansive, and for contrasts above 1.5% the function is compressive. Now suppose that in the Low SF condition, the response of the off-frequency mechanisms corresponds to the compressive part of the non-linearity. The high SF grating at low contrast

150
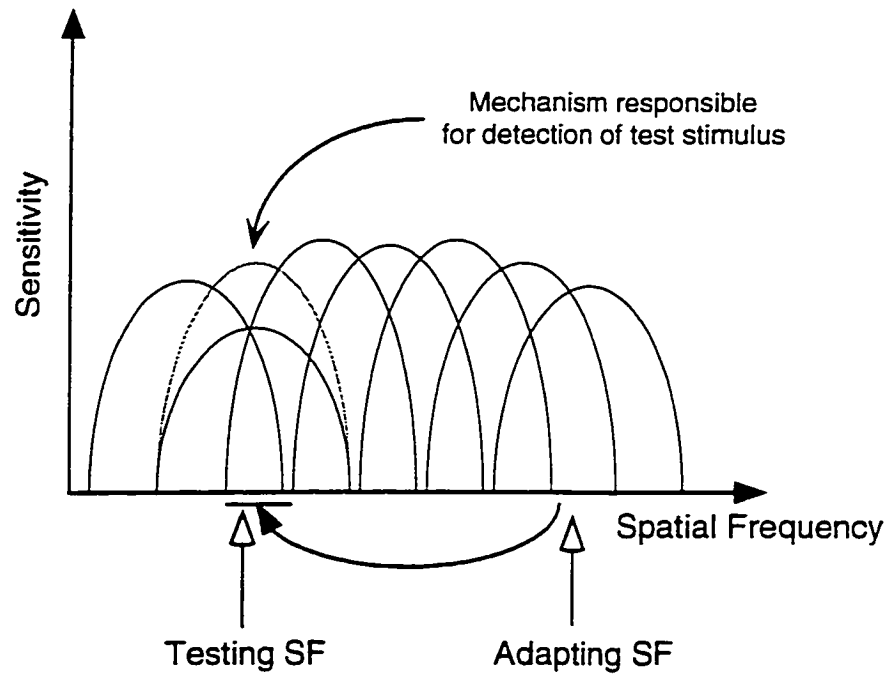
Figure 8.14: How low-grade inhibition can result in increases in discriminability (or, conversely, lower discrimination thresholds). The open arrows indicate the testing and adapting frequencies, while the filled arrow indicates the cross-spatial frequency inhibitory interaction.

will, according to DeValois (1977), inhibit all spatial frequency channels, including those at and around 0.5 cpd. The effect of this inhibition on the on-frequency mechanism will be as described above, and accounts for the detection data. The effect on the off-frequency mechanism will be to shift the tuning function of the mechanism down towards lower response ranges. If one supposes that this shift in response is such that the response by the off-frequency mechanism to the two stimuli now lies "below the knee" of the contrast nonlinearity, then the same physical difference in the stimuli will result in greater response differences, thus in greater discriminability (conversely a given discriminability will be achievable with a smaller physical stimulus difference). This argument is described graphically in Figure 8.14.

How can this framework be extended to deal with the high-contrast effects seen in Experiment 8.4? There are two possible mechanisms which can be invoked to

151

explain why discrimination performance can worsen when high contrast high SF gratings are used. The first is that the inhibition is strong enough to completely suppress the response of the off-frequency to at least one of the two gratings. In such a case (illustrated in Figure 8.15), discrimination can be considered to reduce to a detection task by the off-frequency mechanism of one of the two stimuli, and that task can be made increasingly difficult by increasing the amount of inhibition applied to to that mechanism. This explanation has the admittedly weak backing of the subjective reports of some observers, who claimed that with the high contrast high SF mask, while they felt they could see the low SF grating (the on-frequency mechanism is still sensitive enough), they were unable to judge how fast it was going (the off-frequency mechanisms, according to this explanation, were supressed). The schematic makes it clear that subtractive inhibition is assumed in this case. If instead, divisive inhibition is assumed, then the discrimination performance can be affected without the detectability necessarily vanishing. This is illustrated in Figure 8.16. In this case, the effect of the inhibition is to decrease the slope of the affected mechanism, so that a greater change in stimulus in needed in order to yield a given change in mechanism response.

8.10   Summary

The aim of this research was to investigate whether the visual system could integrate across spatial frequency bands when doing speed perception. Indeed, my first understanding of these effects (Ascher, Welch, & Festa, 1996; Ascher, Welch, & Grzywacz, 1997) was that indeed this was occurring, and that the high contrast inhibition was due to either an attentional effect, contrast-based segregation of the signals into two distinct signal pools, or to some unexplained sampling artifact. The fact that performance is equally good or better when constant speed high SF masks are used as when

152
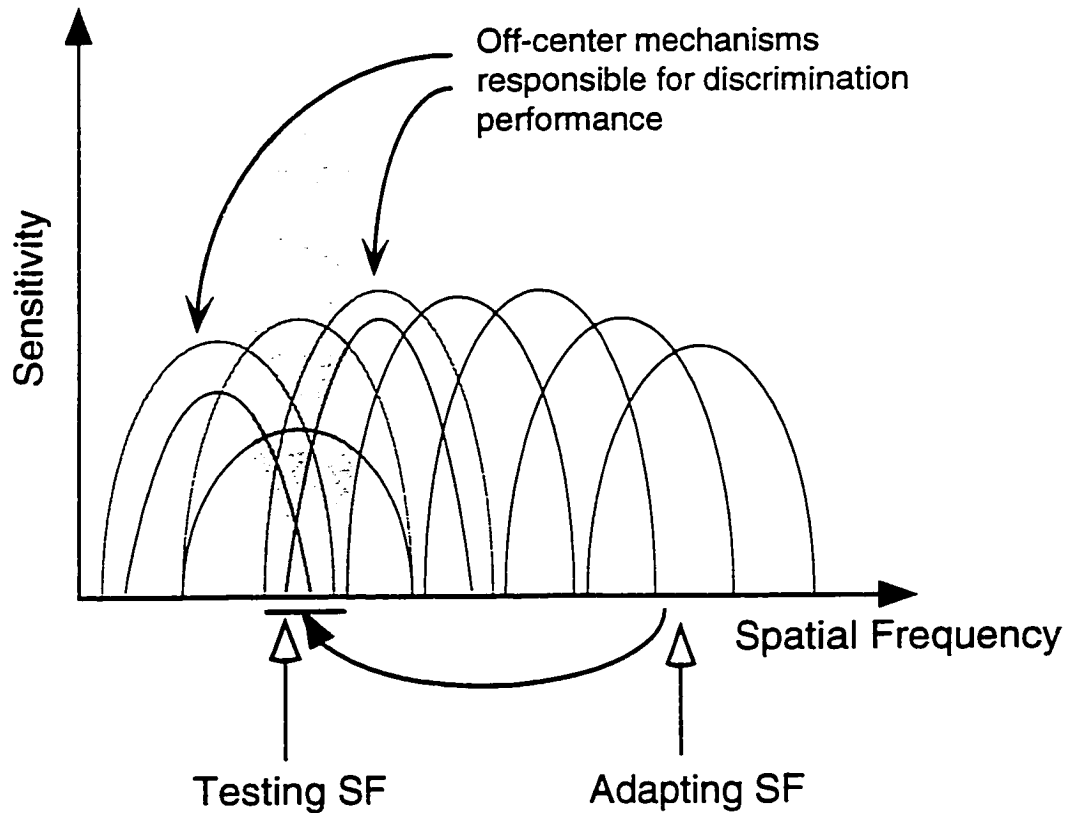
Figure 8.15: Schematic of how large inhibitions of the off-frequency mechanisms can result in a decrease in performance, using subtractive inhibition. Several mechanisms sensitive to the adapting stimulus have reduced sensitivities, which results in the *slope* of their sensitivity to decrease at the frequency of the adapting stimulus.
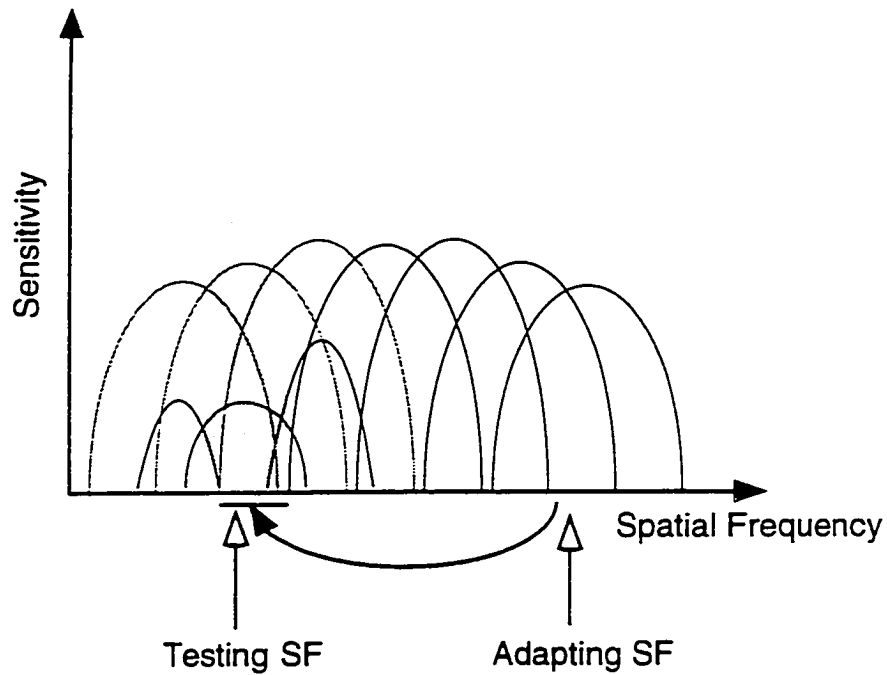
Figure 8.16: Schematic of how large inhibitions of the off-frequency mechanisms can result in a decrease in performance, using divisive inhibition. Several mechanisms sensitive to the adapting stimulus have reduced sensitivities, which results in the *slope* of their sensitivity to decrease at the frequency of the adapting stimulus.

perfectly correlated high SF gratings are used argues to my mind convincingly that if there is such an integration, it is not the main source of the improvements in performance that we report. Consequently, a model was presented, which, by appealing to the effect of the contrast nonlinearity as modulated by inter-mechanism inhibition, accounts for both patterns of behavior. That said, some of the data requires further discussion given the analysis presented above. Recall that in Experiment 8.3, different observers behaved differently for the uncorrelated stimuli and for the anti-correlated stimuli. In other words, the fact that the high SF grating varied its speed across trials did have a deleterious effect on performance. It should be noted that

154

these displays are somewhat odd, in that the gratings appear to "cohere" on some trials (when the speeds of the two gratings happen to be similar), while in other trials in the same block, appear quite distorted (when the speeds of the two gratings happen to be quite dissimilar). Some observers, however, were not affected by these misleading trials. It is possible that with practice, the apparent loss in performance due to the speed signal yielded by the high SF stimulus could disappear. Regardless, the fact that a high SF grating is affecting low SF grating performance *in either direction* is sign of cross-channel interactions.

All of the experiments described above are consistent in one main point, that is that, at least in the context of a speed discrimination task, and under these admittedly extreme conditions of low contrast and short durations, processing of translating gratings is not performed by independent mechanisms. A qualitative explanation of the data was presented, using two well-established aspects of processing of spatial stimuli, namely the interaction across spatial frequency mechanisms first identified by DeValois (1977), and the nonlinear contrast compression first argued for by Nachmias and Sansbury (1974), and subsequently explored in detail by, e.g. Legge and Foley (1980), Legge (1981), Wilson et al. (1983). Briefly, high spatial frequency gratings (3 and 6 cpd) inhibit the response of mechanisms tuned to low spatial frequencies (0.5 cpd). This inhibition results in the increase in detection thresholds reported in Experiment 8.5. This inhibition also acts on the off-frequency mechanisms responsible for the speed discrimination performance. Low-contrast high SF gratings result in a slight shift in the operating range of the off-frequency mechanism, so that the difference in the mechanism's response to nearly discriminable stimuli is enhanced by the accelerating portion of the contrast nonlinearity, yielding the improvements in discrimination performance noted in Experiments 8.1, 8.2, 8.3, and 8.6. High contrast high SF gratings result in large inhibition of the low SF off-frequency mechanism, which effectively lose the ability to detect the low SF grating, thus yielding

155

the decrements in performance noted in Experiment 8.5.

Chapter 9

A Revised Model of
Speed Perception

How can one incorporate the psychophysical results described in Chapters 4 through 8 into a new computational model of speed perception? While the experiments described earlier were all designed in the context of existing models, and especially the experiments in the last two chapters were aimed at testing predictions which stemmed from the models, their conclusions are difficult to use when it comes to choosing between, e.g., gradient models and correlational models. The adaptation results argue against the Watson and Ahumada (1985) model, but the other models can be accomodated quite well. Recent data from Jackson and Welch (personal communication) argues against two-channel models, as mentioned in Chapter 2, but further work is needed to more fully specify the exact channel shapes. Finally, the spatial frequency integration question is left unanswered by the results of Chapter 8, which, while interesting, do not help identify the *mechanism* used by the visual system to compute speed.

Given this conundrum, this chapter cannot hope to make any firm conclusions regarding *the* speed model which will supplant all others. However, I take this

opportunity to present a model which has emerged as a new take on an old model, and which, it is hoped, will allow the future quantitative testing of a model with hopes of good matches between psychophysical results and simulations[1].

9.1 Outline

The model of speed perception presented here can mostly be considered to be a "psychophysically correct" version of that presented in Grzywacz and Yuille (1990). The general conceptual framework of that model is used, but the mathematically convenient filter shapes which that model used are replaced with psychophysically derived functions, and the estimation strategy is further refined. The purpose of these manipulations is that while the model in Grzywacz and Yuille (1990) is proven to arrive at the mathematically correct answer, it is not a model of discrimination, being noise-free. The model presented here incorporates both peripheral and central noise sources, with the hopes that psychophysical discrimination data can be replicated, mostly thanks to the careful choice of filter functions. The estimation strategy used is a maximum likelihood method, which is quite logically a popular one.

The first section will give a formal presentation of the filters used in the model, followed by the proposed mechanism for speed estimation within a spatial-frequency channel. The following section introduces the noise processes, which allow the model to produce noisy estimates, which are needed to yield imperfect speed estimates, a crucial step in reproducing human data. Finally, a set of less formal proposals

---

[1]The model presented in this chapter is the result of discussions between Dr. Grzywacz, Prof. Welch and me over the last two years. It is presented here not so much as a individual's contribution (which it is not), but rather as my current thoughts regarding how an interesting speed model could work. The good ideas in the model are the result of a collaborative effort, but I am to blame for the more speculative proposals in the later part of the chapter.

158

for extensions to the model are presented, in order to account for some of the data presented in the earlier chapters.

9.2 The Noise-free theory

The noise-free theory can be decomposed into three aspects, which will be considered in turn. A general description of the filter set being used, both their number and their general properties. Candidates functions used for the spatial and temporal tuning shapes for the filters are then presented, based on some of the psychophysical results presented in Chapter 2. Finally, the speed estimation procedure within a spatial frequency band is presented.

9.2.1 Derivation of the Filter Response

The estimation of speed is assumed to be performed by spatiotemporal filters applied to the retinal input. The visual stimulus is filtered by a set of filters tuned both to spatial and temporal frequencies. There are I sets of spatial-frequency tuned functions, and J sets of temporal-frequency tuned functions. There are thus $I \cdot J$ filters, each with a given spatial- and temporal-frequency tuning functions (see Figure 10). Existing data (Wilson et al., 1983, Section C.2) suggests that I is at least 6 and J is at least 3. Each filter $F_{i,j}$ is a function of the stimulus $\mathbf{r}$ and time t, and can be characterized by its best spatial and temporal frequencies $\Omega_{\mathbf{r},i}$ and $\Omega_{t,j}$: $F_{i,j}(\mathbf{r}, t : \Omega_{\mathbf{r},i}, \Omega_{t,j})$ with $i \in \{1 \cdots 6\}$ and $j \in \{1, 2, 3\}$. For the purposes of description, and because in the experiments proposed here motions are only along a single direction, only filters with a single preferred direction and considered, but the mathematical analysis can be performed identically across preferred directions. The stimulus S is a function of space $\mathbf{r}$ and time t: $S(\mathbf{r}, t)$.
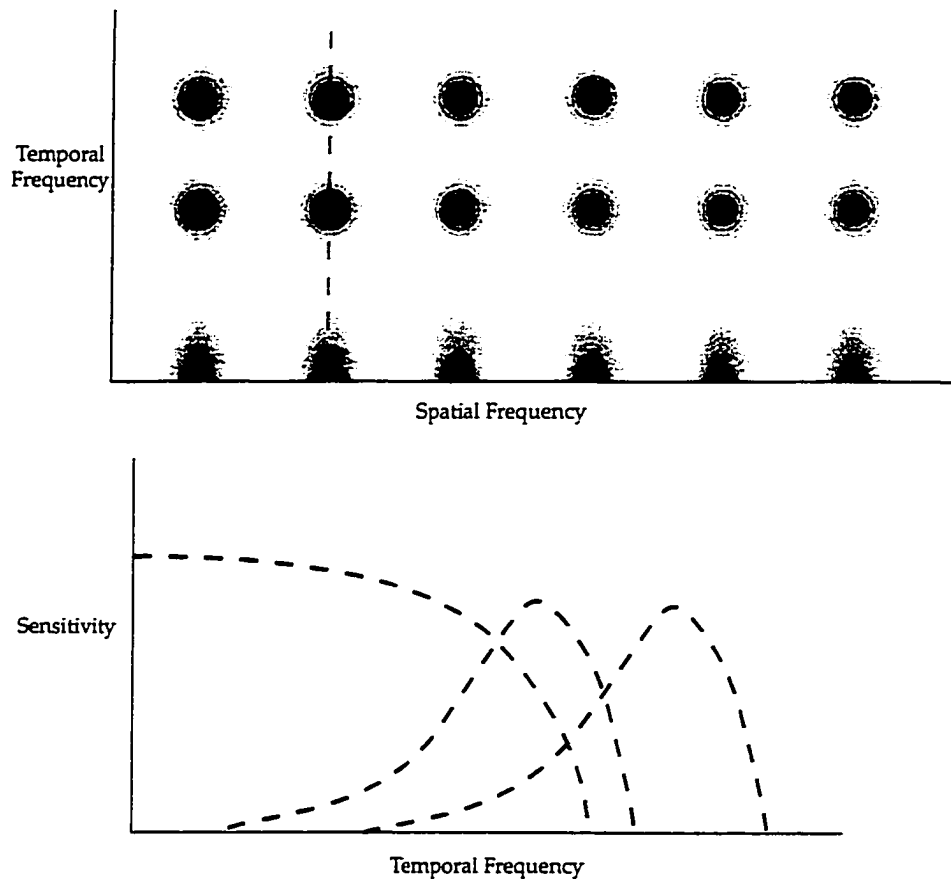
Figure 9.1: Schematic representation of the $F_{i,j}$ filters. In the top panel, each filter is represented by its peak of sensitivity (shaded "bumps"). Depicted are three different best TFs ($J = 3$) and six different best SFs ($I = 6$). The bottom panel plots the TF sensitivities of the filters for a given SF (along the dashed line in the top panel). Three different filters are visible (dashed curves), corresponding to the low-pass and two band-pass channels.

Assumption 1 (Small Receptive Fields) *The receptive fields of the detectors are small enough that the motion of the stimulus in front of any given detector can be considered a linear translation.*

If the receptive fields are assumed to be small enough, the stimulus S can be rewritten as a function of $\mathbf{r} - \mathbf{v}t$, where $\mathbf{v}$ is the speed of the stimulus, that is, stimulus $= S(\mathbf{r} - \mathbf{v}t)$

160

Let us first consider the case of a translating sinusoidal wave. A translating sine wave of amplitude A can be described as: $Ae^{i\omega_r(r-vt)}$, where $\omega_r/2\pi$ is the spatial frequency of the sinewave. Because the stimulus is translating at a velocity v, $\omega_r \cdot v = \omega_t$, where $\omega_t/2\pi$ is the temporal frequency.

Assumption 2 (Linearity) *The filters are linear.*

· If the filters are assumed to be linear, then the response of a filter F to a given stimulus S can be rewritten as the convolution $F * S(r, t)$. The output of a filter to a sinewave stimulus, therefore, is:

$$\text{Sine filter output} = A \iint dr\, dt\, e^{i\omega_r(r-vt)} F(r_0 - r, t_0 - t) \tag{9.1}$$

$$= A\tilde{F}(\omega_r, \omega_t : \Omega_r, \Omega_t) \tag{9.2}$$

where $\tilde{F}$ is the fourier transform of F. The *response* of the filter to be the magnitude of its output is defined to be $A|\tilde{F}|$.

In the Fourier domain, any translating stimulus can be described as the product of a spatial component (corresponding to the texture of the stimulus) times a translation component (corresponding to the motion of the stimulus):

$$\text{General stimulus} = g(\omega_r) \cdot \delta(\omega_r \cdot v + \omega_t) \tag{9.3}$$

where δ is the Dirac delta function, zero everywhere but where its argument is zero, and with an integral of 1. The response of a filter to this general translating stimulus is therefore:

$$\text{Response} = \left| \iint d\omega_r\, d\omega_t\, g(\omega_r)\delta(\omega_r \cdot v + \omega_t) \tilde{F}(\omega_r, \omega_t : \Omega_r, \Omega_t) \right| \tag{9.4}$$

161

This equation follows from Equations 6 and 7 through the following reasoning: If the response to a single sine wave is like in Equation 6, then for a general stimulus, that is, with multiple sinewave components, the linear portion of the response is the superposition of all individual sinewave responses weighted by the spectral distribution of the stimulus, that is, Equation 7. Using the fact that the integral of the delta function is 1, one gets:

$$\text{Response} = \left| \int d\omega_{\mathbf{r}}\, g(\omega_{\mathbf{r}})\, \tilde{F}(\omega_{\mathbf{r}}, -\omega_{\mathbf{r}} \cdot \mathbf{v} : \Omega_{\mathbf{r}}, \Omega_t) \right| \tag{9.5}$$

Assumption 3 (Locally Constant Spatial Frequency Spectrum) *The receptive fields of the detectors have spatial frequency tunings that are narrow enough that the spatial spectrum of the stimulus in front of any given detector can be considered constant.*

This assumption, as well as Assumption 1, cannot be valid in general, since they depend on the stimulus. However, one can postulate that these assumptions correspond to knowledge that the visual system has about the statistics of images. In other words, that these assumptions are normally true for real images and when the stimuli do not fulfill them, predictable visual illusions should occur.

Thus, the texture component can be taken out of the integral, yielding:

$$\text{Response} = \left| g(\Omega_{\mathbf{r}}) \int d\omega_{\mathbf{r}} \tilde{F}(\omega_{\mathbf{r}}, -\omega_{\mathbf{r}} \cdot \mathbf{v} : \Omega_{\mathbf{r}}, -\Omega_t) \right| \tag{9.6}$$

9.2.2 Filter Shapes

The filter functions used will have considerable impact on performance. These filters need to be direction selective. The first assumption is that the spatial profile corresponds to that of a Gabor filter G,

162

$$G(\mathbf{r} : \Omega_{\mathbf{r}}) = \frac{1}{2\pi\sigma} e^{\frac{|\mathbf{r}|^2}{2\sigma^2}} e^{i\,\Omega_{\mathbf{r}}\cdot\gamma} \tag{9.7}$$

where $\Omega_{\mathbf{r}}$ is the spatial frequency of the carrier sinewave, σ is the standard deviation of the modulating gaussian and γ defines the spatial orientation of the Gabor (Gabor, 1946; Daugman, 1985). Using Gabor filters is not only in agreement with physiology (Daugman, 1985), but also is the ideal filter for fulfilling Assumptions 1 and 3 simultaneously. This filter is the one among all physical filters that minimizes the product of spatial bandwidth by spatial frequency bandwidth (Daugman, 1985).

The temporal profile used will be a linear combination of two temporal functions T_1 and T_2, the specific contribution of each being determined by the location in the receptive field. This spatial dependence of the temporal functions is based on the analyses of simple-cell data by Kontsevich (1995) and Reid, Soodak, and Shapley (1991). Based on the Reid et al. (1991) data, the filters can be modeled as:

$$F(\mathbf{r}, t) = G(\mathbf{r}) \cdot [s(\mathbf{r} \cdot \mathbf{n})T_1(t) + (1 - s(\mathbf{r} \cdot \mathbf{n}))T_2(t)] \tag{9.8}$$

where $\mathbf{n}$ is a spatial vector which determines the direction along which the T_1/T_2 combination occurs, s is a sigmoid which flattens $[-\infty, \infty]$ down to $[0, 1]$ such that when $\mathbf{r} \cdot \mathbf{n}$ is large and positive, $s(\mathbf{r} \cdot \mathbf{n})$ is close to 1 and the linear combination is mostly composed of T_1, when $\mathbf{r} \cdot \mathbf{n}$ is large but negative, $s(\mathbf{r} \cdot \mathbf{n})$ is close to 0 and the linear combination is mostly composed of T_2, and $s(0) = 0.5$.

The shapes of T_1 and T_2 will determine the sustained or transient characteristics of the filters. These functions will be based on Alpha functions, such the sustained and transient functions h_0 and h_1 used by Watson (1986) to fit temporal contrast sensitivity data:

163

$$h_0 \;=\; \xi h(t, n_1, \tau) \tag{9.9}$$

$$h_1 \;=\; \xi h(t, n_1, \tau) - \eta h(t, n_2, \kappa\tau) \tag{9.10}$$

where

$$h(t, n, \tau) = u(t)(\tau(n-1)!)^{-1}(t/\tau)^{n-1}e^{-t/\tau} \tag{9.11}$$

and $u(t)$ is the unit step function. Watson (1986) found that values of $\tau = 4.94$, $\xi = 1.0$, $n_1 = 9$, $n_2 = 10$, $\eta = 1.0$ and $\kappa = 1.33$ fit the data well (see Figure 9.2).
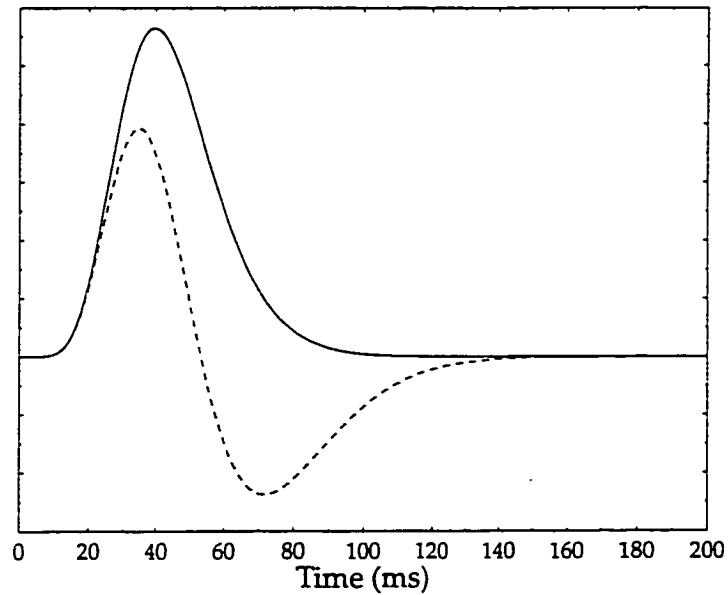

Figure 9.2: The Alpha functions h_0 (solid line) and h_1 (dashed line) used by Watson (1986).

The filters are now completely specified, and the response of each filter can be obtained for any given stimulus. To compute the response of a filter, one needs to

164

get the integral over $d\omega_r$ of the Fourier transform of the filter, F. First let us identify the Fourier transform $\tilde{F}$:

$$\tilde{F} = \int d\mathbf{r}\, dt\, e^{i(\omega_r \cdot \mathbf{r} + \omega_t t)} F(\mathbf{r}, t) \tag{9.12}$$

$$= \tilde{T}_1(\omega_t) \int d\mathbf{r}\, e^{i\omega_r \cdot \mathbf{r}} G(\mathbf{r}) s(\mathbf{r}, \mathbf{n}) + \tilde{T}_2(\omega_t) \int d\mathbf{r}\, e^{i\omega_r \cdot \mathbf{r}} G(\mathbf{r})(1 - s(\mathbf{r}, \mathbf{n})) \tag{9.13}$$

The integral needed is:

$$\int d\omega_r (\tilde{F}) = \int d\mathbf{r}[\tilde{T}_1 \cdot G \cdot s(\mathbf{r} \cdot \mathbf{v}) + \tilde{T}_2 \cdot G \cdot (1 - s(\mathbf{r} \cdot \mathbf{v}))] \int d\omega_r\, e^{i\omega_r \cdot \mathbf{r}} \tag{9.14}$$

Because

$$\int d\mathbf{r}\, \delta(\mathbf{r})\, e^{i\omega_r \cdot \mathbf{r}} = 1 \tag{9.15}$$

the integral

$$(1/2\pi)^2 \int d\omega_r\, e^{i\omega_r \cdot \mathbf{r}} = \delta(r) \tag{9.16}$$

since this integral is the conjugate of the inverse Fourier transform of 1. Hence,

$$\int d\omega_r (\tilde{F}) = (2\pi)^2 \int d\mathbf{r}[\tilde{T}_1 \cdot G \cdot s(\mathbf{r} \cdot \mathbf{v}) + \tilde{T}_2 \cdot G \cdot (1 - s(\mathbf{r} \cdot \mathbf{v}))]\delta(\mathbf{r}) \tag{9.17}$$

$$= (2\pi)^2 G(0)[\tilde{T}_1 \cdot s(0) + \tilde{T}_2(1 - s(0))] \tag{9.18}$$

$$= (2\pi)^2 G(0)\frac{1}{2}[\tilde{T}_1(-\omega_r \cdot \mathbf{v}) + \tilde{T}_2(-\omega_r \cdot \mathbf{v})] \tag{9.19}$$

Referring back to Equation 11, one can see that the response of the filter output to a given stimulus is given by:

165

$$\text{Response} = |g(\Omega_r)\frac{\pi}{\sigma}[\tilde{T}_1(-\boldsymbol{\omega}_r \cdot \mathbf{v}) + \tilde{T}_2(-\boldsymbol{\omega}_r \cdot \mathbf{v})]| \tag{9.20}$$

$$= \kappa|\tilde{T}_1(-\boldsymbol{\omega}_r \cdot \mathbf{v}) + \tilde{T}_2(-\boldsymbol{\omega}_r \cdot \mathbf{v})| \tag{9.21}$$

where κ is a positive constant.

9.2.3 Speed Estimation

Equation 9.21 is an exact formula for the filters' outputs. The next task is to define how the filters are combined to yield a measure of speed. (The individual filters are not speed tuned; their response depends on the spatial frequency, contrast and temporal frequency of the stimulus.)

First an estimate of speed in each spatial frequency band will be obtained, and then a best fit of the speed estimates across the spatial frequencies will be performed, using a maximum-likelihood estimation procedure.

The response of a given filter can be influenced by two factors: the speed of the stimulus (the desired measure) and the fraction of the stimulus spatial frequency spectrum which falls within the bandwidth of the filter (function $g(\Omega_r)$ in Equation 9). To obtain the former, one needs to discount the effects of the latter. This can be done by noticing that all of the filters in a given spatial frequency band are equally affected by the texture of the stimulus, but unequally affected by the speed of the stimulus. Note that the texture of the stimulus includes both spatial-frequency effects and contrast effects. By considering simultaneously the outputs of all the filters tuned to the same Ω_r but different Ω_t, one can obtain an estimate of speed unbiased by spatial frequency spectral distributions.

Consider the responses, R_i, of the I filters, F_i, which fall in a given spatial frequency band i. In a noise-free system, these responses are exactly given by the

166

product of a textural component (which is stimulus specific) times the sensitivity of the filters at a given speed. Thus, by normalizing the filter responses across a spatial frequency band (by dividing R_i by $\sum_i R_i$), one obtains a set of normalized responses which correspond uniquely to one speed $|\mathbf{v}|$ (the speed of the stimulus). Therefore, in a noise-free system, obtaining the set of normalized responses would solve the problem of measuring speed.

9.3 Model Implementation: Noise & Numerical Methods

The theory described above outlines a possible scheme for measuring stimulus speed given a set of filters. The purpose of this section is to outline how this scheme accounts or fails to account for human perceptual data. The theory above is perfect— any given spatial frequency band contains enough filters to give perfect estimates of stimulus speed (assuming the stimulus has energy in that spatial frequency band). The model described below describes how the estimation procedure needs to be modified if noisy filter responses are taken into consideration. Since the model is a model of human speed perception, the effect of noise on the model will be compared with human performance by comparing the result of noisy simulations with psychophysical measures of noise: discrimination thresholds.

The noise in the system is assumed to be characterizable by the addition of Gaussian additive noise to the response of the filters. If M_i is the *measurement* (not normalized) of the output of filter F_i, then

$$M_i = R_i(\mathbf{v} \cdot \Omega_{\mathbf{r}}) + n_i(\sigma_i) \tag{9.22}$$

where n_i represents the noise, characterized by the standard deviation σ_i and ampli-

167

tude $1/2\pi\sigma_i^{1/2}$ of a Gaussian with zero mean.

This simple noise (which is *peripheral noise* within the framework presented in Chapter 4) has relatively complex consequences, since it means that any distribution of responses across a bank of filters even with identical SF tuning functions, cannot be unambiguously attributed to a single speed, since the response is due to a combination of the normal filter response, and to noise.

The estimation of speed in a given spatial frequency channel must therefore be made on the measured values M_i instead of the "true" responses R_i. Let us begin with the expected responses $E_i(\mathbf{v})$ of the filters for speed $\mathbf{v}$. These responses are normalized in the sense that they do not include the dependence on texture and contrast embodied by κ in Equation 9.21. Fortunately, the effect of κ is the same for all filters within a single spatial-frequency band. The only thing that changes from filter to filter is the amount of noise added, since noise is random and independent across filters. Therefore, for all $i \in \{1 \cdots J\}$

$$M_i = \kappa E_i(\mathbf{v}) + n_i. \tag{9.23}$$

For each choice of $\mathbf{v}$, this is a linear system of J equations with $J + 1$ unknowns (κ and the n_i). If somehow κ is known from independent estimates by the visual system, then one can solve for the n_i; this strategy was called "the extra-information strategy" by Grzywacz and Yuille (1990). Otherwise, one must compute the n_i for each choice of κ and $\mathbf{v}$. For each such choice, each computed n_i has an associated probability p_i, which can be computed from Equation 9.22.

Consequently, the probability of κ and $\mathbf{v}$ given the measurements M_i is

$$P(\kappa, \mathbf{v}|M_i) = \prod_{i=1}^{J} p_i(\kappa, \mathbf{v}|M_i). \tag{9.24}$$

(Actually, to be exactly Bayesian, this probability must be weighted by the *a priori*

168

probabilities of κ and $\mathbf{v}$. In other words, if the visual system knew what speeds, textures, and contrasts are more likely to occur in the world, then it could take this information into account to compute $P(\kappa, \mathbf{v}|M_i)$. However, since this information is not available to us, it is assumed for Equation 9.24 that all speeds, textures, and contrasts are equally likely.)

We postulate that to estimate speed, the visual system finds the $\mathbf{v}$ and κ that maximizes the probability $P(\kappa, \mathbf{v}|M_i)$ in Equation 9.24.

This probability can then further be used as a measure of the "confidence" that the speed of the stimulus is indeed $\mathbf{v}$. For example, if the probability is very close to 1 for a specific value of $\mathbf{v}$, then this means that very small amounts of noise were needed to convert the expected responses of the filters into the measured responses. If the probabilities are small for all $\mathbf{v}$'s, it is because all values of $\mathbf{v}$ required improbable amounts of noise to make the expected and measured responses match.

To work with Equation 9.24, the n_i (noise amplitudes) need to be converted into the probabilities p_i. This conversion is performed by assuming that the visual system knows the σ_i in Equation 9.22. However, this does not mean that σ_i must be constant. On the contrary, a large body of evidence suggests that at suprathreshold contrasts, contrast-noise increases with background contrast (Legge & Foley, 1980) and that the slope of this increase depends on spatial and temporal frequencies (Burbeck & Kelly, 1981; Phillips & Wilson, 1984). This noise-contrast relationship is commonly known as the dipper function. Two assumptions are made about the dipper function: 1) That the noise increases not as a consequence of absolute contrast in the image (which is very hard to define for arbitrary images), but as a consequence of the filter responses (that is, neural noise). Because these responses are proportional to sine wave contrasts for a large range of contrasts, this assumption is consistent with the body of evidence on the dipper function. 2) That the visual system knows the dipper function when response, rather than contrast, appears in the horizontal axis. With

169

these assumptions in hand, for each choice of v and κ, one has the expected response $\kappa E_i(\mathbf{v})$ and therefore σ_i.

The procedure outlined above yields an estimate of speed and the confidence in that speed for a given set of filters with common spatial-frequency tuning. There are two possible ways to obtain a speed estimate for the entire filter bank.

One way of obtaining an estimate of speed using information from all spatial-frequency channels is to perform a weighted regression of the estimates at each spatial-frequency band, weighted by the confidence in that estimate yielded by the procedure above. Thus the spatial-frequency channels with the most "believable" information (most likely due to richer stimulus energy in their bandwidths) would be given greater weight. This strategy, while likely to work and to be efficient in terms of computational costs, is not the most mathematically optimal. It breaks the principle of least commitment—information about the probability density function in each channel is lost if all that is considered for the system-wide estimation is the peak of those distributions.

An alternative strategy is to perform the probability computation on the output of all the filters in all the channels for any given velocity and combinations of κ's (κ's could be different for different spatial-frequency bands). This will yield a unique probability density function for the entire filter bank. Computationally, this can be performed by multiplying all of the channel-specific probability density functions by one another just as the filter-specific functions were multiplied to yield the channel-specific functions.

The peak of the resulting function can then be used to simulate perceived speed.

170

9.4 Extensions

The model just outlindes provides a mechanism for obtaining perceived speed in such a way as to include peripheral noise (i.e. noise at the level of the filters). This noise allows the model (in theory) to provide both exact matches (which human observers generally do, given enough signal), as well as measures of discrimination, which can be compared with the literature mentioned at several places in this dissertation. For example, simulations using an implementation of the model as formulated could be compared to many of the experiments reviewed in Chapter 2, as well as the new data in Chapters 4, 5 and 6. However, it should be clear that the model is unable to deal with the adaptation data of Chapter 7, and that while the model will yield improvements for stimuli with spatially rich spectra, it will not match the data of Chapter 8, especially the last crucial experiment showing interactions between SF channels. This last section will present possible extensions to the model which would possibly allow it to account for this data.

9.4.1 Dynamics

The model presented above is a steady-state model. It does not incorporate any changes to the response of the filters as a function of past exposure. There are two known general effects which it can therefore not model. The first is the effect of recent exposure to long-lasting stimuli, which yields adaptation effects as described in Chapter 7. Two types of adaptation are frequently used, subtractive and divisive inhibition. Fairly exhaustive (and exhausting) studies would be needed to provide data which could allow the disambiguation between these two, if the study of Mandler and Makous (1984) is any indication. Their study yielded a good model fit, but required setting different adatability levels for each of the two channels. It would be interesting to see if their data could be fit using the model just presented and their

171

adaptation mechanism, and whether the differences in the temporal functions used would yield a different answer regarding the type of inhibition and/or the adaptability levels for the two channels.

A second type of long-term change should be considered as well. That is, it is quite clear when doing speed experiments that observers undergo long-term plasticity. They both learn how to do the task, as well a they get better. The spatial frequency integration effects of Chapter 8 were especially susceptible to learning. While there is a great deal of literature on learning systems, there is too little data available on speed perception learning to usefully constrain the model, so no long term plasticity modification is proposed at this stage.

9.4.2 SF Interactions

The results of Chapter 8 indicated that a speed model, while traditionally anchored in the world of temporal vision, needs to take into consideration many results from the world of spatial vision. Thus interactions (in this case inhibitory) between spatial frequency channels should be taken into account. It is hard to know how one should model such interactions, however, since the data in Chapter 8 are to my knowledge the first showing an effect of cross-SF interactions on speed perception.

9.5   Conclusion

This chapter has sketched a new model of speed perception which incorporates psychophysically derived filter shapes, computationally strong mechanisms for computing veridical speed, and a natural mechanism for dealing with filter noise which can in turn yield discrimination thresholds. Extensions have been briefly described which could allow the model to fit the data presented in the previous chapters which could not be accounted for by the basic model. Whether the model matches the data or

not will have to wait for its implementation and simulations.

Conclusion

10.1 Summary

This dissertation has presented a set of experiments concerned with aspects of speed perception. These experiments were motivated by a desire to better understand how speed perception occurs, especially within the context of plausible computational models thereof. The results from each experiment are summarized briefly, followed by a cautionary note on the scope of these results.

The first experiment addressed the important and often overlooked result of Bowne (1990), which argued that temporal frequency discrimination (along with spatial frequency and orientation discrimination, which we are not concerned with here) is not affected by stimulus contrast. The generality of these results was questioned due to some methodological concerns, namely the long duration of the stimulus, and the lack of duration and contrast jitter. My results indicate that, unlike in Bowne (1990), speed discrimination *is* affected by stimulus contrast. This means both that the models of speed discrimination which assume that detector response is monotonically related to stimulus contrast *can* be used and that the parameters of the spatiotemporal channels derived from experiments manipulating stimulus contrast can be used in the study of speed perception.

The second and third experiments report preliminary measurements on the spatial and temporal extent over which speed discrimination can increase. The temporal

174

integration period of the speed mechanism is shown to be between 120 and 200 msec, consistent with previous psychophysical and physiological results. Stimulus durations longer than 200 msec do not result in significantly better performance. In contrast, integration of information over spatial extent occurs up to at least 12 deg of visual angle. Furthermore, as first pointed out by Bowne (1990), the speed sensitive units appear to be elongated along the direction of motion.

The fourth set of experiments showed that the lowpass temporal channel, which is sensitive to stimuli with low temporal frequencies can be adapted by exposure to stationary stimuli, and is an active participant in the computation of speed. After adaptation to stationary gratings, the perceived speed of moving gratings was higher than it would have been otherwise, consistent with some but not all speed models in the literature. This increase in perceived speed is not due to afterimages, spatial frequency adaptation, or contrast adaptation.

The fifth set of experiments tested the ability of the speed system to integrate information across a wide range of spatial frequencies. While an increase in speed discrimination performance was obtained for compound gratings over single gratings, subsequent testing revealed that this increase in performance is not "integration" but rather the consequence of temporal frequency independent interactions between spatial frequency channels. The question of whether speed computation can use information from a broad range of spatial frequencies is left open.

Finally, Chapter 9 presents a sketch of a new model of speed perception, which incorporates psychophysically and physiologically motivated spatiotemporal filters and a computationally optimal speed estimation strategy.

175

10.2 Scope Limitations

Models survive because they account for data elegantly, or become discarded because they need as many knobs and switches as there are data sets to fit. Thus the qualities and limitations of the model sketched in Chapter 9 will be made obvious as it is further developed and tested. Experimental data, however, lasts, provided it is viewed within the appropriate context.

The absolute numbers of discrimination performance reported in this dissertation indicate, by the standards of speed discrimination experts, rather poor performance. As has been mentioned repeatedly, this is deliberate, and the result of careful parameter hunting. Finding stimulus parameters where small stimulus manipulations can yield measurable effects is a slow (and painful, from the subject's point of view) process. When comparing data from these studies with other similar studies, it is important to compare the stimulus durations and contrasts, and whether the displays have jittered durations and/or contrasts.

As mentioned in the early chapters, all of the stimuli I used were luminance gratings, viewed under mesopic/photopic lighting conditions, foveally. Extrapolating any of the specific results presented here to the domains of scotopic, peripheral or color vision is best avoided. Also as discussed at length in the Introduction, the characteristics of speed discrimination depends on the task. More precisely, the mechanisms by which humans compute retinal speed, whether based on estimates of SF and TF, direct speed estimates, or estimates of position and duration, will depend on the speed, the temporal frequency and the stimulus duration.

One final stimulus dimension needs to be mentioned, and that is the spatial arrangement and extent of the stimulus. As pointed out by Snowden (1997), perceived speed of a grating depends on the size of the grating. Along similar lines, Metha, Bex, and Makous (1997) found that the perceived speed of spatially localized drifting grat-

176

ings depended on whether the orientations of the patches was such that the patches formed a translating, expanding or rotating pattern, so that expansion patterns were perceived to move 30% faster than rotating and translating patterns. Similar results regarding the perceived speed of expanding, contracting and rotating dots have been reported by Geesaman and Qian (1996, in press). Clearly these spatial interactions (possibly due to biases related to ego-motion) are missing in all of the models of speed perception described above, and should be accounted for by a comprehensive speed model.

10.3   A Note on Physiology

The Preface discussed speed perception as a field where physiology and psychophysics can be related, along with studies of computational models. Yet this dissertation has only briefly and infrequently mentioned physiological results. This is deliberate, and reflects my feeling that, as far as speed is concerned, the physiological cards have yet to be played. Almost all computational modellers, from Marr and Ullman (1981) on, cite the results of physiological studies, be they from X and Y cells in the cat retina, magno and parvo cells in the LGN or V1, or cortical cells in early or late visual cortex. And yet all of these modellers appeal to this data in order to argue for their models, in opposition to the competing models present in the literature. This is no doubt because of a selection bias on the part of the modellers. There are many cells in the brain, and many exploratory studies of their behaviors. With enough patience, reports of a wide variety of cellular behaviors can be found in the literature. Marr and Ullman (1981) audaciously proposed a set of theoretically motivated facts about visual cells, and categorized the importance of each prediction's ability to falsify the model if it were disproved. More recently, Simoncelli and Heeger (1997) make four similar proposals. Although these proposals are derived from the theoretical model,

177

their eventual validation would not provide clear counter-evidence against the other models of speed perception, as the predictions are mostly related to the computation of stimulus direction and not speed. The reason for this paucity of theoretically-motivated physiological data for speed is quite simple to explain. Firstly, most of the physiological experiments on cell properties have used stimuli such as bars and dots, mostly for traditional reasons. The use of spectrally complex (and poorly controlled) stimuli, however, has obscured the distinction between temporal frequency and speed, and while many physiological papers refer to the speed-tuning of a cell, their methods do not allow one to differentiate between speed-tuning and temporal frequency tuning. Secondly, if all of the models of speed agree on one fact, it is that is that speed cannot be read from a single cell's firing rate. It is implicit in the distribution of several cells, in ways which are either simple (as in the simplest ratio gradient model) or complex (as in the Grzywacz and Yuille (1990) model). Thus it will probably take the use of some of the more recent multielectrode recording techniques to disentangle between these competing views.

10.4   Closing Comments

The topic of this dissertation was chosen because it appeared to provide fertile ground for interactions between modeling and psychophysical approaches to a common problem. Most of the headway reported in this dissertation has been in the shape of psychophysical experiments, which, inspired by the existing models of speed perception, provide more data regarding the actual process of human speed perception. It is now time for the models to provide quantitative accounts of how these data come to bear, something they are far from being able to do at this stage. None of the published models account for some of the non-veridical speed perception data, such as the contrast effect (Thompson, 1982) and the duration effect (Giaschi & Anstis,

178

1989). Neither do they account for any of the data on speed discrimination, since they are almost all noise-free models. The model presented in the last chapter provides the beginning of a framework for such study, but clearly many issues remain unsolved. It will be interesting to see whether a single model can provide a concise and compelling explanation for the varied types of effects which have been reported, such as those reported in this dissertation.

References

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

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

Anderson, S., & Burr, D. (1985). Spatial and temporal selectivity of the human motion detection system. *Vision Research, 25*(8), 1147–1154.

Anderson, S. J., & Burr, D. C. (1987). Receptive field size of human motion detection units. *Vision Research, 27*(4), 621–635.

Ascher, D., Welch, L., & Festa, E. K. (1996). Adaptation to stationary gratings results in an increase in apparent speed of moving gratings. In *Investigative ophthalmology and visual science (suppl.)* (Vol. 37, p. S916). Fort Lauderdale, FL.

Ascher, D., Welch, L., & Grzywacz, N. M. (1997). Integration across spatial frequency channels in speed discrimination. In *Society for neuroscience, 26th annual meeting* (p. #349.6). Fort Lauderdale, FL.

Barlow, H. B., & Levick, R. W. (1965). The mechanism of directional selectivity in the rabbit's retina. *Journal of Physiology, 173*, 477–504.

Blackwell, H. (1946). Contrast thresholds of the human eye. *Journal of the Optical Society of America, 36*, 624–643.

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

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

Bowne, S. (1990). Evidence for speed-tuned motion units. In *13th european conference on visual perception*. Paris, France.

Bowne, S. F. (1990). Contrast discrimination cannot explain spatial frequency, orientation or temporal frequency discrimination. *Vision Research, 30*(3), 449–461.

Bowne, S. F., McKee, S. P., & Levi, D. M. (Personal Communication). *Speed-tuned mechanims are required for speed discrimination.*

Burbeck, C. A., & Kelly, D. H. (1981). Contrast gain measurements and the transient/sustained dichotomy. *Journal of the Optical Society of America, 71*, 1335–1350.

Burr, D., Ross, J., & Morrone, M. (1986). Seeing objects in motion. *Proceedings of the Royal Society, London, B, 227*, 249–265.

Burt, P. (1983). Fast algorithms for estimating local image properties. *Computer Vision Graphics Image Processing, 21*, 368–382.

Campbell, & Robson. (1968). Application of fourier analysis to the visibility fo gratings. *Journal of Physiology (London), 197*(3), 551–566.

Carandini, M., & Heeger, D. (1994). Summation and division by neurons in primate visual cortex. *Science, 264*, 1333-1336.

Carandini, M., Heeger, D., & Movshon, J. (1996). Cerebral cortex. In E. Jones & P. Ulinski (Eds.), (Vol. XII. Cortical Models). New York: Plenum Press.

Carlson, V. R. (1962). Adaptation in the perception of visual velocity. *Journal of Experimental Psychology, 64*, 192–197.

Clymer, A. B. (1973). *The effect of seen motion on the apparent speed of subsequent test velocities: speed tuning of movement.* Unpublished doctoral dissertation, Columbia University, NY.

Dao, J. (1994). Contrast sensitivity measures and accuracy of image stabilization systems. *J Neuroophthalmol, 14* (4), 234–249.

Das, A., & Gilbert, C. D. (1995). Receptive field expansion in adult visual cortex is linked to dynamic changes in strength of cortical connections. *Journal of Neurophysiology, To appear.*

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

DeAngelis, G., Ohzawa, I., & Freeman, R. (1992). The organization of suppression in receptive fields of neurons in the cat's visual cortex. *Journal of Neurophysiology, 68*, 144–163.

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

DeValois, K. (1977). Spatial frequency adaptation can enhance contrast sensitivity. *Vision Research, 17*, 1057–1065.

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

Fennema, C. L., & Thompson, W. B. (1979). Velocity determination in scenes containing several moving objects. *Computer Graphics and Image Processing*, *9*, 301–315.

Fermi, G., & Reichardt, W. (1963). Optomotorische reaktionen der fliege. *Musca Domestica L. Kibernetik.*, *2*, 15–28.

Foley, J. M., & Legge, G. E. (1981). Contrast discrimination and near-threshold discrimination in human vision. *Vision Research*, *21*, 1041–1053.

Foster, K. H., Gaska, J. P., Nagler, M., & Pollen, D. A. (1985). Spatial and temporal frequency selectivity of neurones in visual cortical areas v1 and v2 of the macaque monkey. *Journal of Physiology*, *365*, 331–363.

Gabor, D. (1946). Theory of communication. *Journal of the Institute of Electrical Engineering*, *93*(429), 429–457.

Geesaman, B., & Qian, N. (1996). A novel speed illusion involving expansion and rotation patterns. *Vision Research*, *36*, 3281–3292.

Geesaman, B., & Qian, N. (in press). The effect of complex motion pattern on speed perception. *Vision Research*.

Giaschi, D., & Anstis, S. (1989). The less you see it, the faster it moves: shortening the "on-time" speeds up apparent motion. *Vision Research*, *29*(3), 335–347.

Gibson, J. J. (1937). Adaptationwith negative aftereffect. *Psychological Review*, *44*, 222–244.

Gilbert, C. D. (1993). Rapid dynamic changes in adult cerebral cortex. *Current Opinion in Neurobiology, 3*, 100-103.

Gilbert, C. D., & Wiesel, T. N. (1990). The influence of contextual stimuli on the orientation selectivity of cells in primary visual cortex of the cat. *Vision Research, 30*, 1689–1701.

Götz, K. G. (1965). Die optischen übertragungseigenschaften de komplexaugen von drosophila. *Kybernetik, 2*, 215–221.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* Hungtington, NY: Robert E. Kriegler Publishing Company.

Grzywacz, N. M., Watamaniuk, N., & McKee, S. P. (1995). Temporal coherence theory for the detection and measurement of visual motion. *Vision Research, 35*, 3183–3203.

Grzywacz, N. M., & Yuille, A. L. (1990). A model for the estimate of local image velocity by cells in the visual cortex. *Proceedings of the Royal Society of London Series B, 239*, 129–161.

Heeger, D. (1987). Model for the extraction of image flow. *Journal of the Optical Society of America, 4*(8), 1455–1471.

Heeger, D. (1993). Modeling simple cell direction selectivity with normalized, half-squared, linear operators. *Journal of Neurophysiology, 70*, 1885–1898.

Heeger, D. J. (1991). Computational models of visual prcessing. In M. Landy & J. A. Movshon (Eds.), (pp. 119–133). MIT Press, Cambridge, MA.

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

Heeger, D. J. (1992b). Normalization of cell responses in cat striate cortex. *Visual Neuroscience, 9*, 181–198.

Heeger, D. J., Simoncelli, E. P., & Movshon, J. A. (1996). Computational models of cortical visual processing. *Proceedings of the National Academy of Sciences, USA, 93*, 623–627.

Hildreth, E. C., & Koch, C. (1987). The analysis of visual motion: From computational theory to neuronal mechanisms. *Annual Review of Neuroscience, 10*, 477–533.

Hoekstra, J., Goot, D. P. J. van der, Brink, G. van der, & Bilsen, F. A. (1974). The influence of the number of cycles upon the visual contrast threshold for spatial sine patterns. *Vision Research, 14*, 365–368.

Holub, R. A., & Morton-Gibson, M. (1981). Response of visual cortical neurons of the cat to moving sinusoidal gratings: Response-contrast functions and spatiotemporal interactions. *Journal of Neurophysiology, 46*, 1244–1359.

Horn, B. K. P., & Schunck, B. G. (1981). Determining optical flow. *Artificial Intelligence, 17*, 185–203.

Howell, E., & Hess, R. (1978). The functional area for summation for sinusoidal gratings. *Vision Research, 18*, 369–374.

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

Hubel, D., & Wiesel, T. (1962). Receptive fields, binocular interaction and functional architecture. *Journal of Physiology, 160*, 106–154.

185

Hubel, D., & Wiesel, T. (1974). Sequence regularity and geometry of orientation columns in the monkey striate cortex. *Journal of Comparative Neurology*, *158*(158), 267–293.

Johnston, A., & Clifford, C. (1995). A unified account of three apparent motion illusions. *Vision Research*, *35*(8), 1109–1123.

Johnston, A., & Clifford, C. W. G. (1995). Perceived motion of contrast-modulated gratings: predictions of the multi-channel gradient model and the role of full-wave rectification. *Vision Research*, *35*(12), 1771–1783.

Johnston, A., McOwan, P. W., & Buxton, H. (1992). A computational model of the analysis of some first-order and second-order motion patterns by simple and complex cells. *Proceedings of the Royal Society B*, *250*, 297–306.

Kelly, D. H. (1979). Motion and vision. i. stabilized images of stationary gratings. *Journal of the Optical Society of America*, *69*, 1266–1274.

Klein, S. A., & Levi, D. M. (1985). Hyperacuity thresholds of 1 sec: Theoretical predictions and empirical validations. *Journal of the Optical Society of America A*, *2*, 1170–1190.

Koch, C., Wang, H., & Mathur, B. (1989). Computing motion in the primate's visual system. *Journal of Experimental Biology*, *146*, 115–139.

Kontsevich, L. L. (1995). The nature of the inputs to cortical motion detectors. *Vision Research*, *35*(19), 2785–2793.

Kulikowski, J. J., & Tolhurst, D. J. (1973). Psychophysical evidence for sustained and transient detectors in human vision. *Journal of Physiology*, *232*, 149–162.

Legge, G. (1981). A power law for contrast discrimination. *Vision Research*, *21*, 457–467.

Legge, G. E., & Foley, J. M. (1980). Contrask masking in human vision. *JOSA*, *70*(12), 1458–1471.

Lehky, S. R. (1985). Temporal properties of visual channels measured by masking. *Journal of the Optical Society of America A*, *2*(8), 1260–1272.

Levinson, E., & Sekuler, R. (1975). The independence of channels in human vision selective for direction of movement. *Journal of Physiology*, *250*, 347–366.

Limb, J., & Murphy, J. (1975). Estimating the velocity of moving images in television signals. *Computer Graphics and Image Processing*, *4*, 311–327.

Mandler, M. B. (1984). Temporal frequency discrimination above threshold. *Vision Research*, *24*(12), 1873–1880.

Mandler, M. B., & Makous, W. (1984). A three channel model of temporal frequency perception. *Vision Research*, *24*(12), 1881–1887.

Marr, D., & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London, B*, *207*, 187–217.

Marr, D., & Ullman, S. (1981). Directional selectivity and its use in early visual processing. *Proceedings of the Royal Society of London Ser. B*, *211*, 151–180.

McCann, J. J., Savoy, R. L., & Hall, J. A. (1973). Visibility of low spatial frequency sine-wave targets: Dependence on number of cycles. *Journal of the Optical Society of America*, *63*, 1297.

McKee, S., & Welch, L. (1985). Sequential recruitment and the discrimination of velocity. *Journal of the Optical Society of America*, *2*(2), 243–251.

187

McKee, S. P., Silverman, G. H., & Nakayama, K. (1986). Precise velocity discrimination despite random variations in temporal frequency and contrast. *Vision Research*, *26*(4), 609–619.

McKee, S. P., & Welch, L. (1989). Is there a constancy for velocity? *Vision Research*, *29*(5), 553–561.

Metha, A. B., Bex, P. J., & Makous, W. (1997). Different apparent speeds for optic flow. *Investigative Ophtalmology and Visual Science*, *38*(4), S1167.

Mostafavi, H., & Sakrison, D. J. (1976). Structure and properties of a single channel in the human visual system. *Vision Research*, *1*, 957–968.

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

Nachmias, J., & Sansbury, V. R. (1974). Discrimination may be better than detection. *Vision Research*, *14*, 1039–1042.

Naka, K., & Rushton, W. (1966). S-potentials from colour units in the retina of fish (cyprinidae). *Journal of Physiology*, *185*, 536–555.

Nestares, O., & Heeger, D. J. (in press). Modeling the apparent frequency-specific suppression in simple cell responses. *Vision Research*.

Newsome, W. T., Gizzi, M. S., & Movshon, J. A. (1983). Spatial and temporal properties of neurons in macaque mt. *Investigative Ophtalmology and Visual Science (ARVO)*, *24*, 106.

Olzak, L. A., & Thomas, J. P. (1986). Handbook of perception and human performance. In (Vol. I). John Wiley and Sons.

188

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

Pantle, A. J. (1978). Temporal frequency response characteristic of motion channels measured with three different psychophysical techniques. *Perception & Psychophysics, 24* (3), 285–294.

Pelli, D. (1981). *Effects of visual noise.* Unpublished doctoral dissertation, Cambridge University, England.

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

Poston, T., & Stewart, I. (1978). *Catastrophe theory and its applications.* London: Pitman.

Quinn, P. C., & Lehmkuhle, S. (1983). An oblique effect of spatial summation. *Vision Research, 23,* 655–658.

Regan, D., & Beverley, K. (1983). Spatial-frequency discrimination and detection: comparison of postadaptation thresholds. *Journal of the Optical Society of America, 12,* 1684–1690.

Regan, D., & Beverley, K. (1985). Postadaptation orientation discrimination. *Journal of the Optical Society of America A, 2,* 147–155.

Reichardt, W. (1961). Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. In *Sensory communication.* New York: Wiley.

Reid, R. C., Soodak, R. E., & Shapley, R. M. (1991). Directional selectivity and spatiotemporal structure of receptive fields of simple cells in cat striate cortex. *Journal of Neurophysiology*(2), 505–529.

Reisbeck, T. E., & Gegenfurtner, K. R. (1997). Velocity tuned mechanisms in human motion perception. In *Investigative ophthalmology and visual science* (Vol. 38, p. S376). Fort Lauderdale, FL.

Robson, J. (1966). Spatial and temporal contrast sensitivity functions of the visual system. *Journal of the Optical Society of America, 56*, 1141–1142.

Robson, J. (1988). Linear and nonlinear operations in the visual system. *Investigative Ophtalmology and Visual Science Supplement, 29*, 117.

Robson, J., & Graham, N. (1981). Probability summation and regional variation in contrast sensitivity across the visual field. *Vision Research, 21*, 409–418.

Schrater, P., & Simoncelli, E. (1995). Biases in speed perception due to motion aftereffect. In *Investigative opthalmology and visual science supplement (arvo)* (Vol. 36). Fort Lauderdale, FL.

Schrater, P. R., Knill, D. C., & Simoncelli, E. P. (1997). Local translation detection: Evidence for velocity tuned pooling of spatio-temporal frequencies. In *Investigative opthalmology and visual science supplement (arvo)* (Vol. 38). Fort Lauderdale, FL.

Scott, T. R., Jordan, A. E., & Powell, D. A. (1963). Does the visual after-effect add algebraically to objective motion of the test stimulus? *Journal of Experimental Psychology, 66*.

Seiffert, A. E., & Cavanagh, P. (1997). Position displacement, not velocity, is the

cue to motion detection of texture-defined patterns. *Investigative Ophtalmology and Visual Science (ARVO)*, *38*(4), S377. (1775)

Simoncelli, E. P., & Heeger, D. J. (1994). A velocity-representation model for MT cells. In *Investigative opthalmology and visual science supplement (arvo)* (Vol. 35, p. 1827). Fort Lauderdale, FL.

Simoncelli, E. P., & Heeger, D. J. (1997). A model of neural responses in visual area MT. *Vision Research*. (Accepted for publication)

Simoncelli, E. P., Heeger, D. J., & Adelson, E. H. (1992). A computational model for perception of two-dimensional pattern velocities. In *Investigative opthalmology and visual science supplement (arvo)* (Vol. 33). Sarasota, FL.

Smith, A. T. (1987). Velocity perception and discrimination: relation to temporal mechanisms. *Vision Research*, *27*(9), 1491–1500.

Smith, A. T., & Edgar, G. K. (1990). The influence of spatial frequency on perceived temporal frequency and perceived speed. *Vision Research*, *30*(10), 1467–1474.

Smith, A. T., & Edgar, G. K. (1991). The separability of temporal frequency and velocity. *Vision Research*, *31*(2), 321–326.

Smith, A. T., & Edgar, G. K. (1994). Antagonistic comparison of temporal frequency filter outputs as a basis for speed perception. *Vision Research*, *34*(2), 253–265.

Smith, R. A. (1970). Adaptation of visual contrast sensitivity to specific temporal frequencies. *Vision Research*, *10*, 275–279.

Smith, R. A. (1971). Studies of temporal frequency adaptation in visual contrast sensitivity. *Journal of Physiology (London)*, *216*, 531–552.
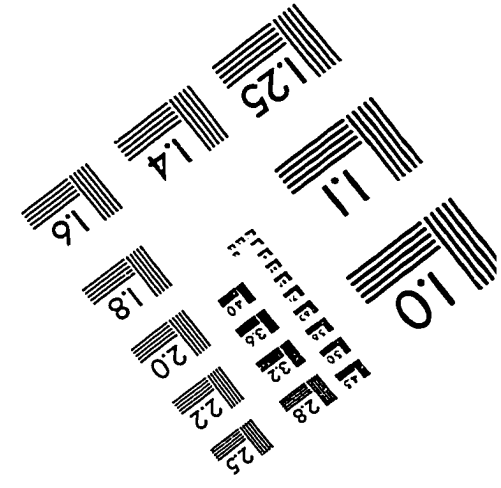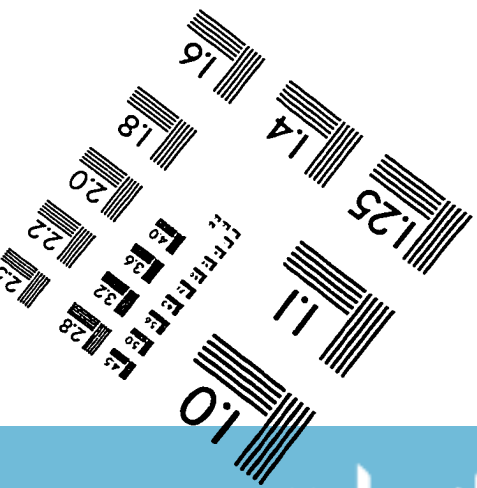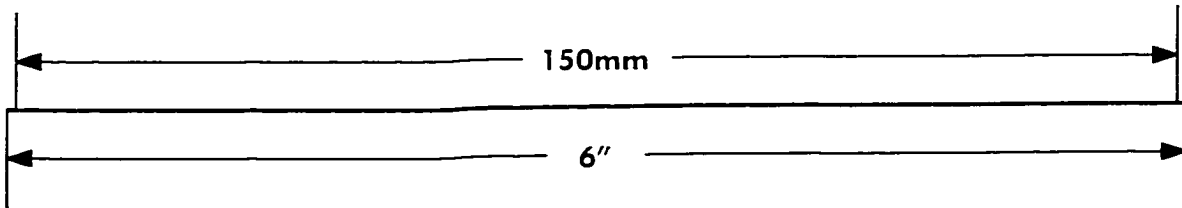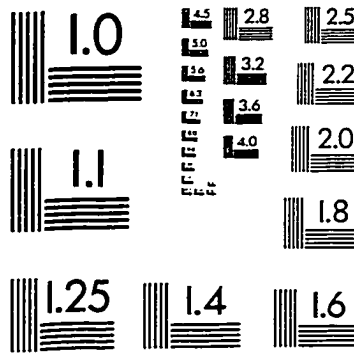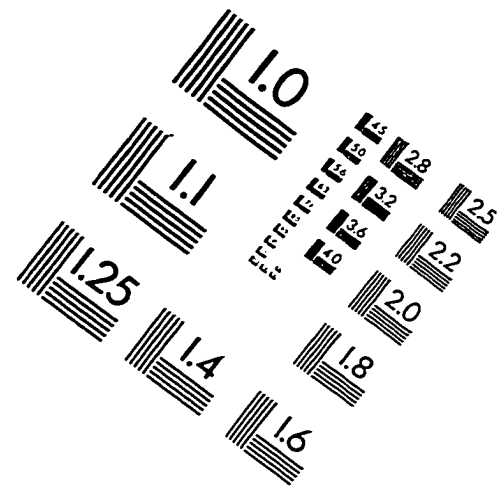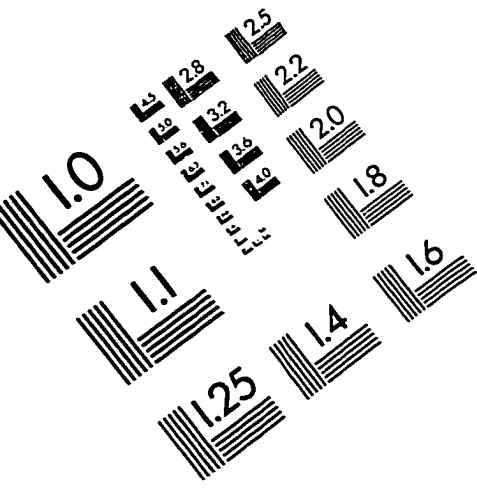
Snowden, R. (1997). Perceived speed: effects of field size and background texture. *Investigative Ophtalmology and Visual Science, 38*(4), S1167.

Stone, L. S., & Thompson, P. (1992). Human speed perception is contrast dependent. *Vision Research, 32*(8), 1535–1549.

Stromeyer, C. F., & Julesz, B. (1972). Spatial-frequency masking ini vision: critical bands and spread of masking. *Journal of the Optical Society of America, 62,* 1221-1232.

Thomas, J. P., & Gille, J. (1979). Bandwidths of orientation channels in human vision. *Journal of the Optical Society of America, 69,* 652–660.

Thompson, P. (1976). *Velocity after-effects and the perception of movement.* Unpublished doctoral dissertation, University of Cambridge, England.

Thompson, P. (1981). Velocity after-effects: the effects of adaptation to moving stimuli on the perception of subsequently seen moving stimuli. *Vision Research, 21,* 337–345.

Thompson, P. (1982). Perceived rate of movement depends on contrast. *Vision Research, 22,* 377–380.

Tolhurst, D. J., & Heeger, D. J. (in pressa). Comparison of contrast normalization and hard threshold models of the responses of simple cells in cat striate cortex. *Visual Neuroscience.*

Tolhurst, D. J., & Heeger, D. J. (in pressb). Contrast normalization and a linear model for the directional selectivity of simple cells in cat striate cortex. *Visual Neuroscience.*

Tootle, J., & Berkley, M. (1983). Contrast sensitivy for vertically and obliquely oriented gratings as a function of grating area. *Vision Research, 23,* 907-910.

Tulunay-Keesey, U., & Jones, R. M. (1980). Contrast sensitivity measures and accuracy of image stabilization systems. *Journal of the Optical Society of America, 70*, 481–488.

Turano, K., & Pantle, A. (1985). Discontinuity limits for the generation of visual motion aftereffects with sine- and square-wave gratings. *Journal of the Optical Society of America, 2*(2), 260–266.

Turano, K., & Pantle, A. (1989). On the mechanism that encodes the movement of contrast variations: velocity discrimination. *Vision Research, 29*, 207–221.

Van Santen, J. P. H., & Sperling, G. (1984). A temporal covariance model of motion perception. *Journal of the Optical Society of America A, 1*, 451–73.

Van Santen, J. P. H., & Sperling, G. (1985). Elaborated reichardt detectors. *Journal of the Optical Society of America A, 2*, 300–20.

vanEssen, D. C., Newsome, W. T., & Maunsell, J. H. (1984). The visual field representation in striate cortex of the macaque monkey: asymmetries, anisotropies, and individual variability. *Vision Research, 24*(5), 429–448.

Watamaniuk, S. N., McKee, S. P., & Grzywacz, N. M. (1994). Detecting a trajectory embedded in random-direction motion noise. *Vision Research, 35*, 65–77.

Watamaniuk, S. N., & Sekuler, R. (1992). Temporal and spatial integration in dynamic random dot stimuli. *Vision Research, 32*, 2341–2347.

Watson, A., & Robson, J. (1981). Discrimination at threshold: Labelled detectors in human vision. *Vision Research, 21*, 1115–1122.

Watson, A. B. (1986). Handbook of perception and human performance. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), (Vol. I). John Wiley and Sons.

193

Watson, A. B., & Ahumada, A. J. (1985). Model of human visual-motion sensing. *Journal of the Optical Society of America A*, *2*, 322–341.

Watson, A. B., Ahumada, A. J., & Farrell, J. (1986). Window of visibility: psychophyiscal theory of fidelity in time-sampled visual motion displays. *Journal of the Optical Society of America A*, *3*(3), 300–307.

Watson, A. B., & Pelli, D. G. (1983). Quest: a bayesian adaptive psychometric method. *Perception and Psychophysics*, *33*, 113–120.

Watt, R., & Morgan, M. (1984). Spatial filters and the localization of luminance changes in human vision. *Vision Research*, *24*, 1387–1397.

Werkhoven, P., Snippe, H. P., & Toet, A. (1992). Visual processing of optic acceleration. *Vision Research*, *32*(12), 2313–2329.

Wilson, H. (1986). Responses of spatial mechanisms can explain hyperacuity. *Vision Research*, *26*, 453–469.

Wilson, H., & Gelb, D. (1984). Modified line-element theory for spatial-frequency and width discrimination. *Journal of the Optical Society of America A*, *1*(1), 124–131.

Wilson, H., McFarlane, D., & Phillips, G. (1983). Spatial frequency tuning of orientation selective units estimated by oblique masking. *Vision Research*, *23*(9), 873–882.

Wohlgemuth, A. (1911). On the after-effect of seen movement. *British Journal of Psychology, Monograph Supplement*, *1*, 1–117.

Zanker, J. (1990). Theta motion: A new psychophysical paradigm indicating two levels of visual motion perception. *Naturwissenschaften*, *77*, 243–246.

Zanker, J. (1993). Theta motion: A paradoxical stimulus to explore higher order motion extraction. *Vision Research, 33*, 553–569.

Zeki, S. (1969). Representation of central visual fields in prestriate cortex of monkey. *Brain Research, 14*, 271–291.

Zipser, K., Lamme, V., & Schiller, P. (1996). Contextual modulation in primary visual cortex. *The Journal of Neursocience, 16*(22), 7376–7389.

IMAGE EVALUATION
TEST TARGET (QA-3)

150mm

6"