LOCAL MOTION DETECTION: COMPARISON OF HUMAN AND MODEL OBSERVERS

Paul Schrater

A DISSERTATION

in

Neuroscience

Presented to the Faculties of the University of Pennsylvania in Partial Fulfillment

of the Requirements for the Degree of Doctor of Philosophy

1998

David Knill, Ph.D.

Supervisor of Dissertation

Michael Nusbaum, Ph.D.

Graduate Group Chairperson

ABSTRACT

LOCAL MOTION DETECTION: COMPARISON OF HUMAN AND MODEL OBSERVERS

Paul Schrater

David Knill

We investigated the mechanisms of visual motion processing in humans. Previous research had shown the human visual system decomposes moving images by a set of spatio-temporal frequency selective mechanisms which do not unambiguously encode the velocity of moving patterns. By pooling the outputs of these mechanisms which have frequency selectivity lying on a common plane in spatio-temporal frequency space, the visual system can encode pattern velocities. The goal of this research was to investigate whether the visual system uses such pooling.

We designed a novel set of stimuli which are optimal for pattern motion detectors. By comparing detection performance on these stimuli against a set of control stimuli, we found evidence for pooling. A trial by trial perturbation analysis of this data was used to determine the observers' pooling strategies, which showed at least two distinct kinds of frequency weighting, narrow band in orientation and broad band with weights restricted to a common plane. Finally, using subthreshold summation experiments we found that fourier power lying a common plane is additively pooled while power not a common plane is subadditively pooled.

The thesis provides the first unambiguous psychophysical evidence for a set of visual mechanisms specialized to encode pattern velocity.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In this chapter we frame the general problem of computing image motion biologically and motivate the experiments.

1.1 Image Motion

One of the most striking aspects of visual experience is the ease with which we make sense of motion in the world. The apparent ease with which motion is perceived can obscure the fact that the visual system is faced with a formidable set of problems in processing visual motion. At the physical level, movement in the world causes changes in the distribution of light across the retina. Somehow the visual system extracts information about the environment from the pattern of these changes. In order to study how this occurs, we need a description of the information available at the retina, for which we will make some simplifications. Ignoring wavelength information and adopting a ray model for light, the light distribution on the retina can be described by specifying the intensity at each point (x, y) on the retina at each moment of time, t: $I(x, y, t)$.

We refer to the instantaneous light distribution on the retina $I(x, y)$ as the retinal image, which is formed by the projection of light from points in the world onto the retina. When objects in the world move relative to the observer, the projected points move within the retinal image. The movement of the projected points on the retina is referred to as the *motion field*. Several authors have shown that by having access to the motion field many properties of the world can be estimated, including relative depth, surface shapes, and object motions [71, 70, 101, 68].

The motion field is not directly available to the observer. What is available to the observer are the changes in intensity at each point in the image. However, if the intensity of a projected point remains constant as the point moves, then the change in the image intensities can be used to compute the motion field. One of the simplest ways to measure the positional change of an image point is to approximate the path by a set of local translations. If a projected point (x, y) at time t_1 moves to (x', y') at time t_2, and the intensity of that point remains constant for the duration of the movement, then the result of the movement is to transfer the image region from one region of the retina to another region $I(x', y') = I(x(t_1 + \Delta t), y(t_1 + \Delta t))$. For small time durations Δt, the change in position can be approximated by a translation, so that the image at time t_2, $I(x(t_2), y(t_2))$ is given by:

$$I\left(x(t_2), y(t_2)\right) = I(x(t_1) - v_x t, y(t_1) - v_y t) \tag{1.1}$$

where v_x and v_y are functions of image position and time, and $t = t_2 - t_1$. The approximation of image changes by estimates of local image translations at each retinal location is commonly called the *optic flow*

field. Use of the optic flow approximation by the visual system can be motivated by the simple relationship between optic flow and the movement of points in the world which holds when image intensities remain constant, and by the fact that local translations are the simplest approximation of motion induced image changes.

However, care must be taken when interpreting optic flow in terms the motions of points in the world, since in computing optic flow we assume that the intensity image of objects remain fairly constant as they move. Several authors have shown that this assumption is frequently violated in real scenes [130, 99]. The problem is that the luminance profile of an object can change quite substantially for reasons other than the object's motion, the most important of these being: changes in reflectance due to changes in object pose, changes in the illumination condition, object occlusion, and shadowing. Thus under many circumstances the motion field can only be accurately estimated by combining the optic flow measurements with estimates of other scene attributes (e.g. light source direction, object reflectances, etc.).

Local image translations are also useful outside the context of inferring the motion of points in the world. Another distinct motivation for their use by the visual system stems from their utility as an efficient code for time-varying images. This motivation follows from the hypothesis that the goal of the early visual system is to efficiently encode the time varying image [16, 137, 37]. In this view, what determines early motion processing is the statistical structure of the received luminance signal and not in the relationship between the signal and motions in the world. The encoding of the time varying image amounts to finding an approximation to $I(x, y, t)$ which captures most of the significant behavior of the function (i.e. allows approximate inversion of the code) while eliminating redundancies in the code. Interestingly, the consideration of efficient coding within the context of eye movements[1] has led to the idea that the visual system should make local translation measurements[37]. Eckert et al.[37] show that the time varying image can be decomposed into an optic flow component and a 'stationary' component[2], and that the optic flow representation efficiently encodes the redundancies in the time varying image introduced by eye movements.

Thus, we see that the measurement of local translations of the retinal image can be motivated by two different assumptions of the goal of early motion processing. These theoretical motivations are corroborated by the abundant evidence that local image translations are sufficient for the performance of large set of perceptual tasks. Observers can be forced to rely exclusively on motion information by using stimuli composed of moving dots which are randomly spatially distributed in which each individual dot has a limited duration of exposure. Using these stimuli, investigators have shown that observers can use local motion information to estimate the observer's direction of heading [133], object geometry [59, 75, 35], depth relationships [108], and define boundaries between objects [19, 108], to name only a few. In addition, McKee and others [82, 20] have argued that the visual system explicitly represents image velocities from the fact that observers are highly sensitive to changes in speed and direction of moving images. Finally, Schrater & Simoncelli have found evidence for the representation of image velocities by the visual system in a set of adaptation experiments [111].

In addition to psychophysical evidence, electrophysiological recording in monkey visual area MT have revealed neurons with many of the properties expected of mechanisms which measure local image translations [80, 8, 10, 90].

All of these finding suggest that the visual system may make measurements of local image translations.

[1] particularly tracking eye movements

[2] the stationary component represents all the changes in luminance not accounted for by local image translations

1.2 Measuring Local Translations

In the previous section the measurement of local image translations by the visual system were motivated and it was suggested that the visual system may make such measurements. In this section we address how these measurements might be made. A large number of measurement schemes have been published (see Simoncelli [112], Nakayama [92], or Koenderink [125]) for reviews), however, most proposed measurement methods can be classified as belonging to one of three basic types: [125, 112]

• **Matching methods** Matching methods estimate image translations by attempting to match local image regions or characteristic image features at subsequent instants of time [100, 127, 49, 24, 27, 88, 89].

• **Gradient methods** Gradient methods use derivatives of image intensity over space and time to form estimates [61, 79, 55, 146].

• **Frequency-based or Filter-based methods** Frequency-based methods is an umbrella term for a collection of methods developed by considering the problem in the spatio-temporal frequency domain. These methods fall into two different categories - amplitude based [56, 53, 112] and phase based [46].

Simoncelli [112] and others have shown that there are numerous connections between these various methods and that all of these methods can be described in a common framework based on spatio-temporal filtering. In the remainder of the thesis we will focus on frequency based methods for two reasons: 1) The description of image translations in the frequency domain is fundamental and simple. The reason that so many different estimation methods can be put into a common framework is that they each make use of the structure of the signal, which is evident in the frequency domain. 2) The evidence for the existence of mechanisms which act as frequency selective filters in the visual system is quite substantial (see Graham, 1989 for a review). Thus a translation estimation method which uses spatio-temporal filtering can be naturally implemented in the visual system. The next section describes the structure of translating signals in the frequency domain.

1.2.1 Image translations in the frequency domain

Translating images have a simple structure in the spatio-temporal frequency domain which is illustrated in figure 1.1 for the temporal and one spatial dimensions. At the top of the figure a random texture is illustrated, with the black arrow indicating the direction of the translation. If we stack up the images of the texture at a set of discrete time points, we will have a 'movie' of the translation. The x-t slice of this movie is shown in the lower left side of the figure. Notice that the slice has the appearance of a oriented pattern. The orientation θ_{x-t} is the result of the shearing transformation caused by translation, and it is directly related to the speed: $\theta_{x-t} = \tan^{-1}(1/v)$. Because of the dominant orientation, the Fourier decomposition of the x-t slice consists of the set of sinusoids with the same orientation in x-t, differing only in phase and frequency magnitude[3]. Thus, the Fourier transform of the x-t signal has an amplitude spectrum constrained to lie on a line through the origin whose slope represents the orientation of the pattern, shown in the bottom right hand side of figure 1.1.

A more intuitive way to understand the Fourier representation is to consider a one dimensional intensity signal along x which is drifting with uniform speed. We can decompose the signal into its set of component sinusoids each of which drifts at the same speed. Having the same speed means that each undergoes the same spatial displacement in the same duration. For a speed $v = 1$ deg/sec, a sinusoid with a spatial

[3]Frequency magnitude refers to $(\omega_x^2 + \omega_t^2)^{\frac{1}{2}}$. It is the same concept as wave number and spatial frequency magnitude in the space domain.

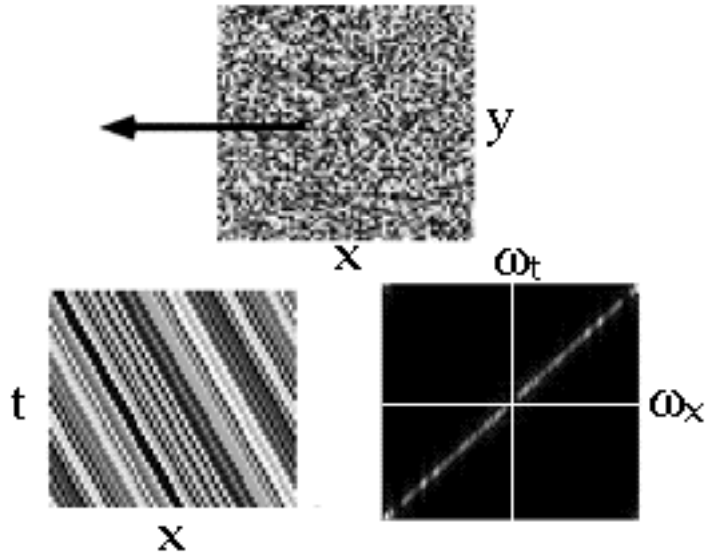

Figure 1.1: Fourier analysis of a translating image in 2-D. **a**, Random intensity image translating to the left (as indicated by the arrow). **b**, x-t slice of the image sequence shown in a. The pattern has a characteristic orientation which uniquely specifies the speed of the translation. **c**, Amplitude spectrum of the Fourier transform of the x-t slice. All the non-zero coefficients lie on a line through the origin. Deviations from the line are caused by the use of a finite size fourier transform.

frequency of 2 cyc/deg must move 2 cycles per sec, similarly 4 cyc/deg, 4 cycles per sec. In general $v = -\omega_t/\omega_x$, which is the equation of a line in (ω_x, ω_t) passing through the origin[4]: $\omega_t = -v\omega_x$, with slope $-v$. The angle of the line is $\theta_{\omega_x,\omega_t} = \tan^{-1}(v) = \pi - \theta_{x,t}$

Although reintroducing the neglected spatial dimension complicates matters, the description is a simple generalization of the x-t analysis. An image sequence can be represented as a block of data in (x,y,t) space. The data block representation is illustrated on the right hand side of figure 1.2, for a movie depicting a forest scene translating to the right. The front x-y face shows the image of the trees. The top x-t face shows the evolution over time of the luminance pattern at the top of the x-y image. If we take x-t slices at different y positions, a similar oriented pattern would be revealed. Thus, in three dimensions the signal can be described as a set of luminance fibers all with the same 3-D orientation. y-t slices of the data cube for different x positions are all broad band, illustrated by the visible y-t face of the cube.

The Fourier transform of the data cube is illustrated on the right hand side of the figure. In three dimensional spatio-temporal frequency space, translating signals are constrained to lie on a *plane* passing through the origin [138]: $\omega_t = -(v_x\omega_x + v_y\omega_y)$. This is the natural generalization of the line in 2-D and many of the properties of the 2-D analysis carry over. The orientation of the plane uniquely specifies the velocity of the translation, with the speed given by the angle between the velocity plane and the (ω_x, ω_y) plane.

An intuitive motivation for the result that the spectral power of a translating image lies on a plane can be given by considering the previous 2-D analysis and the properties of the x-t and y-t slices of the data cube. From the 2-D analysis, we know that an image translating in the x direction will have spectral power which lies on a line in (ω_x, ω_t). Since all of the x-t slices have the same oriented structure, the averaging of $(\omega_x, \omega_y, \omega_t)$ over ω_y results in a spectrum which lies on the same line. Thus the spectrum in $(\omega_x, \omega_y, \omega_t)$ must be either a plane or a line. Since the y-t slices are broadband, it must be a plane. Because this result is central to the thesis, a proof is sketched below:

[4]This can also be derived using dimensional analysis, speed = deg/sec, tf = cycles/sec, sf= cycles/deg.

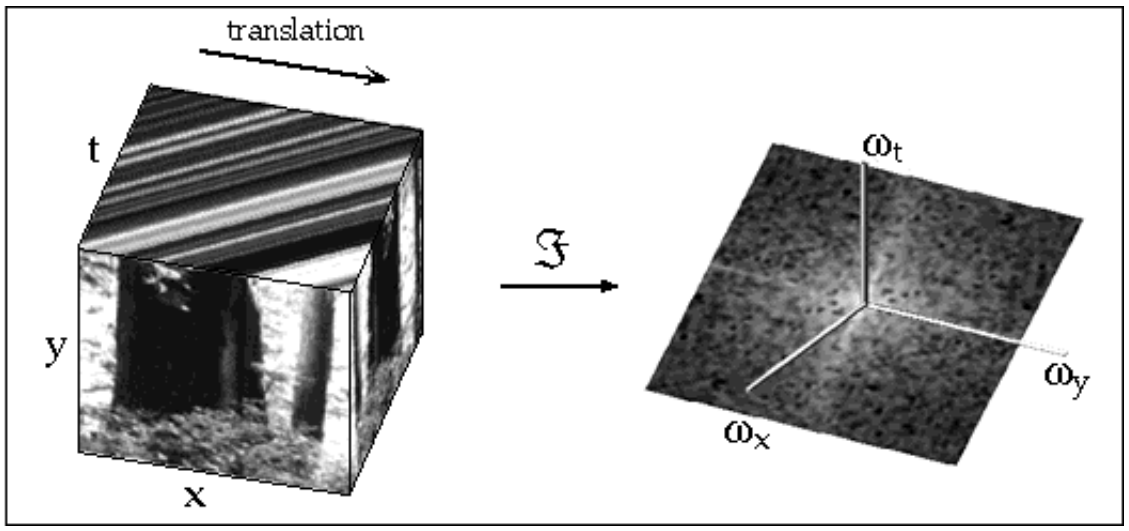

Figure 1.2: Illustration of the information in a translating image. a, Movie showing a forest scene translating to the left is represented as a data cube. This is a representation of the intensity information in the retinal image (the x-y plane) over time. The leftward motion can be inferred from oriented lines on the x-t face. b, The amplitude spectrum of the 3-D Fourier transform of the data cube is rendered as gray levels in $(\omega_x, \omega_y, \omega_t)$ space after the DC component is removed. For a translation, the non-zero Fourier amplitudes are constrained to lie on a plane through the origin. The slant and tilt of this plane uniquely specify the speed and direction of the translation.

For a globally translating image, $I(x, y, t)$ can be rewritten as $I(x - v_x t, y - v_y t)$. This signal can be written as the sum of sinusoids which have a constant spatial spectrum, with the translation causing a phase shift in each sinusoid:

$$I(x(t), y(t)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A(\omega_x, \omega_y) \exp(2\pi i(\omega_x(x - v_x t) + \omega_y(y - v_y t))d\omega_x d\omega_y \tag{1.2}$$
$$I = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A(\omega_x, \omega_y) \exp(2\pi i(\omega_x x + \omega_y y)) \exp(-2\pi i(\omega_x v_x t + \omega_y v_y t))d\omega_x d\omega_y$$

The 3D Fourier Transform of this expression is:

$$\begin{aligned}\mathcal{F}\{I\} &= \int_{x,y,t} \exp(-2\pi i(\omega_x x + \omega_y y + \omega_t t)) \\ &\quad \left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A(\omega_x, \omega_y) \exp(-2\pi i(\omega_x(x - v_x t) + \omega_y(y - v_y t)))d\omega_x d\omega_y\right) dx\, dy\, dt\end{aligned} \tag{1.3}$$

The spatial sinusoids in the transform and the signal multiply to one, and switching order of integration using Fubini's theorem:

$$\begin{aligned}\mathcal{F}\{I\} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A(\omega_x, \omega_y) \\ &\quad \left(\int_t \exp(-2\pi i(\omega_t + \omega_x v_x + \omega_y v_y)t)d_t\right) d\omega_x d\omega_y\end{aligned} \tag{1.4}$$

The inner integral evaluates to:

$$\mathcal{F}\{I\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A(\omega_x, \omega_y)\delta(\omega_t + \omega_x v_x + \omega_y v_y)d\omega_x d\omega_y \tag{1.5}$$

Recall that $\delta(0) = 1$, and that $\delta(x \neq 0) = 0$. Thus the transform is non zero for

$$\begin{aligned} \omega_t + \omega_x v_x + \omega_y v_y &= 0 \\ \vec{u} \cdot \vec{\omega} &= 0 \end{aligned} \tag{1.6}$$

which is an equation for a plane in frequency space $(\omega_x, \omega_y, \omega_t)$, with the normal vector to the plane given by $\frac{\vec{u}}{||\vec{u}||}$ where $\vec{u} = (v_x, v_y, 1)$. Equation 1.5 also shows that the amplitude spectrum on the plane is just the spectrum of the initial spatial image. What changes when a image translates is that the sinusoids at each spatial frequency drift with a temporal frequency consistent with the pattern motion. This is illustrated in figure 1.3.

Figure 1.3A illustrates a random texture translating to the left. The spectrum of this image sequence will be constrained to lie on a plane, illustrated by dashed lines in figure 1.3B. Pairs of points on the plane represent drifting gratings, and the translating texture movie can be constructed by summing up all of these gratings, weighted by their complex amplitudes. To illustrate the decomposition of the translating texture into its component gratings, six points lying on the plane were chosen, illustrated by pairs of gray balls. The spatial and temporal properties of the gratings can be determined by looking the projection of the frequency point onto the spatial frequency plane and onto the temporal frequency axis. The projection lines are shown for one of the frequency points in gray. The length of the projection onto the spatial frequency plane gives the spatial frequency magnitude and the spatial orientation θ of the projection. The projection onto the ω_t axis gives the temporal frequency, and the angle ϕ between the frequency vector and the spatial frequency plane is related to the speed at which the grating drifts in the direction of its orientation. The six gratings are depicted in figure 1.3D, with the lengths of the arrows representing the speed of the gratings in their normal directions.

An important point is that the frequency description of translations subsumes the 'intersection of constraints' rule for determining the direction of motion of a coherent pattern made up of several one dimensional components. Figure 1.3C illustrates the 'intersection of constraints' construction [2]. The figure represents the motion in three grating's normal directions as vectors in velocity space. Because each grating is a one dimensional signal, the motion in the direction parallel to the grating's orientation is ambiguous, a property often called the 'aperture problem'. Thus a grating can have any motion which preserves the speed orthogonal to the grating's orientation, shown as the three gray 'constraint' lines. The point of intersection of the gray lines is the velocity of the coherently moving pattern. This velocity is identical to the velocity indicated by the common plane which contains the set of frequency points. The importance of this fact will be clearer when we discuss translation detectors. Detectors which estimate velocity by pooling information across planes in frequency space automatically 'solve' the aperture problem because they are intrinsically estimating pattern motion[5].

Having determined the structure of a signal which is translating globally, it is easy to describe the structure of a local translation. We can model localization of a signal as multiplication by a smooth window in space and time $W(x, y, t)$. The localized signal spectra is just the spectrum lying on a plane convolved with the fourier transform of the smooth function W, since multiplication in the space domain results in convolution in the frequency domain. For functions which are sufficiently smooth, the effect will be to 'fuzz out' the plane. For instance if the window function is Gaussian, then the localized spectrum will be the non-zero points on the plane blurred by a Gaussian function. Thus for local translations, the plane becomes a 'pancake'.

[5]Except for signals which are intrinsically one dimensional, for which there is no well defined pattern motion. For instance, a grating is represented by a pair of points in frequency space which can be fit by the infinite family of planes which intersect these two points.

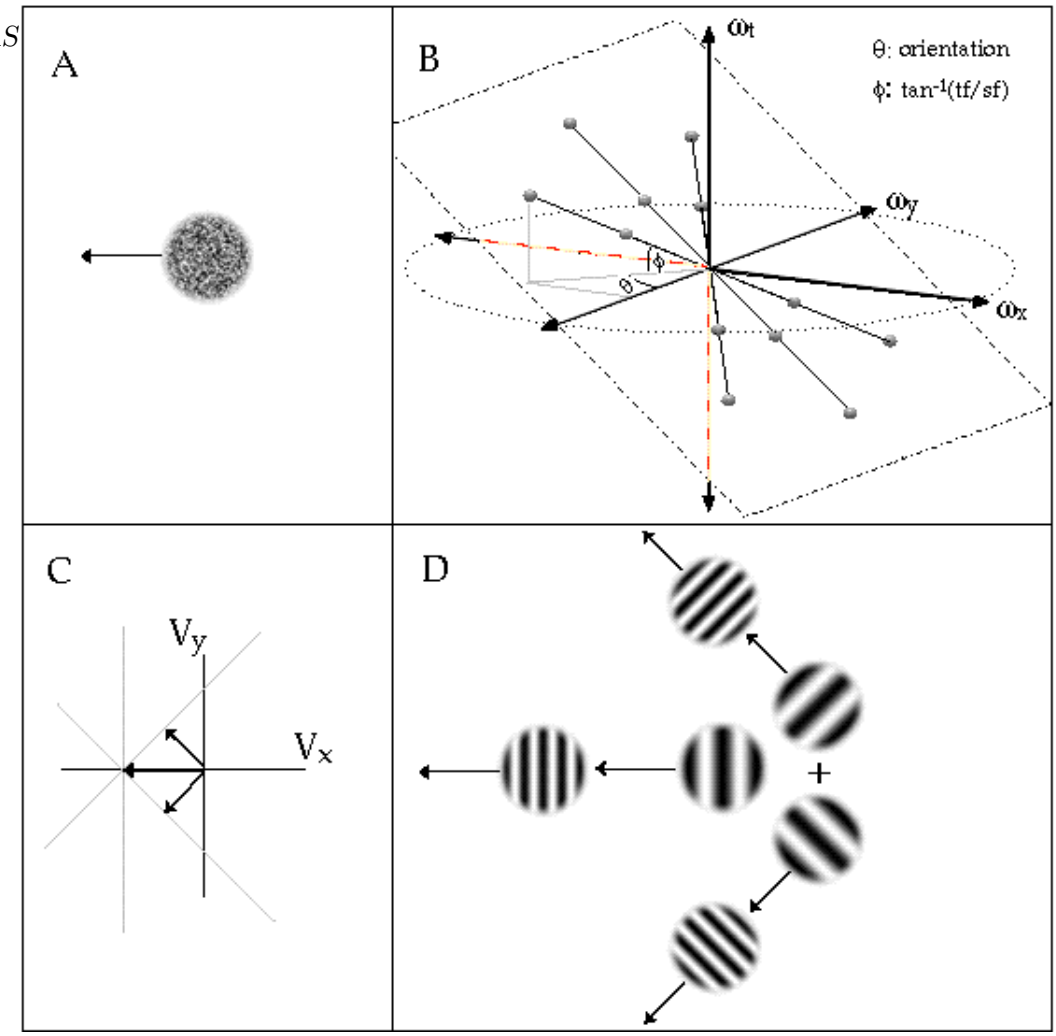

Figure 1.3: Illustration of properties of translating images in the frequency domain. **A**, Random intensity image translating to the left (as indicated by the arrow). **B**, Translating pattern in the frequency domain. Six frequency points on the plane are selected, represented by pairs of gray balls on opposite sides of the origin. The properties of the grating can be determined from the projections onto the frequency axes, shown in gray. θ determines the spatial orientation of the grating, and ϕ determines the speed in the direction orthogonal to the grating orientation. **C**, The points on the plane obey the 'intersection of constraints' rule. **D**, The six points from B shown as gratings. See text for details.

1.2.2 Models for translation detection

Having characterized the structure of image translations, we can discuss more readily the methods the visual system might use to estimate image velocities. When analyzed in the fourier domain, every successful method will estimate velocity by extracting information which lies on a common plane, but the structure of the estimator will vary depending on the task the visual system is optimized for. For instance, assume the visual system is optimized for the task of detecting a translating image whose spatial and temporal characteristics are exactly known buried in Gaussian white noise. Since the optimal detector for this task is the matched linear filter, then in this case we would expect the visual system to exhibit the properties of this detector.

What can we say about the task of estimating image velocities a priori? 1) Velocities can only be estimated in regions which have non-constant luminance profiles. 2) A generic image velocity estimator must be able to process a large class of different spatial patterns. 3) Under most circumstances the

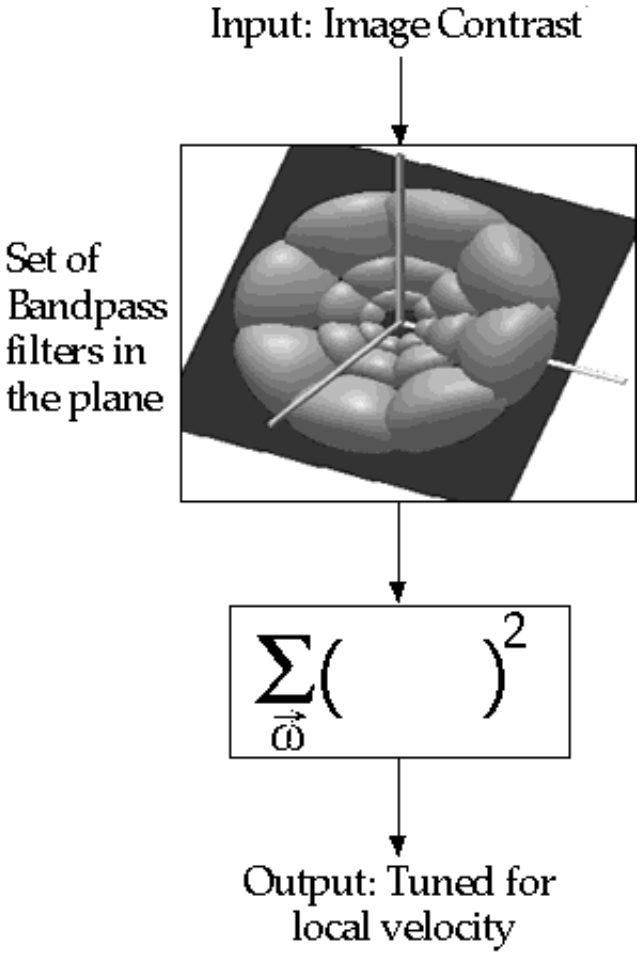

Figure 1.4: Planar power detector model for estimating image velocities.

estimator will have considerable uncertainty about the properties of the translating image. Chief among these are uncertainty about initial phases and the amplitude spectrum of the signal.

Properties 1 & 2 suggest that an image velocity estimator, in the absence of any a priori information about the spatial structure, should pool fourier information across the plane consistent with the expected image structure. Studies of natural image statistics suggest a reasonable model of the expected image structure is isotropic with a $1/f$ fall off in frequency magnitude f [45]. Thus, we might expect a velocity detector to tile a plane in Fourier space with bandpass filters which span all spatial orientations and have roughly one-octave frequency magnitude bandwidths (due to the $1/f$ fall off). If we further assume that the visual system's task is to detect translating images buried in Gaussian white noise, then Property 3 suggests a specific form for the velocity estimator: the optimal estimator in this case is a *planar power detector*, which sums the squared local fourier amplitude on the plane. The power detector is optimal for signals with unknown phase and amplitude buried in white noise [58, 128]. Although the background noise in the average motion detection task is unlikely to be either exactly Gaussian or white, the structure of a detector which is optimized for more complicated background noise can be expressed as extensions of the power detector model. In addition, the power detector model is the simplest model which can handle the expected kinds of uncertainty.

In the last ten years, several models for translation detection have been developed which are essentially planar power detectors [56, 53, 112]. Each of these models estimates local velocities by pooling the spectral power information across planar regions of frequency space. A version of this model is illustrated in figure 1.4. Depicted is a set of band pass filters which intersect a common plane. The outputs of these

filters are squared and summed. The result produces a mechanism which is tuned for a particular velocity. A population of these mechanisms each tuned for different velocities can code for several aspects of local motions, such as providing an estimate of image velocity which is (nearly) independent of contrast from the mean or peak of the population, representing multiple motions, the uncertainty in the velocity of a motion, and the distribution of motions present[112].

Many motion detection models other than the planar power detector have been proposed. Most of these, however, do not satisfy the criteria set out above, and hence are not generic translation detectors. For instance most models make particular assumptions the spatial structure of signals. This class includes 'motion energy' detectors which assume signals are one-dimensional [1], feature matching in which particular features are determined [88, 89], and matched filter models[143], among others. However, when these models are actually implemented, they typically combine the estimates from detectors sensitive to different patterns. When this combination is least squares and enough different patterns are averaged across, the resulting algorithms strongly resemble planar power detectors[112]. Thus the planar power detector model is useful in that it represents the optimal estimator of image translations under the assumptions of uncertainty and gaussian noise, and because it represents a natural approximation to a large set of motion detection models which estimate the velocity of generic patterns.

1.3 Thesis Overview

The goal of the thesis is to investigate the utility of the planar power detector as a model of motion processing in the visual system.

In chapter 2 I introduce a novel set of stimuli for which planar power detectors are the ideal observers. The stimuli are used detection experiments which qualitatively test two predictions made by the model: 1) observers should efficiently pool spectral power across spatial orientation when all the power lies on a common plane. 2) Planar configurations of spectral power should be more detectable than non-planar configurations. The results are consistent with the planar power detector model.

In chapter 3 a kind of perturbation analysis is used to analyze the data from chapter 2 to determine what frequency pooling strategies observers used in detecting the stimuli. Frequency space is divided into a set of bands and weights for each band are estimated. The results show that observers are able to use at least two distinct kinds of frequency weighting, one narrow band in orientation and one broad band in spatial orientation but which restricts its weights to a common plane. In addition, significant negative weights appear for all the stimuli, which suggests that the planar power detector model needs to be extended to handle some inhibitory interactions between frequency bands.

In chapter 4 we test the prediction of the model that power which lies on a common plane is *additively pooled*. We perform three tests of additivity using sets of three band-pass stimuli. In two of the tests the band pass stimuli lie on a common plane, and in the third one of the band-pass stimuli is off the common plane specified by the other two. For the configurations on a common plane subject's pooling is described by an additive law, while for the off-plane configuration pooling is significantly subadditive.

In chapter 5 we state the general conclusions of the thesis from the experimental results.

Chapter 2

Detection Experiments

2.1 Introduction

In the last chapter we introduced a model for local translation detection we called a 'planar power detector'. The purpose of this chapter is to experimentally test several qualitative properties of the model. The experiments involve comparing detection performance across a novel set of stochastic stimuli buried in white noise.

The chapter is divided into several sections. In the first section, the basic properties of the model to be tested are outlined. The stochastic stimuli are then introduced, and the ideal observer for their detection is discussed. In the next section the experimental logic is explained and predictions are generated. The remaining sections present the results of the experiments and a general discussion.

2.1.1 Properties of model

In this section we will describe several basic properties of the planar power detector which are experimentally testable.

Recall the basic structure of the model: the detector additively pools power around planes in spatio-temporal frequency space (fig. 1.4). The detector can be implemented as a two-stage computation [56, 53, 112]. In the first stage, motion energy is computed by squaring the output of spatio-temporal frequency tuned bandpass filters (e.g. Gabor filters). In the second stage, translation velocities are estimated by pooling the outputs of motion energy detectors whose peak frequencies lie on common planes.

The critical features of the model are (1) that the inputs to the pooling process are squared (or similarly rectified) filter outputs which provide estimates of spectral power within different pass-bands, (2) that the inputs to the pooling process span a range of spatial orientations, and (3) the inputs have preferred spatio-temporal frequencies lying in a plane.

The presence of planar power detectors in the visual system predicts observers should be able to efficiently pool across planar regions of frequency space. This observation can be re-expressed as two general hypotheses outlined below, which will be more precisely formulated in the predictions section. The first hypothesis is that human observers will be better at detecting signals whose amplitude spectrum lies around a plane (planar signals) than signals which do not (non-planar signals). In addition, planar signals contain power at all orientations. The second hypothesis is that human observers should be at least as efficient at detecting signals whose power which include all spatial orientations as those which are narrow band in orientation. This will be subsequently referred to as the hypothesis of perfect pooling across orientation.

2.1.2 Stimuli and Task

The experimental paradigm is based on a simple idea. We choose a task and stimuli matched to the properties of planar power detectors and have subjects maximize their performance on the task. Performance will then be limited by the mismatch between the visual mechanisms and decision rule used by the visual system and the optimal mechanisms and decision rule. Measuring this mismatch gives a basis for inferring the plausibility of the planar power detector model for human vision (see [138] for a similar experimental paradigm). First we will explain the matched task and stimuli, after which we will describe the experimental logic.

Because the detectors only depend on signal power, they are insensitive to the phase spectrum of the signal. Thus all stimuli with the same power spectra will stimulate the detectors equally. In particular, we show in Appendix A the detectors are ideal for detecting signals in white noise whose phase spectra are random but whose average power spectra are equal to the detectors' spectra . A simple method to construct stochastic stimuli whose expected power spectra have a particular shape is to pass spatio-temporal Gaussian white noise through a filter with the desired shape. Thus, we can create stimuli matched to planar power detectors by passing white noise through a filter which preserves the frequencies lying around a plane. Stimuli constructed by filtering white noise are spectrally flexible because filters can be designed which have nearly any spectral shape. Although spatially filtered noise stimuli have been frequently used before in vision research[69, 87, 81, 63, 65], to our knowledge the use of spatio-temporally filtered noise stimuli in vision is novel.

We constructed three types of stochastic signal stimuli by passing spatio-temporal Gaussian white noise through three different configurations of band-pass filters. The first type of signal stimulus had a power spectrum confined to a single pass-band, which we term *Component* stimuli since they form the components of the other two stimuli. Level sets for the Component filter are shown in fig.2.1a. The *Planar* signals had a power spectrum confined to an annulus in frequency, which was created by passing white noise through the sum of a set of Component filters at different orientations in a common plane (fig.2.1b). The *Scrambled* signals had a power spectrum which was a scrambled version of the Planar signal, in which every other component filter in the annulus has the sign of its temporal frequency inverted (fig. 2.1c). Subjects detected these stochastic signals added to white noise in a two-interval forced-choice discrimination between signal-plus-noise vs. noise alone. Noise contrast energy was fixed and the total signal power (energy) was varied to find thresholds.

Examples of the stimuli are depicted in space-time as data cubes beneath the filters which created them in fig. 2.1d-f. Consider the example Planar signal shown in fig. 2.1e. In the space domain Planar signals are bandpass noises which have a spatio-temporal correlation structure consistent with a particular velocity. The stimuli have the phenomenal appearance of lumpy textures which non-rigidly drift rightward, somewhat like the movement of a shallow rocky stream. The x-y face of the cube shows an example of the spatial appearance of the texture. To imagine the motion, look at the x-t face of the cube. Luminance elements from the top of the x-y face shift rightward as time progresses into the page. However, unlike a rigid translation, the luminance elements fade in and out of existence and drift rightward for variable amounts of time.

These stimuli have three properties which make them much better suited to addressing questions about the pooling of Fourier energy than those previously used. The stimuli are matched to hypothesized detectors, they can be created with any spectrum by changing the filters, and they require a non-linear mechanism to process them. The last property stems from the fact that all the information is contained in the covariance of the signal. Linear filters compute weighted averages of a signal, which make them sensitive to the local average or mean function of the stimulus. The filtered noise stimuli have the property

that any weighted average is zero in expectation [1], which means the performance of a linear matched filter detector (ideal for exactly known deterministic signals) on these signals is chance [129].

In contrast, the vast majority of previous studies have used stimuli which have properties ill-suited to address hypotheses involving pooling of spectral energy. The two most common types of motion stimuli, drifting sinusoidal gratings[2] and moving fields of random dots, represent spectral extremes. Drifting sinusoids have their spectral power concentrated at a single spatio-temporal frequency and thus a single spatial orientation, while drifting dots are isotropic (all orientations are equally likely) and have their spectral power concentrated around a plane. In neither stimulus type is the frequency content adjustable, making them unsuited to questions about spectral pooling.

A third commonly used type of stimulus which is spectrally adjustable are sinusoidal plaids, consisting of drifting sinusoidal gratings summed together. Since sinusoidal plaid stimuli have two spectral components, questions about pooling can be addressed with this paradigm. However, plaid components are detected independently at threshold (i.e. they show no summation) [135] which has forced investigators to use less direct perceptual measures to study pooling. It has been shown that for a broad range of component gratings and for long enough viewing durations, the perceived direction and speed of the plaid pattern motion is given by the 'intersection of constraints' rule discussed in chapter 1 [2, 90, 72, 115, 18]. This shows that under most conditions the visual system is able to extract the actual velocity of a translating pattern. However, the studies do not address whether there exist mechanisms which are *specialized* for extracting image velocities. The perceptual results could be generated by any process which leads to a veridical percept, for which specialized planar power detectors are not necessary. Indeed several other hypotheses about plaid stimuli processing have been generated, including feature tracking [6, 145] and non-linear 1-D methods[3][43, 44, 147, 26]. In short non-optimal choices in stimuli have led previous studies to determine very little about spectral pooling in local motion processing, except the suggestion that some pooling must occur to explain the perceptual results with plaids.

2.2 Experimental Logic

In the main experiment we compare detection performance on Component, Planar and Scrambled signals. These comparisons were chosen to control for generic sources of task inefficiency so that our inferences are based on signal specific sources of task inefficiency. Generic sources of inefficiency include internal noise and generic spatial and temporal subsampling[4]. When observers detect signals which differ only in their configuration of power, it is reasonable to assume that the generic sources of inefficiency remain relatively constant. Given this assumption, the relative efficiency of detecting stimuli with different spectra provide a measure of how well the observer can use ('pool') all the signal power which factors out generic sources of inefficiency. Because the ideal detectors for the stimuli are power detectors matched to the signal spectrum, the primary source of suboptimal pooling of power in the signal is the mismatch between signal and internal filter spectra[5]. Thus in principle, we can use comparisons of performance across stimuli type to infer aspects of the filtering properties of the visual system.

One of the two experimental hypotheses outlined above is that observers should efficiently pool power across spatial orientation for frequency bands which lie in a common plane. To test this possibility, we

[1] in statistical parlance, the stimuli are zero mean Gaussian processes

[2] often windowed by a Gaussian function

[3] commonly called non-Fourier processing [29, 28].

[4] e.g. subsampling due to the cone array, the use of a discrete neural temporal code (spiking), etc.

[5] The term 'internal filter' is a convenient way of describing the power spectrum the observer pools over in detecting the signal.

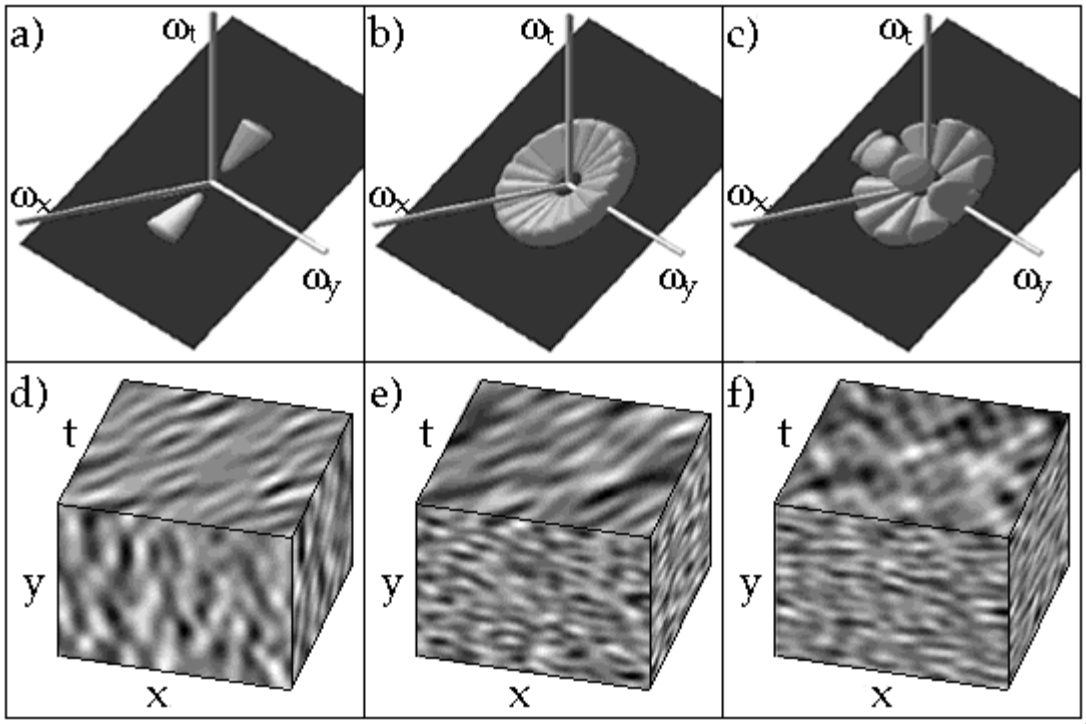

Figure 2.1: Filters used to generate experimental stimuli and the data cube representations of stimuli made by the filters. The top panels depicts level sets (65% of peak response) of the three different filters used to generate stimuli for this paper. The bottom panels are data block representations of the stimuli produced by passing spatio-temporal Gaussian white noise through the filters in the top panels. **a)** The 'Component' filter, which is a spatially and temporally bandpass filter, rotated to lie in the plane depicted in gray. **b)** The 'Planar' filter produced by rotating a set of 10 'Component' filters from a by multiples of 18 deg within the common plane shown in gray. These filters are arranged to tile an annular region of a plane in frequency space which specifies a rightward translation. **c)** The 'Scrambled' filter, which is produced by reflecting every other band pass filter in the 'planar' filter about the temporal frequency axis. **d)** The 'Component' stimulus produced by passing spatio-temporal Gaussian white noise through the filter shown in **a**. The x-y face shows the spatial characteristics of the stimuli: band-pass and oriented in the y direction. The x-t face shows the rightward motion information as oriented structure localized in time and which appear randomly across the image. **e)** The 'Planar' stimulus produced by passing spatio-temporal Gaussian white noise through the filter shown in **b**. The x-y face shows the spatial characteristics of the stimuli: band-pass and isotropic. The x-t face shows the rightward motion information as oriented structure localized in time and which appear randomly across the image. **f)** The 'scrambled' stimulus produced by this filter, as shown in the lower middle panel, has local oriented structure in the x-t slice consistent with both leftward and rightward motions.

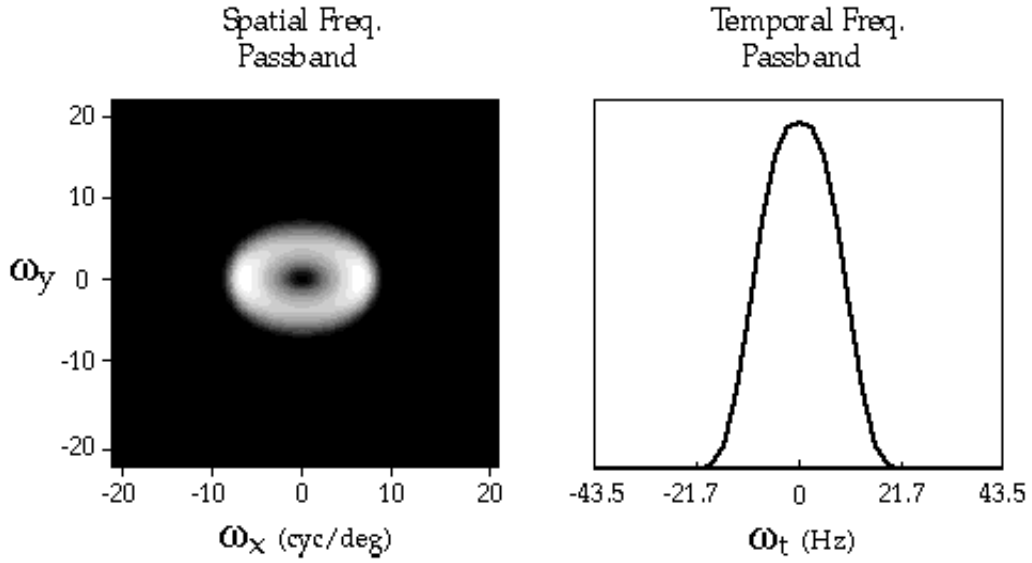

Figure 2.2: The average spatial and temporal frequency structure of both the Planar and Scrambled filters. The left side shows the spatial frequency amplitude spectrum of both filters averaged across temporal frequency. The right panel shows the temporal frequency amplitude spectrum of the filters, averaged over both dimensions of spatial frequency.

compared efficiencies for Component and Planar stimuli. Component stimuli have power concentrated around a single orientation while Planar stimuli have power at all orientations. The prediction can be rephrased as: observers should lose no more information detecting Planar stimuli than detecting Component stimuli. Let ν_{Pl} and ν_C represent the subject's efficiencies for Planar and Component stimuli respectively. Then the prediction can be expressed as: $\nu_{Pl} \geq \nu_C$, where $\nu_C = \nu_{Pl}$ represents the case in which the visual system is able to pool across orientation with no loss of information. On the other hand, if $\nu_C \gg \nu_{Pl}$ then it is unlikely that the visual system has detectors specialized for detecting planar configurations of power.

The other hypothesis states observers should be much more efficient at detecting planar configurations of power than non-planar configurations. We tested this hypothesis by comparing detection efficiencies Planar and Scrambled stimuli. Recall that Scrambled stimuli are constructed from Planar stimuli (fig 2.1b) by inverting ω_t for every other bandpass component. These stimuli represent the most non-planar configuration of power which can be constructed from the Planar stimuli's ten bandpass components which preserves an important spectral property: the spatial and temporal frequencies are matched except for the sign of some of the temporal frequencies. This equivalence should minimize the effect that anisotropies in the observer's spectral sensitivity might have on the performance comparison. The spectral equivalence also controls for the use of non-motion cues present in the stimuli. Both stimuli have the same average spatial 'lumpiness' and temporal flicker cues, because the filters have identical spatial frequency structure averaged across temporal frequency, and also have identical temporal frequency structure, averaged across spatial frequency. The spatial and temporal frequency structure of both filters is shown in fig. 2.2. Thus, if subjects are using the spatial appearance or temporal flicker of the stimuli as cues to detection, we would expect identical detection performance for the two stimuli.

2.3 Methods

2.3.1 Definitions

The following definitions are used to describe the physical properties of the stimuli. The *Power Spectral Density* is the squared Fourier amplitude spectrum of the stimulus, i.e. if the stimulus has a luminance distribution L(x,y,t) and $F(\omega_x, \omega_y, \omega_t) = \mathcal{F}\{L(x,y,t)\}$ is the 3-D Fourier transform of L, the power spectral density is $|F(\omega_x, \omega_y, \omega_t)|^2$.

The *energy* of a stimulus is the integrated power spectral density, $\int_{\vec{\omega}} |F(\omega_x, \omega_y, \omega_t)|^2 \, d\vec{\omega}$.

The *filter bandwidth* with respect to frequency coordinate variable ω_u is measured by taking the second moment of the spectrum of the normalized filter, $B_{\omega_u} = \int_{-\infty}^{\infty} \omega_u^2 \cdot |F_n(\vec{\omega})| \, d\vec{\omega}$, where $F_n = \frac{|F(\vec{\omega})|}{\int_{-\infty}^{\infty} |F(\vec{\omega})| \, d\vec{\omega}}$.

The *equivalent flat filter* is derived in Appendix A and is a rectangular filter which produces performance equivalent to a given non-rectangular filter. The equivalent flat filter is used to determine the effective number of samples used by the human observer.

The *equivalent background noise* is the additional background noise required to make the ideal observer perform at the same level as the human observer. It is a kind of 'referring the noise to the input' [5, 96]. Formulas for the equivalent background noise are derived in appendix A.

The *sampling efficiency* is a measure of the number of samples effectively used by the observer.

2.3.2 Apparatus

Stimuli were displayed on a Radius 20" grayscale monitor at nominal 12 bit grayscale precision by a Macintosh PowerPC. The display had a P104 phosphor and a frame rate of 75 Hz. Monitor MTF was measured and the central region used for display was verified to be linear for all but the 3 highest screen harmonics. Custom software was used to display the stimuli which employed VideoToolbox software written by Denis Pelli and David Brainard. Denis Pelli's Video Attenuator was used to achieve 12 bit precision. The attenuator sums together the 24 bit r,g, and b channel DAC outputs with simple resistors to produce a nominal 12 bit signal to the monitor [98]. Custom code was employed to rapidly transform floating point luminance images into 24 bit r,g, b inputs for the attenuator. Subject's heads were kept stationary during the experiment using a chin/head-rest.

2.3.3 Stimuli

Signal stimuli were produced by digitally filtering spatio-temporal gaussian white noise with spatio-temporal linear filters using the discrete fourier transform. Spatio-temporal gaussian white noise was produced by transforming a double precision pseudo-random number generator with a rectangular distribution. After filtering, the stimuli were truncated at $\pm$3.5 standard deviations. The probability of exceeding those bounds after filtering is much less than 4.7x10^{-4}. Since the stimuli have about 1.3x10^5 samples, only a few values were truncated on a given stimulus. Signal stimuli were added to either unfiltered background noise, or filtered background noise. Unfiltered background noise was spatio-temporal gaussian white noise which was bounded by systematically resampling any values which exceeded $\pm$2 standard deviations. The filtered background noise was produced in the same manner as the signal stimuli. Both signal and noise stimuli had a mean luminance of (25) cd/m^2 which was the same as the background luminance.

Stimuli dimensions were 64x64 pixels in space and 32 frames in time. The viewing distance was 44 cm, and the stimuli subtended 2.2 deg. Subjects viewed the monitor binocularly in a darkened room, and fixated a spot 1.3 deg above the center of the stimulus. Stimuli were modulated in space by a circularly symmetric 'smooth pillbox' window and in time by a 'smooth box' function. The 'smooth box' function

$W(r)$, $r = \sqrt{x^2 + y^2}$ is given by:

$$W(r) = \begin{cases} 1, & l + w < r < u - w \\ 0, & l < r > u \\ \cos[\pi(r - l)/w], & \text{left transition} \\ \cos[\pi(u - r)/w], & \text{right transition} \end{cases}$$

where l and u are the left and right boundaries of the box function, respectively. The aperture window had a radius of $r = 1.1$ degrees, with transition width $w = 0.26$ degrees. The temporal window had an identical form, replacing r with t. For the temporal window stimulus onset and offset transition durations were $w = 0.053$ seconds. The fourier transform of this window function is a blurred sinc function, which can be approximated by a gaussian with sigma of 0.034 cyc/deg (1 pixel in fft). For the stimuli employed, windowing the stimuli changes the spectral density by at most 8% at any given frequency and produces an average change less than 3%. Performance calculations which assumed no windowing differed from performance calculations which included windowing by less than 1%, hence the effect of windowing on the signal shape was ignored. Windowing does reduce the total energy available to the observer. This reduction was factored into the ideal observer calculations.

2.3.4 Subjects

Three subjects took part in the experiments, one being the first author (PS) and the other two being undergraduates who were experienced psychophysical subjects but were naive to the purpose of the experiments (ML and AS). All three observers had corrected to normal vision. Viewing was binocular with natural pupils.

2.3.5 Procedures

Data were collected using the method of constant stimuli for two temporal interval forced choice discrimination between signal plus noise and noise alone. Typically 5 or 6 different energy levels were chosen, and these were uniformly intermixed during the sessions. Within a session, one type of signal was randomly selected and presented, and subjects were told which signal type was present at the start of each session.

Before this data was collected, observers had at least three hours practice for detecting each signal type. Practice data was collected using a QUEST[142] adaptive procedure, and subjects were given practice until their last three thresholds were not significantly different. In each practice session, the signal type was randomly selected and used for the entire session. An exception to this procedure were the Gabor filter stimuli, on which data was collected after the completion of the other data sets.

Thresholds were determined by fitting 2 parameter Weibull functions to the detection data using a maximum likelihood procedure. The α parameter was used as a measure of threshold, which is equivalent to reporting the signal to noise ratio which produced 81.1% correct. Error bars for thresholds, slopes, and efficiencies are computed from the inverse numerical Hessian of their likelihood functions, which were cross-validated using a parametric bootstrap procedure. In the bootstrap procedure, 1000 data sets were simulated by sampling from the binomial distributions of the observer's data points. Maximum likelihood fits of the parameters were then generated for each data set. The distribution of fitted parameters was used to estimate the standard error on the parameters. The bootstrap distribution of fitted α and β parameters were also used to compute bootstrap T-tests and ANOVAs [40].

2.3.6 Filters

Three different filters were used to create stimuli. One filter is a rotationally symmetric bandpass filter (the *Component* filter), and two other filters are created by summing together spherically rotated copies of this bandpass filter (the *Planar* and *Scrambled* filters.

Component Filter The Component filter's amplitude spectrum has this functional form in spherical spatio-temporal frequency coordinates:

$$C(\omega_r, \omega_\theta, \omega_\phi) = R(\omega_r)|(\cos(\omega_\theta - \omega_{\theta_0})|^9 |\cos(\omega_\phi - \omega_{\phi_0})|^9 \tag{2.1}$$

Where $R(\omega_r)$ is given by the smooth box function with transition region width of 1.45 cyc/deg, and low-high frequency cutoffs of (0.49,7.6) cyc/deg. This filter has an orientation bandwidth of about 18 degrees, and an ω_ϕ bandwidth of 18 degrees. To interpret ω_θ and ω_ϕ, recall that the plane in frequency space is given by $v_x\omega_x + v_y\omega_y + \omega_t = 0$, thus the normal vector to the plane is given by $\vec{u} = (v_x\, v_y\, 1)/\sqrt{v_x^2 + v_y^2 + 1}$. The angles ω_{θ_0} and ω_{ϕ_0} can be determined from velocity through the normal vector:

$$\omega_{\phi_0} = \frac{\pi}{2} - tan^{-1}(\frac{1}{\sqrt{v_x^2 + v_y^2}}) \tag{2.2}$$

$$\omega_{\theta_0} = \pi + tan^{-1}(\frac{v_y}{v_x}) \tag{2.3}$$

The peak frequency of this filter had $\omega_{\theta_0} = 90\,\text{deg}$, and $\omega_{\phi_0} = 36.9\,\text{deg}$ which corresponds to a grating moving in the negative y direction at a speed of 1.93 deg/sec.

Planar Filter The Planar filter is produced by summing together 10 BandPass filters rotated such that they lie on a common plane. The simplest way to express this is to express the BandPass filter in Cartesian coordinates. Then the Planar filter can be written as the sum of BandPass filters rotated in direction by multiples of 18 deg, then rotated in ϕ to lie in a common plane. Let $\vec{\omega}$ represent the vector $[\omega_x, \omega_y, \omega_t]$, and $\mathbf{R_x}(\phi)$ and $\mathbf{R_t}(\theta)$ denote the 3-D rotation matrices which leave the ω_x and ω_t axes fixed respectively. Then the Planar filter can be expressed as:

$$Pl(\vec{\omega}) = \sum_{i=1}^{10} C\left(\mathbf{R_x}(-\phi_0)\mathbf{R_t}(-\theta_0 - \mathrm{i}\,\pi/10)\,\vec{\omega}\right) \tag{2.4}$$

Scrambled Filter The Scrambled filter's amplitude spectrum is obtained from the Planar filter's by rotating every other component BandPass filter around the ω_t axis by 180 deg.

2.3.7 Ideal Observer calculations

Ideal observer performance was calculated in two ways. The first is from an approximate expression for the performance which is derived in appendix A. The second was a simulation of the ideal observer performance on the stimuli used in the experiment, except that luminance quantization was not modeled. The simulations were performed by computing the energy within the filter in each interval and then choosing the interval which contained the largest energy to compute a binary response. The probability of correct detection at each signal energy was then estimated by performing a thousand trials per signal energy. The energy within the filter for each interval was computed by taking the 3-D fft of the signal plus noise interval and the noise alone interval, multiplying by the optimal filter spectrum, then summing across all the squared complex amplitudes. The simulations showed the approximations to be accurate to much less than 1%.

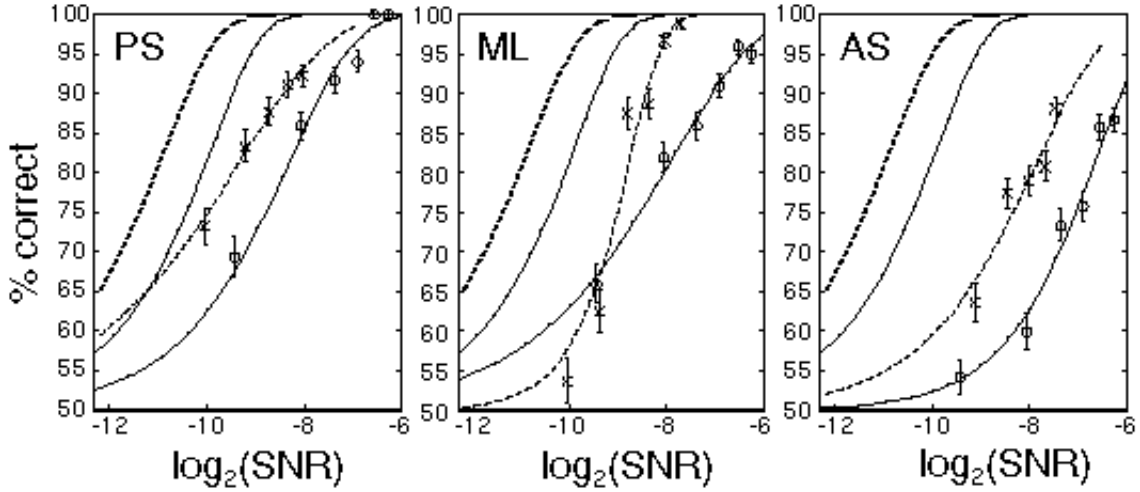

Figure 2.3: Probability correct is plotted as a function of log signal to noise ratio (SNR), base 2, for both Planar (solid lines and circles) and Component stimuli (dashed lines and crosses). The psychometric functions to the left with no data points are the theoretical performances of the ideal observers for Planar stimuli (solid line) and Component stimuli (dashed line).

2.4 Results

The performance of human and ideal observers for detecting Planar and Component signals (see fig. 2.1) are shown in figure 2.3 as a function of signal to noise energy ratio (SNR). The data are plotted on a log base 2 scale, so that a factor of two difference in threshold SNR becomes a one unit shift in the psychometric function. Solid curves plot human and ideal performance for the Planar stimuli, while dashed curves plot performance for the Component stimuli. The data show that detection threshold energies at any % correct are lower for Component stimuli than Planar, except for subject ML for which the psychometric functions cross.

Care must be taken in interpreting this difference in threshold energies. To infer subjects are better at detecting Component stimuli implicitly assumes that Planar and Component stimuli with the same energy should be equally difficult to detect. This assumption turns out to be incorrect. To understand why Planar and Component stimuli matched in energy are not matched in detectability, we need the concept of an ideal observer. An ideal observer is the theoretically optimal observer for a given task given the statistics. Let the signal power spectrum be denoted by $|\mathbf{S}(\omega_x, \omega_y, \omega_t)|^2$, and let $|\mathbf{F}(\omega_x, \omega_y, \omega_t)|^2$ denote the spectrum of the filter which made the signal. For the task of detecting the stochastic stimuli buried in white noise, the ideal observer computes the energy within the filter $\mathbf{F}$:

$$E = \int_{\omega_x,\omega_y,\omega_t} |\mathbf{S}(\omega_x, \omega_y, \omega_t)|^2 |\mathbf{F}(\omega_x, \omega_y, \omega_t)|^2 \, d\omega_x d\omega_y d\omega_t \tag{2.5}$$

on each interval and chooses the interval with the larger energy. It turns out that ideal performance changes with total bandwidth (filter volume) of the ideal's filter, since the number of frequency samples available and the amount of background noise viewed by the ideal's filter changes. Because the expressions for ideal performance are quite complicated, it is useful to examine a case for which a simple result can be achieved. When the filters and signals have rectangular power spectra, a useful approximate expression for d' can be derived [52, 65]:

$$(d')^2 = \frac{2(E/N)^2}{MB_xB_yB_t + E/N} \tag{2.6}$$

where E is the signal energy, N is the background noise power, M the number of signal samples, and the B_u are the bandwidths for of the filter for the u variable. The expression shows that even in this simple case ideal performance is a function of the filter bandwidths, and hence a function of stimulus type. This change in ideal performance across stimulus type indicates that the information available for detection varies with the properties of the signal, which means a straight comparison of thresholds across stimulus type is misleading. To correct for this difference in ideal performance across stimuli, we computed subject's efficiencies for detecting each stimulus type. Efficiencies represent an absolute measure of observer performance on a task which corrects for differences in inherent detectability across stimuli type.

Statistical efficiency ν for the task is defined as the percentage of available samples which are effectively used by subjects to detect the stimuli:

$$\nu = \frac{N_{ideal}}{N_{human}} \tag{2.7}$$

where N_x represents the number of samples required by the observer to achieve a given level of performance. An approximate expression for efficiency is given by the squared ratio of ideal and human observers detection thresholds:

$$\nu = \frac{E^2_{ideal}}{E^2_{human}} \tag{2.8}$$

where E_x are the threshold signal energies for the human and ideal observers. Because this approximation is frequently poor we used formulae derived in Appendix B to compute efficiencies.

Using the ratio of thresholds approximation, the efficiencies can be estimated as $1/4d$, where d is the difference between the subject's and ideal's thresholds on the log scale in figure 2.3. For instance, differences of one, two, and three on a $\log_2$ scale yield efficiencies of about 25%, 12.5%, and 6.25% respectively. Notice the difference in ideal performance for Planar and Component stimuli. The difference means Planar stimuli intrinsically require more energy to be equally detectable. By inspection, the difference in subject's energy thresholds between Planar and Component stimuli is similar to the difference in ideal observer thresholds, indicating the efficiencies are comparable.

To make more straightforward comparisions of ideal and human behavior, we found the best fitting Weibull function to the ideal psychometric function using a least squares criterion. Figure 2.4 shows the Weibull fit parameters α and β for the ideal observer as a function of efficiency. Decreased efficiency for the ideal is modeled by incrementing the ideal's decision variable variance. Surprisingly, the β parameters are approximately constant across efficiency and stimulus type, which means the ideal psychometric functions are nearly shift invariant on a log SNR scale.

The slopes β, of the fitted Weibull psychometric functions are shown in figure 2.5 on the right. The slopes are not significantly different except for the Component condition for subject ML. Since the slopes for the ideal observers are constant across efficiency and stimulus type, we can directly compare their magnitudes. When the slopes are not significantly different, relative performance can be sufficiently summarized by the α parameters, which can be used to compute the efficiencies. Efficiencies were computed from the Weibull α parameters, which are a measure of the threshold signal to noise ratio at 81.1% correct, and are shown in figure 2.5 on the left.

In order to make performance comparisons for subject ML, we computed an average efficiency across the measured data points. For each measured percent correct, we can compute an efficiency with respect to the ideal at the same percent correct. Assuming each measured percent correct is equally informative, averaging across these efficiency estimates is the best measure of the subject's overall performance on a stimulus type. The average efficiencies are shown in figure 2.6 and have the same qualitative trends as the efficiencies computed from the α estimates. Thus, Planar and Component performance is not significantly different in expectation across performance criteria, consistent with the hypothesis that observers can pool

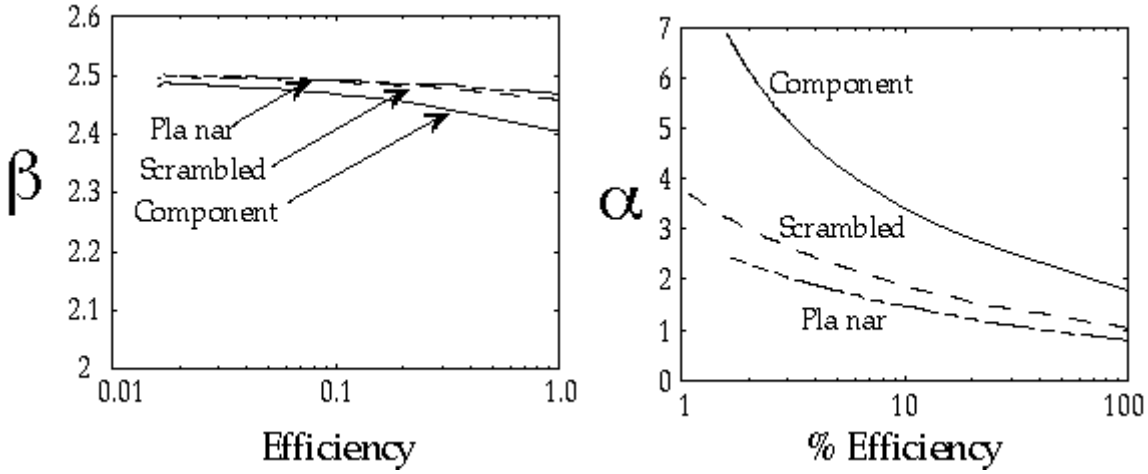

Figure 2.4: Left: Weibull β (slope) parameter estimates for as a function of efficiency for Planar and Component stimuli for the ideal observers. **Right:** Weibull α parameter estimates for as a function of efficiency for Planar and Component stimuli for the ideal observers

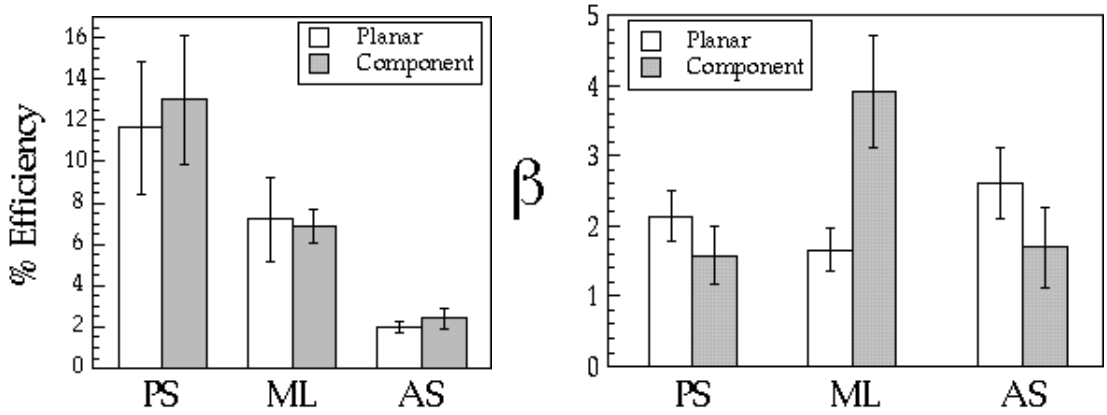

Figure 2.5: Left: Efficiencies for Planar and Component stimuli for three subjects computed from the fitted Weibull α parameter. Error bars represent standard errors of the estimate. **Right:** Weibull β parameter plotted for Planar and Component stimuli. Error bars represent standard errors of the estimate. Slopes are mostly close to the ideal β of 2.5, and are not significantly different across conditions except the Component condition for subject ML.

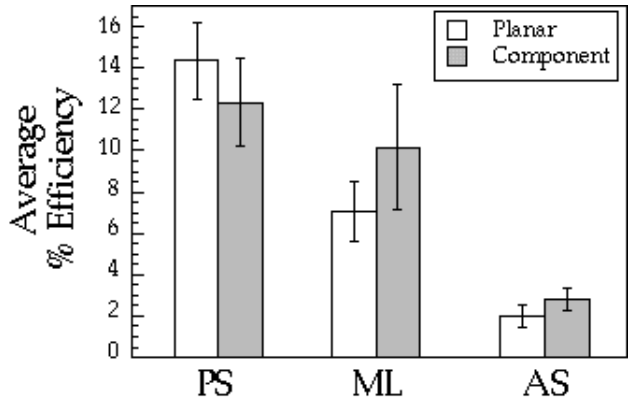

Figure 2.6: Average efficiencies for Planar and Component stimuli for three subjects computed from the fitted Weibull α parameter. Error bars represent standard errors of the estimate.

across planar regions of Fourier space without loss of information.

The results are consistent with perfect pooling across the set of bandpass visual filters which are sensitive to Planar stimuli. To better assess the strength of this inference, we computed predictions for Planar performance using two suboptimal detection strategies: no pooling across the Planar components and probability summation across the set of components. The purpose of the predictions is to give quantitative insight into the kind of performance which can be achieved by suboptimal pooling.

To make performance predictions we assumed that the observer is faced with the same sources of inefficiency in processing components of the Planar stimuli as faced detecting Component stimuli. We estimated the equivalent input noise for Component stimuli, which measures the additional background noise power needed for the ideal observer to perform the same as human observers. The equivalent input noise is then treated as additional background noise which corrupts the suboptimal model performance on Planar stimuli.

To compute predictions for the suboptimal model which does not pool across Planar components, we notice that assuming no pooling is the same as assuming observers can only use information from a single Planar component. The model uses energy within one of the component filters corrupted by the equivalent input noise as a decision variable. Detecting Planar stimuli using the power within one component predicts a reduction of efficiency by a factor of about 5 over performance on Component stimuli, i.e. each Planar component 'sees' about 20% of the available information.[6]

We also estimated the performance for another standard suboptimal model, probability summation across the Planar components. We modeled probability summation by a maximum output rule [96], and used the equivalent input noise to estimate subject inefficiency as before. Since the Planar component filters overlap, the probability summation rule is not as simple as that derived for independent channels. Computing the predictions for both suboptimal models is discussed in detail in Appendix C.

Threshold SNRs and efficiencies for the two stimuli are shown next to the probability summation and component filter predictions in fig. 2.7. For all three subjects, Planar and Component efficiencies are not significantly different (bootstrap T-test, $p<0.05$), while efficiencies for Planar stimuli exceed those of either the probability summation or component filter predictions by a factor greater than two. Although the predictions were generated using particular filters, the prediction results suggest that it would be difficult to achieve the efficiencies observed for Planar stimuli using a suboptimal pooling strategy. This suggestion

[6]Since there are ten component filters, we might expect each component to contain about 10% of the available information. However, the overlap of the filters coupled with a change in decision variable variance with filter volume conspire to double this expectation.

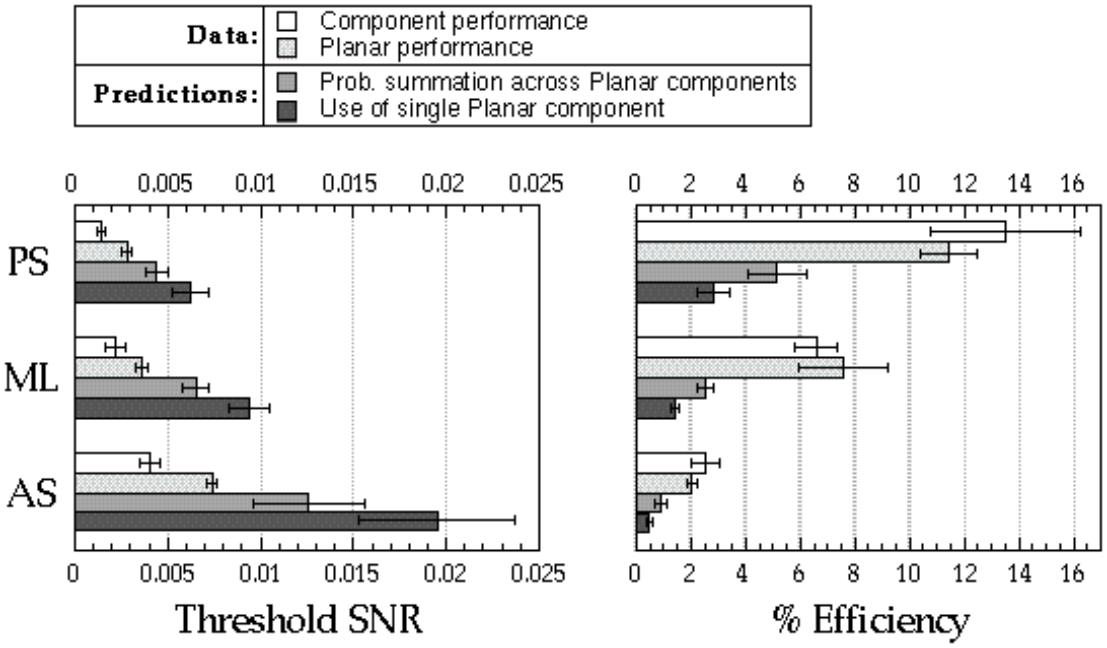

Figure 2.7: Summary of subject performance on Component and Planar stimuli. 81% threshold energies are shown on the left and efficiencies are shown on the right for Component (white bars) and Planar (light gray bars) stimuli. The error bars represent the standard error of the estimates. The two suboptimal pooling predictions for the detection of Planar stimuli are shown as dark gray bars. The error bars for the predictions represent the prediction uncertainty generated by the standard errors for Component detection.

is supported by the fact that all of the existing psychophysical and physiological evidence [138, 143, 83] suggests that spatio-temporal filters in the early visual system are narrow-band in spatial frequency, since these filters would 'see' less of Planar stimuli than Component stimuli. For instance, Watson's 'optimal motion filter' passes 60% of the information[7] in component stimuli but only 14% of Planar stimuli, a 4:1 ratio. Suboptimal schemes like probability summation lose more information the narrower the filter bandwidths are and lose the least information when the filters are closest to the Planar filter spectrum. Thus we would not expect probability summation over Watson's filters to overcome the 4:1 ratio to produce equal performance on Planar and Component stimuli.

The results show that on average the same percentage of information is lost for both Planar and Component stimuli, even though in the Planar case the information is spread across orientation. This suggests that observers can use a planar pooling strategy.

2.4.1 Planar vs. Scrambled

The previous result suggests the visual system can efficiently pool across planar configurations of power. In this section we control for the possibility that generic (rather than plane-specific) pooling processes could explain detection performance on Planar stimuli. In other words, if the visual system can learn to pool efficiently across arbitrary configurations of power, given enough practice, then there is no need to hypothesize specialized planar power detectors. This possibility is suggested by the results of Kersten[65] in the spatial frequency domain, who found efficiencies for detecting spatial noise signals remained constant over a 6 octave range of signal bandwidths, providing evidence that arbitrary pooling may be possible in

[7] In this context, the amount of information means the number of independent and equivalent frequency samples in the stimulus.

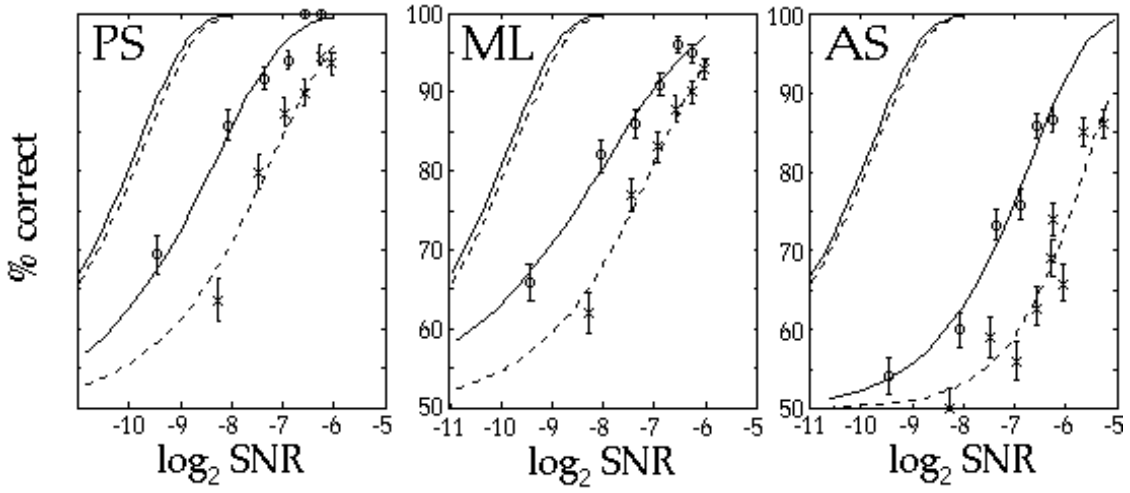

Figure 2.8: Probability correct is plotted as a function of log signal to noise ratio (SNR), base 2, for both Planar (solid lines and circles) and Scrambled stimuli (dashed lines and crosses). The psychometric functions to the left with no data points are the theoretical performances of the ideal observers for Planar stimuli (solid line) and Scrambled stimuli (dashed line).

the spatial domain.

We investigated the possibility that observers may be equally good at pooling energy across non-planar as planar configurations by comparing detection efficiencies for Planar and Scrambled stimuli. Recall that Scrambled stimuli have the most non-planar configuration which can be produced while preserving the spatial and temporal structure of the Planar stimuli. The matching spatial and temporal structure controls for the possibility that a difference in performance is simply due to differences in the visual system's sensitivity to the frequencies in the two stimuli. Thus if the visual system uses a generic pooling strategy we would expect equivalent performance on Planar and Scrambled stimuli.

Psychometric data for Planar and Scrambled stimuli are shown in figure 2.8. Planar data is replotted from figure 2.3.

The slopes of the fitted Weibull psychometric functions are shown in figure 2.9 on the right. The slopes are not significantly different for the three observers. As before, this allows us to summarize our data using the threshold parameters. Efficiencies were computed from the Weibull α parameters, which are shown in figure 2.9 on the left. The results show that Planar efficiencies are consistently higher than Scrambled. The efficiencies for Planar stimuli are more than twice those for Scrambled stimuli for all three subjects. Thus, the visual system is not able to perform arbitrary efficient pooling.

To better quantify how much pooling occurred for Scrambled stimuli, we compared Scrambled performance with the predictions from the no-pooling and probability summation suboptimal strategies. The predictions were made in the same manner as for Planar stimuli: the equivalent input noise performance on Component stimuli was used to predict the effects of using a single Scrambled component and probability summation across these components.

Comparisons of predicted and actual efficiencies at 80% correct are shown in figure 2.10. The results show that efficiencies for Scrambled stimuli are processed more efficiently than predicted by a single Component filter for all three subjects, but less efficiently than predicted by probability summation for two subjects. Observer ML performance is not significantly different from probability summation. Thus, although Scrambled performance is well below Planar performance, there is evidence for some pooling which is within the range expected for a system which detects the components independently.

The result indicates that the visual system is not as sensitive to all non-planar configurations as planar

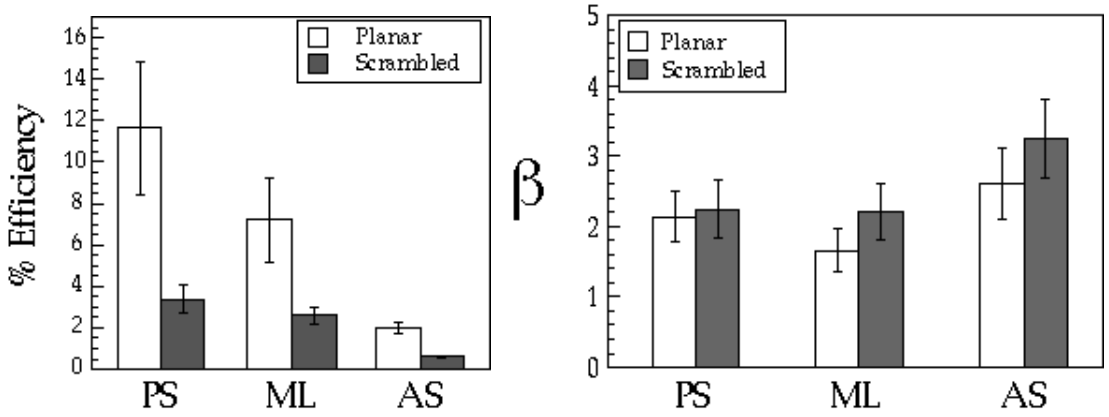

Figure 2.9: Left: Efficiencies for Planar and Scrambled stimuli for three subjects computed from the fitted Weibull α parameter. Error bars represent standard errors of the estimate. **Right:** Weibull β parameter plotted for Planar and Scrambled stimuli. Error bars represent standard errors of the estimate. Slopes are close to the ideal β of 2.5, and are not significantly different across conditions.

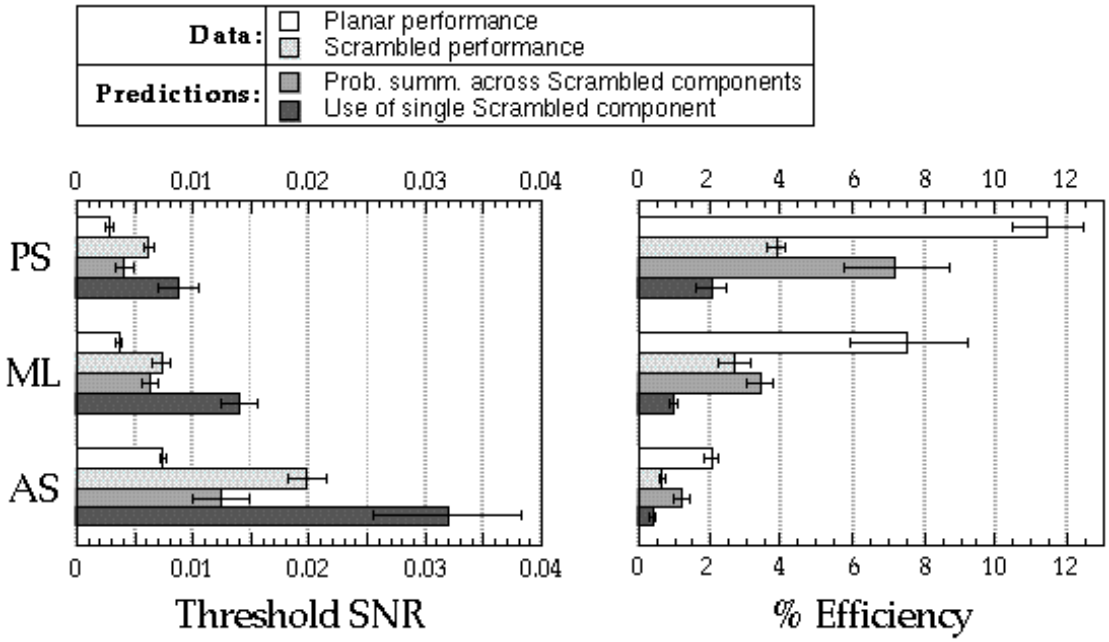

Figure 2.10: Subject performance on Scrambled stimuli compared to predicted performance and performance for Planar stimuli.

configurations. Since the Scrambled stimuli are matched in spatio-temporal frequency content and in spatial and temporal structure, this conclusion cannot easily be attributed to a difference in spectral sensitivity to the two stimuli. Despite the large difference between Planar and Scrambled efficiencies, the visual system is not insensitive to Scrambled stimuli, since efficiencies are higher than those predicted from a single component.

2.5 Experiment 2

In experiment 1 we compared detection performance for Planar and Component stimuli and remarked that the near equivalence of efficiency suggests observers pool information across orientation in a plane. Implicit in this inference is the assumption that the observer used different internal filters for detecting Planar and Component stimuli which were reasonably matched to the spectra of the stimuli. However, it is possible that the visual system relies on bandpass detectors to detect planar and component stimuli which are not matched to either stimuli's spectra. If these bandpass detectors process Planar and Component stimuli with about the same efficiency, then the equal detection efficiencies for Planar and Component stimuli could be explained without hypothesizing specialized planar power detectors. Note that this hypothesis does not require Planar and Component stimuli to be equally processed by a single bandpass detector. For instance, there could be two bandpass detectors each of which is more sensitive to one of the stimuli than any other detector in the visual system, and such that the sensitivities of these two detectors cause the two stimuli to be processed with equal efficiency. In this section we test this possibility.

2.5.1 Experimental Logic

For a non-matched bandpass filter to explain the Planar detection results, the processing efficiency of the detector must be higher than the measured Planar efficiency, since it needs to overcome the additional information loss caused by the mismatch. Let ν_s^f denote the efficiency for detecting a signal with spectrum s using an internal filter with spectrum f, and $m(s, f)$ represent the fraction of the total signal information preserved by the internal filter. Then the efficiency of the observer relying on the mismatched internal filter can be expressed as:

$$\nu_s^f = m(s, f)\nu_f^f \tag{2.9}$$

In words, the efficiency of the observer using the filter is the product of the maximal efficiency obtainable using the filter, ν_f^f, and the fraction of signal information preserved by the filter. Thus, if the observer is using a mismatched internal filter to detect the Planar and Component signals, then the observer should be more efficient at detecting stimuli matched to this internal filter by a factor of $1/m(s, f)$.

What sort of bandpass internal filters could explain Planar performance? In order to produce high relative efficiencies on Planar stimuli the bandpass filter should be better matched to Planar stimuli than Component stimuli. This means that the bandwidths of this filter must be large enough that the filter 'sees' more additional background noise processing Component stimuli than it discards signal power processing Planar stimuli. Such filters do exist, hence the hypothesis is viable. A special role is played by the bandpass filter which can best process the Planar stimuli. This filter is better matched to the planar spectrum than any other filter narrow band in orientation. If we make stimuli with spectra matched to this filter, then any internal filter narrow band in orientation which could explain the Planar results must be better matched to these stimuli than Planar stimuli. Implicit in this prediction is an assumption that the sensitivity of the observer is reasonably isotropic, an issue we will return to in the discussion.

To test this prediction, we found the Gabor bandpass filters which could best process the Planar stimuli. Gabor filters were chosen because of their frequent use in modeling the early visual system

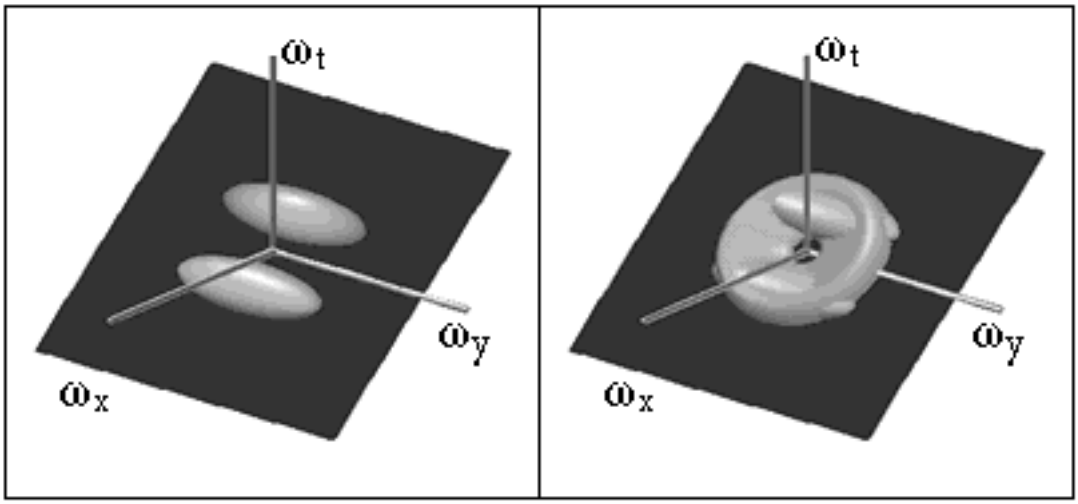

Figure 2.11: The Gabor filter which optimally processes Planar stimuli. The right panel contains the 50% level sets of the optimal Gabor filter. The method for finding this filter is given in Appendix D. On the right is an image of the optimal Gabor filter intersecting the Planar filter, to illustrate the fit.

[31, 32, 56, 110, 74, 143] and their simple parametrization. The 50% level sets of a resulting filter is shown in fig. 2.11 in the first panel. The filter is unique except that it may be rotated within the Planar filter spectrum. We created a set of control stimuli by passing white noise through the optimal Gabor filter shown in the figure, which was chosen to produce stimuli with the same direction of motion as the Planar stimuli. Under the assumption that observer's sensitivities are isotropic, filter bandwidths become the principle determinate of a bandpass filter's processing efficiency for Planar stimuli. Under isotropy, single band-pass filters with bandwidths large enough to potentially detect Planar stimuli better than Component stimuli would be better matched to Gabor stimuli than Planar stimuli. Thus, if performance on the planar stimulus is based on a single pass-band filter, we would predict higher detection efficiencies for the Gabor control stimuli than Planar stimuli.

2.5.2 Methods

The Gabor filter used is a sinusoid enveloped by a gaussian window which is given by:

$$G(\vec{\omega}) = exp(-\frac{1}{2}((\vec{\omega} - \vec{\omega_0})^T \Lambda^{-1} (\vec{\omega} - \vec{\omega_0})) \tag{2.10}$$

The parameters of the Gabor filter were determined by a fitting procedure described in Appendix D. $\vec{\omega_0}$ was set to $[x, y, t]$ and Λ is a diagonal matrix, with non-zero entries given by the squared bandwidths $[\sigma_x, \sigma_y, \sigma_t]$ = (2.4 cyc/deg, 6.2 cyc/deg, 5.0 Hz). The filter is centered at (2.73 cyc/deg, 0 cyc/deg, 5.3 Hz).

The experimental procedures were the same as for the previous experiment. Subjects were given at least two hours training before data collection using a QUEST[142] procedure to track performance and estimate thresholds. At the end of the training period the difference between the last four thresholds was not significantly different from the group mean (bootstrap ANOVA, $p < 0.05$.). The data for this experiment were collected after the data for experiments 1 & 3. [8]

[8] Although not critical it is noteworthy that subject performance was initially much better and leveled out faster for all of the stimuli used in experiments after the Planar, Scrambled, and Component.

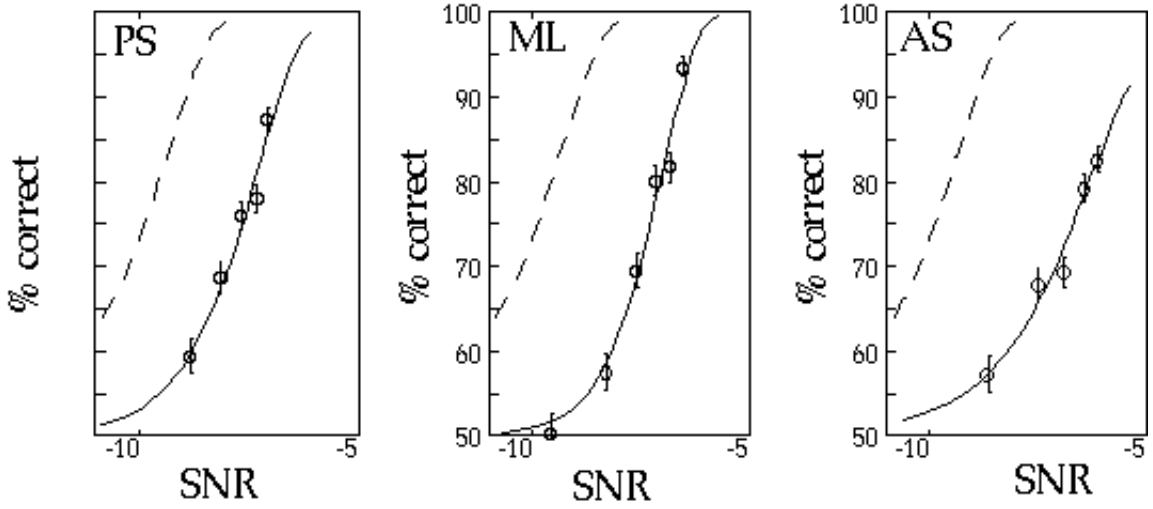

Figure 2.12: Performance on Gabor stimuli is shown for three subjects. Dashed line represents ideal observer performance. See figure 2.8 for details.

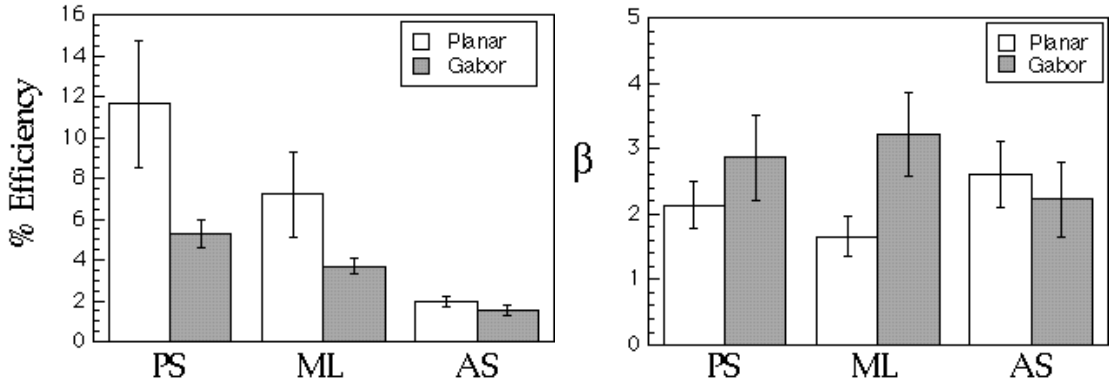

Figure 2.13: Left: Efficiencies for Planar and Gabor stimuli for three subjects computed from the fitted Weibull α parameter. Error bars represent standard errors of the estimate. **Right:** Weibull β parameter plotted for Planar and Gabor stimuli. Error bars represent standard errors of the estimate.

2.5.3 Results

The slope parameters, β, of the Gabor and Planar psychometric functions are shown in fig. 2.13 on the right. The slope parameters are not significantly different using a bootstrap T-test at the p¡0.05 level, so the data were summarized by the efficiency computed from the threshold parameter. Efficiencies for Planar and Gabor stimuli are shown on the left in fig. 2.13. The results show that Planar efficiencies are consistently higher than Gabor, not consistent with the idea that Planar stimuli are being processed by single bandpass filters.

The discrepancy between the results and the hypothesis that the Planar results can be explained by a single bandpass filter is accentuated by considering the predicted efficiency for Gabor stimuli. The prediction was generated by determining how much information would be lost in processing the Planar stimuli using the optimal Gabor filter. The intersection of this Gabor filter and the Planar filter are shown in the second panel of fig. 2.11. Visually, the Gabor stimulus appears to cover about 2/3 of the frequencies in the Planar stimuli, which suggests the filter should discard about 33% of the information in Planar stimuli. The actual maximum processing efficiency of 65.7% is close to this visual estimate. Thus if the

visual system were using a bandpass filter with bandwidths similar to this Gabor filter to detect the Planar stimuli, the filter must be operating with a minimum efficiency of $\frac{1}{0.657} \simeq 1.5$ times the Planar efficiency. An internal bandpass filter with different bandwidths would have to be even more efficient to explain the results. Since performance on Gabor stimuli shows the opposite trend, the results argue strongly against a single bandpass filter model.

2.5.4 Discussion

The result indicates that the observers are not using a bandpass filter which is narrow band in orientation to detect the stimuli, under the assumption that the visual system's sensitivity is reasonably isotropic.

The assumption of isotropy is needed to cover the following possibility. The Gabor stimuli we chose does not cover all of the frequencies contained in the Planar stimuli (see the second panel in figure 2.11). It is possible that all of the observers are detecting Planar stimuli using the frequencies in the side bands excluded from the Gabor stimuli. However, for this hypothesis to be feasible, visual efficiency for the frequencies in these sidebands must be at least 3 times the efficiency for the shared frequencies, since the sidebands constitute only 1/3 of the frequencies in Planar stimuli. This possibility is unlikely given previous data which do not show much anisotropy in visual sensitivity[64, 123], and by the fact that the shared frequencies are much closer to measured optimal temporal frequencies (5.5 Hz, compared with Watson & Turano's 5 Hz[143].) than the sideband temporal frequencies (which are less than half that).

While we framed the discussion in terms of the a single filter detecting the stimuli, the results generalize to probability summation across a set of band-pass filters as well. Consider a bank of bandpass filters which intersect both the Planar and Gabor stimuli, and which are better matched to the Gabor stimuli than the Planar stimuli. The vector of energies across the bank of filters can be transformed to a new basis in which each of the energy measurements are independent. In the new energy bases, the Planar energy vector will have more independent samples but less energy per sample than the Gabor energy vector. We can also rotate these bases so that the vector components are equal. Because the probability correct for probability summation depends on the product of the detection probabilities of a set of independent samples and these probabilities are less than one, the probability correct for detecting Planar stimuli will be less than detecting Gabor stimuli. That is:

$$p(R_k = 1|Planar) = \prod_{i=1}^{n} p(R_k = 1|E_{Pl_i}) < \prod_{j=1}^{m} p(R_k = 1|E_{G_j}) = p(R_k = 1|Gabor) \tag{2.11}$$
$$\text{since: } p(R_k = 1|E_{Pl_i}) < p(R_k = 1|E_{G_j}) \text{ for } i = j, \text{ and } n > m$$

In eqn. 2.12, E_{X_i} is the energy along the ith independent energy coordinate for stimulus X after rotating to equalize the coordinate components, and n, m are the number of independent samples across the bank of filters for the Planar and Gabor stimuli. The formula leaves out the contributions of irrelevant samples, which is safe as long as the number of irrelevant samples included in the probability summation calculation is relatively constant between the Planar and Gabor stimuli. Thus, probability summation does not change the qualitative predictions. A bank of filters better matched to narrow band stimuli still predicts that Gabor stimuli should be more efficiently processed than Planar stimuli, which was not the case.

Planar stimuli could still be detected by an internal filter which matches the Planar stimuli better than the Gabor stimuli. This filter cannot be narrow band in orientation, hence it must be broad band in orientation. To serve as an alternate hypothesis to planar pooling, this filter must not be to be tuned to planar configurations of power. Thus the filter must be broadband in orientation and not tuned to any plane. This is essentially a description of cylindrical filter, which detects the planar stimuli on the basis of the spatial and temporal structure of the stimuli without using the motion information. This possibility

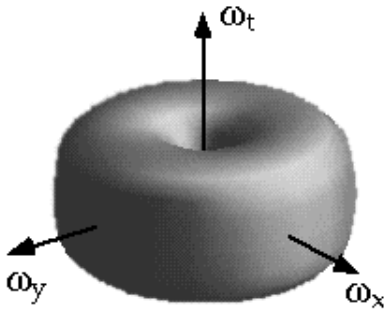

Figure 2.14: The 'Cylinder' filter, which detects Planar stimuli on the basis of spatial and temporal structure, without access to motion information. It is constructed by multiplying the spatial and temporal profiles shown in fig. 2.2. A lumpy background noise was constructed by passing spatio-temporal white noise through this filter.

is unlikely since the use of a cylindrical filter predicts equivalent performance on Planar and Scrambled stimuli, unlike the results.

2.6 Experiment 3

2.6.1 Planar vs. Scrambled, Cylindrical Background Noise

Although the results of the comparison between Planar and Scrambled stimuli suggest that observers do not use a strategy which only uses the spatial or temporal structure, there is still a possibility that the Scrambled stimuli were a particularly poor choice in non-planar comparison stimuli. For instance, it is possible that some inhibitory interactions exist which may serve to cancel parts of the Scrambled signal. In this case it is possible that subjects used a strategy that was not motion specific to detect Planar stimuli, but that this strategy resulted in a large loss of information on the Scrambled stimuli.

To test this possibility we used the following experimental logic. Let us assume that the Planar and Scrambled stimuli are both detected using only the spatial or temporal structure, with additional information being lost in the Scrambled case. This is similar to the observers using an internal filter which is selective to the spatial frequencies in the stimulus at all of the temporal frequencies present in the signal. The internal filter described is cylindrical and is shown in Figure 2.14. The cylindrical filter's amplitude spectrum is made by multiplying the spatial frequency spectrum of the Planar (or Scrambled) filter averaged across temporal frequency with the temporal frequency of the Planar filter averaged across spatial frequency, shown in fig. 2.2. In symbols:

$$Cyl(\vec{\omega}) = \left(\int_{\omega_t} Pl(\vec{\omega})d\omega_t\right) \cdot \left(\int_{\omega_x}\int_{\omega_y} Pl(\vec{\omega})d\omega_x d\omega_y\right) \tag{2.12}$$

We passed white noise through this filter to generate a lumpy dynamic background noise. Observers detected the Planar and Scrambled stimuli added to this lumpy background in two different conditions. In the first condition, the same background energy was used in both intervals of the 2AFC task (Signal Increment condition). In the second condition, the background energy in the noise alone interval is increased to match the signal plus noise energy (Constant energy condition). If subjects are only using the spatial or temporal structure of the Planar or Scrambled stimulus (i.e. a cylindrical internal filter), then the observer's performance is limited by the difference in energy between the two intervals. Thus the hypothesis predicts

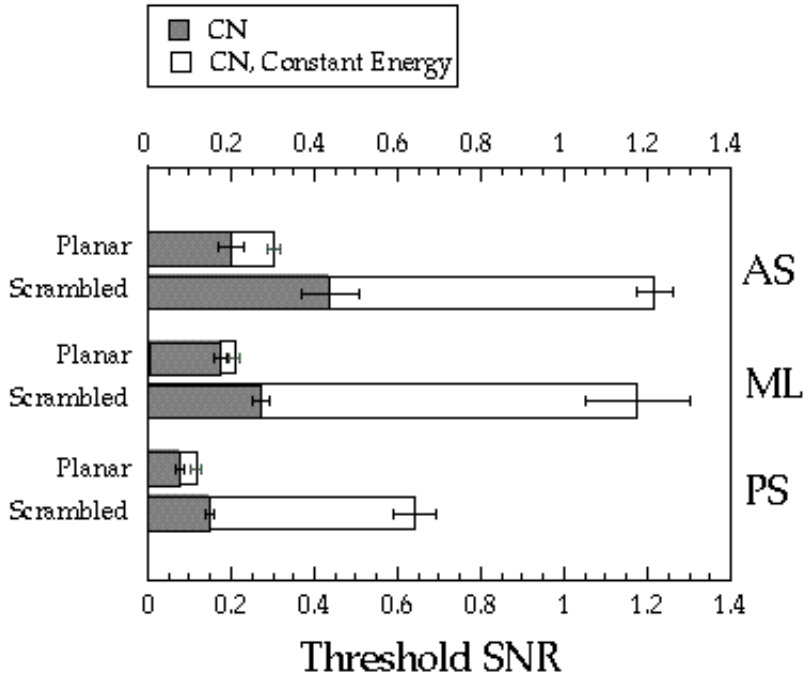

Figure 2.15: Subject performance on Planar and Scrambled stimuli added to Cylindrical background noise for two conditions: Constant Cylindrical noise power (CN) and constant total power in each interval (CN, constant total power).

that observers will be able to perform the task in the signal increment condition, but that performance will go to chance in the Constant energy condition.

Thresholds for 79% correct performance were determined using a transformed 3-down, 1-up staircase procedure [77]. Thresholds are computed as the mean of 20-30 reversals, and 4-5 thresholds were collected for each subject. The standard errors are given by the standard deviation of the thresholds[9]. The mean threshold SNR energies are shown for both conditions in Figure 2.15. Threshold energies for the Planar stimuli are higher for the signal increment condition by 16-50%. However, a difference in threshold is expected for a matched filter power detector. In the Constant energy condition, the difference in energy within the matched filter between the signal plus noise and noise intervals is reduced over the Signal Increment condition by 29% on average. This reduction is large enough to account for the observed increases in Planar thresholds.

Threshold energies for the Scrambled stimuli are much higher for the Constant energy condition, yet observer performance is still well above chance. This suggests that observers are able to use the spatio-temporal structure in the Planar case, but use something closer to an internal cylinder filter in detecting the Scrambled stimuli. However, the fact that performance is above chance on Scrambled stimuli indicates that subjects were able to use some of the spatio-temporal structure in the signal. This result corroborates the previous results which suggest that subjects can selectively and efficiently pool planar configurations of spectral power.

[9]Data for the Signal Increment, but not the Constant energy condition were also collected by the Method of Constant stimuli. The thresholds collected by both methods are quite similar.

2.7 Discussion

2.7.1 Interpretations

The relative efficiencies for observers on Planar, Component, and Scrambled stimuli are consistent with the idea that the visual system has mechanisms which efficiently pools across planes in spatio-temporal frequency space. The functional significance of this pooling is clear in light of the relationship between planes in spatio-temporal frequency space and translations of an image: the results support the idea that the visual system has planar pooling power detectors specialized for processing local translations of an image. This idea, of course, does not exhaust the list of possibilities. Any successful theory, however, must account for the high relative efficiency of Planar stimuli to Scrambled stimuli, and the nearly equal relative efficiency of Planar and Component stimuli. Since the number of comparison stimuli were limited, it is possible that a more complete set will reveal a different or more complicated set of conclusions.

While the relative efficiencies for the task suggest that the visual system is specialized for detecting planar configurations of power, the absolute efficiencies found for this task are lower than those found in other studies, which can be as high as 50-70% [15, 126, 23]. In the next section we discuss some of the possible sources of inefficiency in this study.

2.7.2 Sources of inefficiency

There are at least four distinct sources of inefficiency in processing the stochastic stimuli used here, which include internal filter spectrum/signal spectrum mismatch, generic sampling inefficiency, sensory noise, and the use of a suboptimal detection strategy (e.g., a non-linearity other than squaring, or a non-optimal decision rule such as probability summation). Analogous sources of inefficiency have been identified in previous studies [14, 15, 21, 65, 76, 22, 38]. Statements about the relative role of these sources of inefficiency cannot be made from the data shown here. We can, however, derive some upper bounds on the inefficiencies of these sources, which are presented below.

Since background noise powers were quite high it is likely that sensory/internal noise was 'swamped'[97], and thus played a small role in determining efficiencies. Signal-internal filter mismatch can result in two different kinds of inefficiency. If the internal filter bandwidth is smaller than the signal bandwidth, then the observer loses frequency samples in making the decisions. This is a kind of sampling inefficiency. On the other hand if the internal filter bandwidth is larger than the signal bandwidth, then more background noise is added to the decision variable which is a kind of additional internal noise. It is also possible for the filter to partially overlap the signal and the background noise to produce both types of inefficiency at once. The use of a suboptimal detection strategy can be functionally considered as a kind of sampling inefficiency, in that the available samples are not being optimally used. Since the none of the efficiencies were high and the Planar and Component efficiencies were nearly identical it is likely that generic subsampling, such as those due to the retinal mosaic and the temporal sampling of the signals, was a large source of observer inefficiency. It is easy to see how this could occur since the spatial size and temporal length of the stimuli may exceed the ability of the visual system to represent the signal. For instance, if temporal sampling errors discard half the frames, then efficiencies will be reduced by 50%. Thus the loss of efficiency could be attributed any combination of sampling inefficiencies and additional internal noises.

In figure 2.16, we illustrate the trade off between sampling efficiency and internal noise for the subjects on each of the three stimuli. Estimates of the equivalent input noise are plotted on the abscissa, to represent possible degradation due to internal noise. The equivalent input noise is measured as multiples of the background noise power. Estimates of percent samples used are plotted on the ordinate, representing the sources of sampling inefficiency.

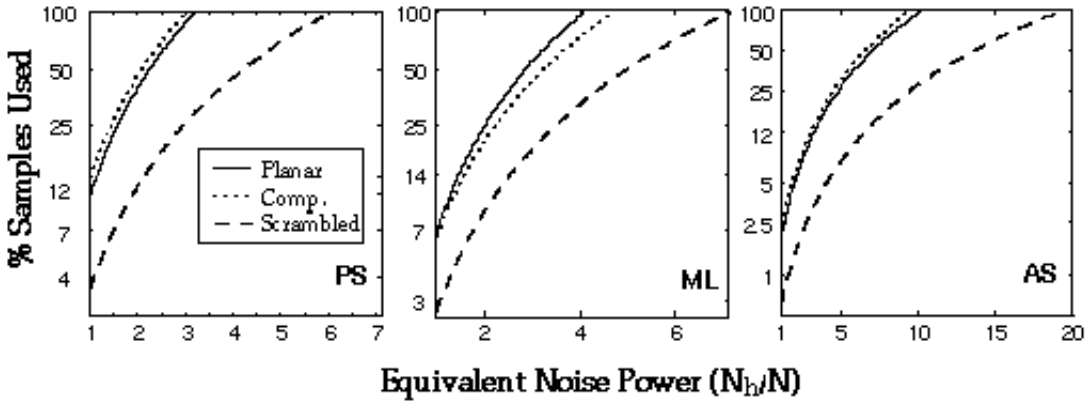

Figure 2.16: The tradeoff between information loss due to sampling inefficiency and internal noise is shown for the three subjects for the Planar, Component, and Scrambled stimuli. Internal noise is represented in units of the background noise power using estimates of the equivalent input noise.

2.7.3 Comparison to other detection studies

The current study is novel in that it considers motion detection for spatio-temporally filtered stimuli. Other motion detection studies have been performed, using much different stimuli. One set of studies most related to the work at hand is due to van Doorn and Koenderink [122, 121, 125], who had subjects detect rigidly translating binary noises added to spatio-temporal binary noise. They found signal to noise r.m.s. contrast ratios of about 0.05 for velocities similar to those used in this study. The threshold expressed in energy units is about $1x10^{-6}$ given the stimulus dimensions, which is quite low. The stimulus, however, is quite large (at 200 x 255 points) which means the ideal observer will be quite good as well. Assuming an integration time of about 150 ms, a rough estimate of their subjects' efficiencies is 0.23 %, which is substantially lower than the efficiencies presented here. However, due to the size of the stimulus, it is unlikely the visual system can use all of the information. If we assume that the subjects are only able to use about 10 frames and about half the 5.2 deg^2 spatial area of the original experiment, which yield stimuli with close to the number of independent samples in our stimuli, then efficiencies increase to about 6%, which is comparable to those found here.

In another set of detection studies, van Doorn and Koenderink [124] divided the display into a series of strips such that half the strips contained noise patterns rigidly translating downward interleaved with half the strips moving upwards. Detection SNR thresholds were measured for these stimuli over a range of different strip widths. They found that at small strip widths subject's detection performance was good and that the stimuli appeared transparent. The authors inferred from these results (and several others) that the visual system processes translations using a bank of filters tuned to image velocity whose properties vary across the retina [120, 119]. The results presented here are consistent with these conclusions and extend them by suggesting a particular form for the velocity filter.

A detection study by Watson and Turano[143] is of interest here. In the study the authors searched across the space of spatio-temporal Gabor stimuli for the stimulus which produced the lowest detection thresholds, i.e. the stimulus within the family that the visual system is most sensitive to when the contrast of the stimulus is dropped until observers can just identify the direction of motion (left-right). They found that the best stimulus did not have a Fourier spectrum best suited to process 1-D velocities (long axis along the velocity line), but instead was aligned with the Cartesian frequency axes, with the largest bandwidth along temporal frequency.

Watson and others have argued that this result constitutes evidence against the idea the visual system

has early mechanisms optimized for computing image velocities. However, Watson's experiments are not well suited to address the question. The ideal detectors for the Watson experiment are signal known exactly (SKE) matched filters instead of power detectors. Thus the subjects will perform best when using a set of linear filters best matched to the Gabor stimuli. Hence, it should come as no surprise that the filters found by Watson are similar to simple cells found in area V1, the last neural locus for which responses are reasonably linear[85].

Watson's results in fact underscore the importance of using stimuli suited to the experimental question. First, stimuli analogous to our Planar stimuli are not contained within the set of Gabor filters, since Gabors are narrow band in orientation. Second, our use of filtered noises forced the visual system to detect the stimuli with non-linear mechanisms. This choice was a conscience effort on our part to push the detection stage beyond V1 simple cells and into something potentially more interesting. The filter bandwidths inferred for the best linear detector need not correspond to the best filter bandwidths for power detectors.

Discussion of how the experimental results might be connected to neural processes is deferred until the end of Chapter 4.

2.8 Conclusions

The results show that observers can efficiently pool fourier power across planar regions of frequency space. In particular, observers are: 1) efficient at pooling spectral power across spatial orientation within a plane, 2) more efficient at detecting planar than non-planar configurations of power, 3) using spatio-temporal structure (i.e. motion) cues rather than spatial or temporal cues alone to detect the stimuli.

These results are consistent with the idea that observers have mechanisms which estimate local velocities using planar power detectors.

2.9 Derivation of Ideal Detector and Performance Approximations: for Signal in White Noise Case

The derivation of the ideal largely follows van Trees (1971), with an extension for unknown signal energy. In the 2AFC task, the subject is presented with two luminance distributions, the signal noise plus white noise and white noise alone, which are both samples from gaussian processes whose mean and covariance functions are known.

$$\begin{array}{lll} H_1 = \text{signal present:} & \mathbf{r}(x,y,t) & = a \cdot \mathbf{s}(x,y,t) + \mathbf{n}(x,y,t) \\ H_0 = \text{noise alone:} & \mathbf{r}(x,y,t) & = \mathbf{n}(x,y,t) \end{array}$$

The constant a determines the contrast of the signal noise $\mathbf{s}$, hence a^2 is proportional to signal energy. We will compute the ideal under two conditions, one in which the signal energy is known exactly on each trial, and one in which the signal energy is not known at all.

The Bayes decision for the 2AFC task is to choose the interval i with the larger likelihood ratio $L(\mathbf{r})_i$:

$$L(\mathbf{r})_1 \underset{2}{\overset{1}{\gtrless}} L(\mathbf{r})_2 \tag{2.13}$$

where the likelihood ratio is the ratio of the conditional probabilities of the the waveform $\mathbf{r}$ given signal present and noise alone conditions:

$$L(\mathbf{r}) = \frac{p(\mathbf{r}|H_1)}{p(\mathbf{r}|H_0)} \tag{2.14}$$

We use the fact that the stimuli are gaussian processes to write down the distribution functions explicitly. Since the mean functions are given by the background luminance, the processes are fully described by the covariance functions. The key step is to transform to a function space in which the signal process is uncorrelated. Since the signal is produced through the action of a linear shift invariant filter, the eigenfunctions of the covariance function are sinusoids. Thus by working in the fourier domain we may write down the distribution functions for equation 2. Let $\mathbf{R}(\omega_x, \omega_y, \omega_t)$ denote the the fourier transform of the signal $\mathbf{r}$. For the purposes of this experiment, R is discretized over a vector of frequencies.

The signal is a gaussian process produced by convolving spatio-temporal white noise by a linear shift-invariant filter, h(x,y,t). We denote the filter's amplitude spectrum by $|\mathbf{H}(\omega_x, \omega_y, \omega_t)|$. The filter excludes the DC term (zero mean) and is linear in phase. The resulting signal process has a zero mean function, and the transform of the covariance function is given by $|\mathbf{H}(\omega_x, \omega_y, \omega_t)|^2$. Since white noise has a flat power spectrum, the covariance of $\mathbf{R}(\omega_x, \omega_y, \omega_t)$ is given by:

$$\begin{array}{lrcl} \text{signal present:} & \mathbf{K}(\omega_x, \omega_y, \omega_t) & = & a^2 \cdot |\mathbf{H}(\omega_x, \omega_y, \omega_t)|^2 + \mathbf{N} \\ \text{noise alone:} & \mathbf{K}(\omega_x, \omega_y, \omega_t) & = & \mathbf{N} \end{array}$$

The distribution of the likelihood function is then:

$$\Lambda(\mathbf{R}) = \frac{\prod_{i=1}^{M} \frac{1}{[2\pi(a^2|\mathbf{H}(\vec{\omega_i})|^2+\mathbf{N})]^{0.5}} \exp(-0.5 \sum_{i=1}^{M} \frac{\mathbf{R_i R_i^*}}{(a^2|\mathbf{H}(\vec{\omega_i})|^2+\mathbf{N})})}{\prod_{i=1}^{M} \frac{1}{[2\pi\mathbf{N}]^{0.5}} \exp(-0.5 \sum_{i=1}^{M} \frac{\mathbf{R_i R_i^*}}{\mathbf{N}})} \tag{2.15}$$

Since monotonic transforms of the likelihood function do not change performance, we work with log likelihoods. Rewriting the equation gives an expression for the ideal receiver when the signal and noise levels are known.

$$\log \Lambda(\mathbf{R}) = \frac{1}{N} \sum_{i=1}^{M} \left(\frac{a^2 \cdot |\mathbf{H}(\omega_{x_i}, \omega_{y_i}, \omega_{t_i})|^2}{a^2 \cdot |\mathbf{H}(\omega_{x_i}, \omega_{y_i}, \omega_{t_i})|^2 + \mathbf{N}} \right) \mathbf{R}_i^2 + k \tag{2.16}$$

The optimal test given in equation 4.18 then becomes

$$\log \Lambda(\mathbf{R})_1 - \log \Lambda(\mathbf{R})_2 \underset{2}{\overset{1}{\gtrless}} 0 \tag{2.17}$$

From equation 2.16 we see the ideal receiver is a generalized Weiner filter with kernel given by

$$\frac{|\mathbf{H}(\omega_x, \omega_y, \omega_t)|^2}{a^2 \cdot |\mathbf{H}(\omega_x, \omega_y, \omega_t)|^2 + \mathbf{N}} \tag{2.18}$$

In the actual experiment, value of a^2 is randomly chosen each interval from one of six or seven values. The ideal can be adjusted to take the randomization into account, by averaging the likelihood function over the set of values of a^2, weighted by their probabilities:

$$\Lambda(\mathbf{R}) = \sum_{j=1}^{n} \Lambda(\mathbf{R}(a_j^2)) p(a_j^2) \tag{2.19}$$

This equation does not appear to produce a simpler receiver, and its performance is difficult to analyze analytically. Fortunately, an extremely good approximation exists. When $a^2 max\mathbf{H}(\vec{\omega_i}) << \mathbf{N}$, the ideal filter kernel reduces to

$$|\mathbf{H}(\omega_x, \omega_y, \omega_t)|^2 \tag{2.20}$$

which is just a filter matched to the expected power spectral density of the signal. This receiver does not use any information about the current signal level, while the first receiver had exact knowledge of the signal level. The performance of decision rule in equation 2.19 uses approximate knowledge of the signal level, hence its performance must lie in between the performance of receivers 2.18 and 2.20. We computed the performance of both the receivers having exact knowledge and no knowledge of the signal level for all the conditions of the experiment, and the results differed by less than 1%.

2.9.1 Performance

Next we derive approximate expressions for the performance of the ideal. Performance of the ideal only depends on the probability:

$$p(\log \Lambda(\mathbf{R}|H1) - \log \Lambda(\mathbf{R}|H0) > 0) \tag{2.21}$$

To find the performance of the ideal, we derive the distribution of the log likelihood on both the signal present and noise alone intervals. Note that each complex frequency sample $\mathbf{R_i}$ is a gaussian vector in which both the real and imaginary parts have identical distributions: $\mathrm{N}(0, \sigma^2/2)$, where σ^2 is given by:

$$\begin{array}{rll} H_1 = \text{signal present:} & \sigma^2 = a^2 \cdot |\mathbf{H}(\vec{\omega_i})|^2 + \mathbf{N} & \\ H_0 = \text{noise alone:} & \sigma^2 = \mathbf{N} & \end{array} \tag{2.22}$$

The distribution of $\mathbf{R_i R_i^*}/(\sqrt{2}\sigma)$ is chi-square with 2 degrees of freedom. Thus the log likelihoods are weighted sums of chi-square distributed random variables. Since the test statistic is the sum over a large number of samples, by the central limit theorem the statistic will be approximately normally distributed with the mean and variance given by the weighted sums of the mean and variance of the samples $\mathbf{R_i}$. Let $\mathbf{R_i}$ have a mean of zero and a variance of v_i. Then the mean and variance of $\mathbf{R_i R_i^*}$ are given by:

$$\begin{aligned} \mu_{R_i} &= 2v_i \\ \sigma^2_{R_i} &= 8v_i^2 \end{aligned} \tag{2.23}$$

Since equations 2.18 and 2.20 are simply weighted sums of $\mathbf{R_i R_i^*}$, using equation 2.22 we can compute the mean and variance of the test statistic:

$$\begin{aligned} \mu &= 2\sum_{j=1}^{M} K(\vec{\omega}_j) v_j \\ \sigma^2 &= 8\sum_{j=1}^{M} K(\vec{\omega}_j)^2 v_j^2 \end{aligned} \tag{2.24}$$

where $K(\vec{\omega}_j)^2$ are the receiver kernels. When the signal is present and the signal level known exactly (using the kernel in equation 2.18, these evaluate to:

$$\mu_{H1} = 2\sum_{j=1}^{M} \frac{|\mathbf{H}(\vec{\omega_j})|^2}{a^2 \cdot |\mathbf{H}(\vec{\omega_j})|^2 + \mathbf{N}} \cdot \left(a^2 \cdot |\mathbf{H}(\vec{\omega_j})|^2 + \mathbf{N}\right) \tag{2.25}$$

$$\begin{aligned}\mu_{H1} &= 2\sum_{j=1}^{M}|\mathbf{H}(\vec{\omega_j})|^2 \\ \sigma^2_{H1} &= 8\sum_{j=1}^{M}|\mathbf{H}(\vec{\omega_j})|^4\end{aligned}$$

on the noise alone condition:

$$\begin{aligned}\mu_{H0} &= 2\sum_{j=1}^{M}\frac{|\mathbf{H}(\vec{\omega_j})|^2}{a^2\cdot|\mathbf{H}(\vec{\omega_j})|^2+\mathbf{N}}\cdot(\mathbf{N}) \\ \sigma^2_{H0} &= 8\sum_{j=1}^{M}\left(\frac{|\mathbf{H}(\vec{\omega_j})|^2\mathbf{N}}{a^2\cdot|\mathbf{H}(\vec{\omega_j})|^2+\mathbf{N}}\cdot\right)^2\end{aligned} \tag{2.26}$$

The performance depends on the difference between the statistics on $H1$ and $H0$, which will also be normally distributed by the central limit theorem with mean and variance:

$$\begin{aligned}\mu &= \mu_{H1}-\mu_{H0} \\ \sigma^2 &= \sigma^2_{H1}+\sigma^2_{H0}\end{aligned} \tag{2.27}$$

Finally the probability correct is given by a cumulative normal function with μ and σ^2 as distribution parameters (using equation 2.21). Letting $\Phi(u,\mu,\sigma^2)$ denote the cumulative normal function integrated up to u, and the binary response by X_i, then the probability correct is given by:

$$p(X_i=1)=\Phi(0,\mu_{H1}-\mu_{H0},\sigma^2_{H1}+\sigma^2_{H0}) \tag{2.28}$$

The case for the matched power detector kernel is similar. Here we simply state the results. It is useful to write out the expression for probability correct in this case in terms of the signal, filter, and noise spectra and power. Let $\langle,\rangle$ represent the inner product of two vectors. The result depends on a set of inner products between spectra. Let $\mathbf{R}(\vec{\omega})$ denote the expected normalized signal spectrum. We define two kinds of inner product:

$$\begin{aligned}H^nR^m &= \langle|\mathbf{H}(\vec{\omega})|^n,|\mathbf{R}(\vec{\omega})|^m\rangle \\ H^n &= \langle|\mathbf{H}(\vec{\omega})|^{(n/2)},|\mathbf{H}(\vec{\omega})|^{(n/2)}\rangle\end{aligned} \tag{2.29}$$

where, for instance, H^2R^2 is the signal energy in the filter. Then the equivalent of equation 2.28 for the matched power detector is:

$$\begin{aligned}&\mu=\mu_{H1}-\mu_{H0}=2a^2H^2R^2 \\ &\sigma^2=\sigma^2_{H1}+\sigma^2_{H0}=8(a^4H^4R^4+2a^2H^4R^2\mathbf{N}+2H^4\mathbf{N}^2)\end{aligned} \tag{2.30}$$

2.9.2 Equivalent Flat Filter

The equivalent flat filter is a rectangular which produces performance equivalent to a non-rectangular filter. We will derive this by setting the efficiency of the two filters equal to 1. In Appendix B, we show that efficiency can be expressed as

$$\nu = \frac{\sigma_{I^2}}{\sigma_{Eq^2}} \tag{2.31}$$

where σ_{I^2} and σ_{Eq^2} are the variances for the ideal and equivalent flat filters respectively. Thus equal performance occurs when

$$\sigma_{I^2} = \sigma_{Eq^2} \tag{2.32}$$

σ_{I^2} is given by equation 2.31. For a flat filter, the filter spectrum $\mathbf{H}(\vec{\omega})$ is 1 within the pass band, and zero elsewhere. We will adjust the width of this pass band until performance is equal. From equation 2.31 we know that the performance of the filter only depends on the signal power a^2, noise power $\mathbf{N}$, and several sums of the signal and filter spectra raised to powers (equation 2.30). For a flat filter equation 2.30 becomes:

$$\begin{aligned} H^n R^m &= \langle |\mathbf{H}(\vec{\omega})|^n, |\mathbf{R}(\vec{\omega})|^m \rangle \\ H^n R^m &= \sum_{j=1}^{M} |\mathbf{H}(\vec{\omega}_j)|^n |\mathbf{R}(\vec{\omega}_j)|^m \\ H^n R^m &= \sum_{j=1}^{M_{Eq}} |\mathbf{R}(\vec{\omega}_j)|^m \\ H^n R^m &= M_{Eq} \frac{\sum_{j=1}^{M_{Eq}} |\mathbf{R}(\vec{\omega}_j)|^m}{M_{Eq}} \\ H^n R^m &= M_{Eq} |\bar{\mathbf{R}}|^m \end{aligned} \tag{2.33}$$

where $|\bar{\mathbf{R}}|^m$ is the mean signal spectrum and M_{Eq} is the number of non-zero frequencies in the flat filter. We now assume that we can change the number of non-zero frequencies without changing the value of the mean signal spectrum, which allows us to adjust M_{Eq} to find an equivalent filter:

$$\begin{aligned} \sigma_{I^2} &= \sigma_{Eq^2} \\ (a^4 H^4 R^4 + 2a^2 H^4 R^2 \mathbf{N} + 2H^4 \mathbf{N}^2) &= M_{Eq} \left(a^4 |\bar{\mathbf{R}}|^4 + 2a^2 |\bar{\mathbf{R}}|^2 \mathbf{N} + 2\mathbf{N}^2 \right) \end{aligned} \tag{2.34}$$

This last equation can be solved for M_{Eq}, which specifies the property of the flat filter we are interested in. Another way to express the equivalent flat filter is the number of chi-square random variables we need to sum together to produce a decision variable with a given variance. This expression has a natural interpretation in terms of sampling: M_{Eq} relates how many identical frequency samples are needed to produce a given level of performance. Although not obvious, M_{Eq} computed this way is always larger the number of frequencies the filter passes, and M_{Eq} is a slow varying non-linear function of the signal power. What the first property means is that to really get equivalent performance with a flat filter, we would have to use a stimulus with more samples.

2.9.3 Equivalent Input Noise

The equivalent input noise is the result of referring the observer's internal noise to the input. We express the equivalent noise as the increase in background noise power needed for the ideal observer to have the same performance as the subject. Using a gaussian approximation for the subject's decision variable, the performance is determined by the variance of the decision variable. Thus, we need to determine the increase in background noise power needed for an ideal observer's decision variable variance to be equal to the subject's variance:

$$\sigma_h^2 \;=\; \sigma_I^2(\mathbf{N}) + \sigma_{internal}^2 \;=\; \sigma_I^2(\mathbf{N} + \Delta\mathbf{N}) \tag{2.35}$$

Where the subject's and ideal's decision variable variances are denoted by σ_h^2 and $\sigma_I^2(\mathbf{N})$ respectively, and $\mathbf{N}$ is the background noise power.

We write down the ideal's variance as a function of $\mathbf{N}$ from equation (2.31),

$$\sigma_I^2(\mathbf{N}) = 8(a^4H^4R^4 + 2a^2H^4R^2\mathbf{N} + 2H^4\mathbf{N}^2) \tag{2.36}$$

Substituting this expression into equation (2.35), we see that the expression is quadratic in $\mathbf{N}$:

$$\sigma_h^2 = \mathrm{a}\mathbf{N}^2 + \mathrm{b}(a_h^2)\mathbf{N} + \mathrm{c}(a_h^2) \tag{2.37}$$

where $\mathrm{a} = 16H^4$, $\mathrm{b} = 16a^2H^4R^2$, $\mathrm{c} = 8a^4H^4R^4$, and a_h^2 denotes the subject's threshold signal power.

In Appendix B, we show how σ_h^2 can be estimated from the subject's threshold. Using this estimate, equation (2.37) can be solved for $\mathbf{N}_h$, and $\Delta\mathbf{N}$ determined. Standard errors for the estimate of $\Delta\mathbf{N}$ were obtained by propagating the error in the estimate of the subject's threshold through the calculation using a Monte Carlo method. A thousand threshold samples were taken from the gaussian approximation to the subject's threshold likelihood function. Equivalent input noises were then computed for each threshold sample, from which the standard error on the equivalent input noise estimate could be computed.

2.10 Appendix B: Efficiency for the Detection of Signal Noises in White Noise

The definition of efficiency comes from information theory, in which it expresses the amount of information used by the human observer relative to the ideal. Let $p_I(s,n)$ represent the probability density of the 2AFC decision variable X_I for the ideal observer and $p_h(s,n)$ of X_h for the human observer. The difference in entropies between these two distributions is a measure of the information used by the human observer. In particular, efficiencies are defined such that

$$log(\nu) \;\;=\;\; \mathrm{H}(X_I) - \mathrm{H}(X_h) \tag{2.38}$$

where $\mathrm{H}(X)$ is the entropy of the random variable X. If X is normally distributed, the entropy is given by $ln(2\pi) + ln(\sigma^2)$. Thus assuming both X_h and X_I can be reasonably approximated by a normal distribution, the efficiency can be expressed as:

$$\nu = \frac{\sigma_{I^2}}{\sigma_{h^2}} \tag{2.39}$$

This definition is an information theoretic interpretation of the standard definition for efficiency of the ratio of ideal to human d'^2. To see this, note that in a generic signal detection model, $d' = s/\sigma$, where s is the signal strength. When the human observer can be modeled as being limited only by the noise in the stimulus and some independently added internal noise, $d'_h = s/\sqrt{\sigma^2 + \sigma_{internal}^2}$. For fixed signal strength s,

$$\nu = \frac{d'^2_h}{d'^2_I} \;=\; \frac{\sigma_I^2}{\sigma_h^2} \;=\; \frac{\sigma_I^2}{\sigma_I^2 + \sigma_{internal}^2} \tag{2.40}$$

The standard approach of computing efficiencies the ratio of squared ideal and observer thresholds can be interpreted in light of this equation. At a given % correct, the signal threshold will be proportional to the decision variable standard deviation, assuming a gaussian decision variable. If the ideal standard deviation does not vary with signal level, then the ratio of the squared thresholds is computing the same quantity as equation (2.40). Because the ideal standard deviation *does* vary with signal level, we cannot use the ratio of squared thresholds to compute efficiencies.

If we assume that human observer's decision variable is approximately normal, then we may use equation (2.39) to construct a measure of efficiency. Since the variance is a function of signal energy, we would underestimate human efficiency if we compared the ideal's threshold to the human observer's threshold, since the ideal observer's threshold is caused by a smaller decision variable variance. What we do instead is to compute the ideal observer's decision variable variance *at the subject's threshold signal energy* which supplies the numerator in equation (2.39). The denominator variance can be estimated assuming the subject has a normally distributed decision variable. Then the subject's variance can be computed from the threshold using the expression for d':

$$\sigma_{h^2} = \mu_{threshold}/{d'}^2_{80\%} \tag{2.41}$$

where $\mu_{threshold}$ is the mean of the ideal's decision variable at the subject's threshold, and ${d'}_{80\%}$ is the value of d' for 80% correct: 1.190233. From Appendix A we showed that the mean of the ideal is proportional to the signal energy within the ideal filter, $\mu_{ideal} = 2a^2\langle|\mathbf{H}(\vec{\omega})|^2, |\mathbf{S}(\vec{\omega})|^2\rangle$, where $\mathbf{H}(\vec{\omega})$ is the ideal filter amplitude spectrum, and $\mathbf{S}(\vec{\omega})$ is the signal spectrum, which are the same, of course. Plugging this in along with an expression for the ideal variance into equation (2.39) gives an expression for the efficiency:

$$\nu = \frac{2c(a_t^4\langle|\mathbf{H}(\vec{\omega})|^4, |\mathbf{S}(\vec{\omega})|^4\rangle + 2a_t^2\mathbf{N}\langle|\mathbf{H}(\vec{\omega})|^4, |\mathbf{S}(\vec{\omega})|^2\rangle + 2\langle|\mathbf{H}(\vec{\omega})|^4, \mathbf{N}^2\rangle)}{(a_t^2\langle|\mathbf{H}(\vec{\omega})|^2, |\mathbf{S}(\vec{\omega})|^2\rangle)^2}$$

where a_t is the subject's threshold energy, and $c = {d'}^2$. This expression has the usual interpretation of the number of samples used by the human observer, which can be shown most easily by replacing the filter spectra with their equivalent flat filters. First, however, we need a sampling interpretation of the estimate of the subject's decision variable variance. We reexpress the subject's variance estimate as the variance predicted by an equivalent flat filter:

$$a_t^2\langle|\mathbf{H}(\vec{\omega})|^2, |\mathbf{S}(\vec{\omega})|^2\rangle)^2/(2c) = M_h\left(a^4|\bar{\mathbf{S}}|^4 + 2a^2|\bar{\mathbf{S}}|^2\mathbf{N} + 2\mathbf{N}^2\right) \tag{2.42}$$

Substituting this expression and the expression for the variance of the equivalent flat filter of the ideal (equation (2.35) into equation (2.42), we get:

$$\begin{aligned} \nu &= \frac{M_I\left(a^4|\bar{\mathbf{S}}|^4 + 2a^2|\bar{\mathbf{S}}|^2\mathbf{N} + 2\mathbf{N}^2\right)}{M_h\left(a^4|\bar{\mathbf{S}}|^4 + 2a^2|\bar{\mathbf{S}}|^2\mathbf{N} + 2\mathbf{N}^2\right)} \\ \nu &= \frac{M_I}{M_h} \end{aligned} \tag{2.43}$$

Note that in the expression $M_h > M_I$ which means that M_h has the interpretation of the additional effective samples needed for the human observer to achieve the same performance as the ideal.

2.11 Appendix C: Performance prediction calculations

In this section we explain how the performance predictions were generated. The component prediction was generated by assuming the observer could monitor only one of the component filters which make up the stimuli. In addition we assume that the decision based on the output of this filter is corrupted by the same amount of internal noise as was estimated from subject's performance on the Component stimuli. Let the bandpass filter be denoted by $|\mathbf{B}(\vec{\omega})|^2$ and the signal spectrum by $|\mathbf{S}(\vec{\omega})|^2$. Modeling the effect of observer internal noise as an equivalent input noise $\mathbf{N_h}$, the output of the filter can be expressed: the bandpass filter

$$\begin{aligned} \text{on } H1\text{: } E^{H1} &= \sum_{j=1}^{M} |\mathbf{B}(\vec{\omega}_j)|^2 \left(a^2 |\mathbf{S}(\vec{\omega}_j)|^2 + \mathbf{N_h}\right) \\ \text{on } H0\text{: } E^{H0} &= \sum_{j=1}^{M} |\mathbf{B}(\vec{\omega}_j)|^2 \left(\mathbf{N_h}\right) \end{aligned} \tag{2.44}$$

The equivalent input noise $\mathbf{N_h}$ is estimated from each observer's performance on the Component stimuli as outlined in Appendix A. Using these equations, expressions for the mean and variance of the decision variable similar to those in equation (2.31) in Appendix A can be derived, from which probability correct and the 80% thresholds were computed. Standard errors for the threshold were computed using the thousand Monte Carlo equivalent input noise samples, whose generation is described in Appendix A.

2.11.1 Probability summation

In the probability summation calculation, we used a decision process which computes the maximum output of a set of bandpass filters and chooses the interval with the larger maximum [96]. We modeled each bandpass filter as above. Let the ith bandpass filter be denoted by $|\mathbf{B}_i(\vec{\omega})|^2$ and the signal spectrum by $|\mathbf{S}(\vec{\omega})|^2$. The output of each filter is:

$$\begin{aligned} \text{on } H1\text{: } E_i &= \sum_{j=1}^{M} |\mathbf{B}_i(\vec{\omega}_j)|^2 \left(a^2 |\mathbf{S}(\vec{\omega}_j)|^2 + \mathbf{N_h}\right) \\ \text{on } H0\text{: } E_i &= \sum_{j=1}^{M} |\mathbf{B}_i(\vec{\omega}_j)|^2 \left(\mathbf{N_h}\right) \end{aligned} \tag{2.45}$$

On each interval, E_i is approximately gaussian, with mean and variance given by equation (2.31) in Appendix A. However, because of filter overlap the E_i are correlated, which needs to be taken into account. This can be accomplished by computing the covariance matrix $\mathbf{C}$ of the E_i, which has elements ij given by $\mathrm{E}\left[E_i E_j\right]$:

$$\begin{aligned} \text{on } H1\text{: } \mathbf{C}_{ij}^{H1} &= \mathrm{E}\left[E_i E_j\right] = \sum_{k=1}^{M} |\mathbf{B}_i(\vec{\omega}_k)|^2 |\mathbf{B}_j(\vec{\omega}_k)|^2 \left(a^2 |\mathbf{S}(\vec{\omega}_k)|^2 + \mathbf{N_h}\right) \\ \text{on } H0\text{: } \mathbf{C}_{ij}^{H0} &= \mathrm{E}\left[E_i E_j\right] = \sum_{k=1}^{M} |\mathbf{B}_i(\vec{\omega}_k)|^2 |\mathbf{B}_j(\vec{\omega}_k)|^2 \left(\mathbf{N_h}\right) \end{aligned} \tag{2.46}$$

Thus, the outputs E_i of the bandpass filters can be treated as multi-dimensional gaussian vectors whose distributions have means $\vec{E}^{H_i}$ and covariances $\mathbf{C}^{H_i}$ on $H1$ and $H0$. Performance is determined by the

probability $p\left(\max_{E_i}(\vec{E}^{H_1}) > \max_{E_i}(\vec{E}^{H_0})\right)$. A Monte Carlo method was used to simulate performance. Ten thousand samples from the multi-dimensional gaussian distributions were taken for both intervals, and responses were generated using by finding the interval with the largest output. Percent correct was determined for 100 different signal power levels, from which the 80% thresholds were determined. Standard errors were computed as above.

2.12 Appendix D: Finding optimal Gabor filters for Planar stimuli

In this section we describe how we determined the parameters of the Gabor filter which would yield the best performance in the presence of the planar signal.

Gabor filters have been frequently used in models of early visual processing. The one-sided amplitude spectrum of Gabor filters in 3-D $(\omega_x, \omega_y, \omega_t)$ frequency domain can be written as:

$$G(\omega_x, \omega_y, \omega_t) = exp(-\frac{1}{2}((\vec{\omega} - \vec{\omega_0})^T \Lambda^{-1} (\vec{\omega} - \vec{\omega_0})) \tag{2.47}$$

Λ was restricted to be a diagonal matrix with the squared ω_x, ω_y, and ω_t bandwidths on the diagonal, since Watson and Turano (1994) found that the optimal Gabor stimulus had zero off-diagonal elements. This leaves 6 parameters to be fit, the bandwidths and the shift vector $\vec{\omega_0}$.

The optimal parameters were found by minimizing the threshold planar signal energy required for the a model detector using a Gabor filter over the 6 dimensional parameter space. The threshold planar signal energy can be found from the expression for d'^2. The function we minimized is the result of solving for signal energy for a fixed d' in terms of the filter parameters. We fixed d' to give a percent correct of 80%.

Let $|Pl(\vec{\omega})|^2$ denote the expected power spectrum of the planar signal and $|G(\vec{\omega})|^2$ the power spectrum of the Gabor filter. In Appendix B, we showed that:

$$d'^2 = \frac{s^2 \langle |G(\vec{\omega})|^2, |Pl(\vec{\omega})|^2 \rangle}{2s^4 \langle |G(\vec{\omega})|^4, |Pl(\vec{\omega})|^4 \rangle + 2s^2 N^2 \langle |G(\vec{\omega})|^4, |Pl(\vec{\omega})|^2 \rangle + N^4 \langle |G(\vec{\omega})|^2, |G(\vec{\omega})|^2 \rangle} \tag{2.48}$$

This equation is a quadratic in s^2:

$$s^2 = \frac{-b - \sqrt{b^2 - 4ac}}{2a} \tag{2.49}$$

Where

$$a = 2d'^2 \langle |G(\vec{\omega})|^4, |Pl(\vec{\omega})|^4 \rangle - (\langle |G(\vec{\omega})|^2, |Pl(\vec{\omega})|^2 \rangle)^2 \tag{2.50}$$

$$b = 2N^2 \langle |G(\vec{\omega})|^4, |Pl(\vec{\omega})|^2 \rangle \tag{2.51}$$

$$c = 2N^4 \langle |G(\vec{\omega})|^2, |G(\vec{\omega})|^2 \rangle \tag{2.52}$$

The optimal bandwidths are (σ_{ω_x}, σ_{ω_y}, σ_{ω_t}) =(2.4 cyc/deg, 6.2 cyc/deg, 5.0 Hz), centered at (ω_x, ω_y, ω_t) = (2.73 cyc/deg, 0 cyc/deg, 5.3 Hz). The resulting filter is shown in fig. 2.11.

Chapter 3

Perturbation Analysis

3.1 Introduction

Since the first filter based models of local motion detection were introduced [142, 139, 1], the idea that the visual system uses a set of spatio-temporal bandpass filters to analyze visual motion information has gained a great deal of support (e.g. [1, 11, 24, 13, 12, 7, 54, 136, 83, 85, 143]). What characterizes these filters is that their frequency selectivity is concentrated around a single spatio-temporal frequency [138]. Several interesting kinds of retinal motion can be encoded by selectively pooling the outputs of these filters, and virtually every motion processing model can be expressed in terms of the outputs of these filters [112]. For instance, the velocity of a translating pattern can be encoded by pooling the filters whose peak frequencies lie in a common plane [138, 56, 53, 112], 'opponent' motion by subtractively pooling filters tuned for opposite directions [127], 'group velocity' by pooling along contours orthogonal to a plane through the origin [47], and gradients in the motion field can be detected by subtractively pooling estimates of retinal velocity across adjacent spatial regions of the retina [141]. In addition, a large class of phenomena have been modeled by allowing nonlinear interactions between the outputs of spatio-temporal filters, including contrast normalization [78, 57], motion induction [94, 66], motion contrast [60], motion transparency [103], and trajectory detection [134], among others.

Each type of filter output pooling corresponds to a strategy for encoding information which is optimal for a certain task. For instance, planar pooling is optimal for detecting translations and 'opponent motion' pooling for discriminating opposite moving gratings. The results of the experiments in chapter 2 (particularly for 'Scrambled' stimuli) showed that subjects are not able to adopt arbitrary detection strategies. Thus the pooling strategies used by the visual system tell us something about the tasks the visual system is optimized to perform.

In this chapter we examine two related issues: What pooling strategies do subjects use in detecting local motions, and How adaptable are the visual system's pooling strategies to the demands of a task? We investigate these questions by focusing on how the visual system combines frequency information to detect four different stimulus types: the Planar, Scrambled, and Component stimuli used in chapter 2, plus one new stimulus type described below. Optimal detection performance on each stimulus type requires a qualitatively different pooling strategy.

We perform a kind of perturbation analysis on the detection data which allows us to infer how subjects weight a set of frequency bands when detecting the stimuli. The analysis we used is similar to several other cue integration studies in vision [151, 67] and audition [4, 48, 105] in that it relies on cue perturbations to derive the weights or relative contribution of the cues. We adapted an approach used by Knill [67] to analyze the data from the last chapter. The relative weight or contribution that each frequency band made

to detection performance was estimated by fitting a linear combination of the energies within the bands to the variations in the subject's trial by trial responses using a psychometric model.

The pattern of weights revealed by the analysis will be compared across stimuli type and also to the ideal weighting strategy for the signal. Both comparisons will be used to infer how well subjects can adjust their frequency pooling strategy to match the properties of the signal. The measured weights will also allow us to make some general inferences about the kinds of tasks the motion system is optimized for.

The rest of the chapter is divided into several sections. First we describe possible weighting strategies the visual system might use to detect the stimuli. Next we describe the weight estimation procedure, its assumptions and conditions for validity. Finally the resulting weights for the four different stimuli are presented and interpreted in terms of the detection strategies consistent with the results.

3.2 Detection Strategies

In the experiments described in this thesis, the information which supports successful detection is contained in the differences in spectral power between the signal plus noise and the noise alone intervals. Subjects could be using any viable detection strategy which makes use of these differences. Rather than attempting to describe every possible strategy in detail, we will restrict our analysis of strategy to the linear weighting of different regions (bands) of spatio-temporal frequency. Because the information for the task is contained in differences in power within frequency bands across the two intervals, the main determinate of performance is the weighting of these bands (e.g. what bands are discarded, what irrelevant bands are included, such as bands without any signal power, and what bands are inappropriately weighted). Concentrating on the linear weighting of frequency space is similar to estimating the first order kernel of a system: it constitutes a first order approximation to the decision strategy employed by the visual system (see Appendix A).

3.2.1 Frequency Decomposition

In order to assign weights to different regions of frequency space, we decomposed frequency space into a set of 13 non-overlapping bands arranged on a sphere. The decomposition is illustrated in figure 3.1. The choice of decomposition was subject to several constraints. The most important constraints were to ensure that the signal spectra lied completely within a small number of bands and that the weights would be interpretable in terms of some general hypotheses about detection strategies outlined below. The number of bands was determined by a trade off between the reliability of the weight estimates (which requires a smaller number of bands) and reducing biases in the weights due to discretizing the subject's spectral weighting function (which requires a larger number of bands)[1]. The compromise of 13 bands is justified in the discussion.

Because most of the signal spectra were concentrated around a single plane in frequency space, the decomposition was centered around this plane. In the figure, the plane is shown in black. The equator of the frequency sphere was chosen to lie in the plane, so that the signal spectra (except Scrambled) are completely contained in bands 1-6. Details of the construction of the filters are contained in Appendix C. Next we discuss how to interpret the bands in the sphere.

Because of the spherical shape of the decomposition, each band is most easily described in terms of the orientations of the set of lines through the origin in frequency space which intersect the band (i.e. in terms of ω_θ and ω_ϕ). Each line through the origin of frequency space represents the set of gratings which have the same spatial orientation and speed. Thus the spherical bands primarily represent a range

[1]The biases induced by using too few bands are simple and predictable, constituting an average over the subject's spectral weighting profile within a band.

Bands	
1-6:	In plane
7-11:	Opponent direction
12-13:	High Temp Freq.

Bands	Spatial Orientation	Temp. Frequency
1	15	+ low
2	45	+ low/mid
3	75	+ mid
4	105	+ mid
5	135	+ low/mid
6	165	+ low
7	60	- low
8	0	- mid
9	90	- mid
10	180	- mid
11	120	- low
12	0	+/- high
13	0	+/- high

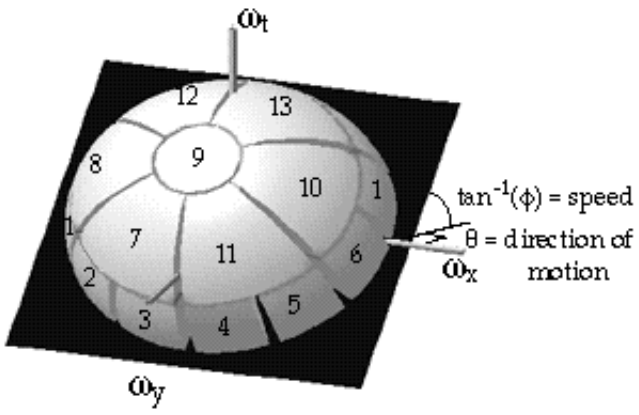

Figure 3.1: Decomposition of spatio-temporal frequency space into 13 non-overlapping bands. The black plane corresponds to the image velocity the stimuli were designed around. θ and ϕ in the diagram refer to the slant and tilt of the plane, from which the direction and speed of the velocity can be computed. Each of the bands is assigned a number which is shown on the visible area of the band. On the left hand are two tables which give some of the properties of the frequencies within each of the bands. The top table divides the 13 bands into 3 groups with qualitatively similar properties. The bottom table gives the mean spatial orientation and temporal frequency range of the gratings within each band. The '+' ('-') denotes gratings which have the same (opposite) direction of motion as the velocity specified by the plane.

of grating orientations and speeds. Each band also includes a range of spatial frequency magnitudes k, however there is less spatial frequency selectivity between bands since the radial extent of the sphere is large. Because grating speed is given by tf/k, where tf is the temporal frequency, the various speeds are largely determined by the range of temporal frequencies within the band. Thus each band can be characterized by the range of orientations and temporal frequencies contained in the band.

The mean spatial orientation and temporal frequency range of each of the bands is listed in the table in figure 3.1. The six bands in the plane represent the frequencies which are consistent with downward translational motion[2], with each of the six having different ranges of spatial orientations and temporal frequencies. Bands 7-11 represent frequencies which intersect planes corresponding to translations in the opposite direction (upward) as the first 6 bands. We refer to these bands as *opponent*. Bands 12 and 13 surround the temporal frequency axis. Thus these bands do not have a specific directionality attached to them, since they include gratings moving in all directions at high temporal frequencies. The pattern of positive and negative weights across these bands, given the signal, can tell us a great deal about the detection strategy used by the subject.

To help the reader to understand how different strategies for detecting the stimuli might appear in the estimated weights, we will outline the predictions made by a number of possibilities.

3.2.2 Ideal strategies

The stimuli we used are ideally detected by power detectors whose spectrum is matched to the signal's spectrum. Thus, the ideal detector positively weights the bands containing the signal and sets the rest of the bands to zero. The weights which we would find were an observer using the ideal strategy will be designated the 'ideal weights'. The ideal weights for the stimuli, given our choice of frequency partition, are shown in figure 3.2. The importance of the ideal weights is that they represent a benchmark against which all strategies can be compared. To the extent that the weights estimated for subjects resemble the ideal weights, we can infer that the filtering properties of the visual system are well suited to the structure of the signal. The salient properties of the weights for each of the four stimuli are: Component stimuli are concentrated around a single spatial orientation (bands 3 & 4), Planar stimuli are broad band in spatial orientation (bands 1-6), Scrambled stimuli intersect most of the bands except 12 & 13, with the largest weights for the horizontally oriented, low temporal frequency bands (1 & 6), and the Plaid stimuli are concentrated around bands 1 & 6.

3.2.3 Subideal strategies

Probability Summation Pooling

Probability summation predicts weights similar to ideal pooling. This is true because the estimation of weights does not require an additive law (see Appendix A). If the visual system's bandpass channels are smaller than the signal spectrum, then probability summation across the bands intersecting the signal spectrum should lead to the exclusive weighting of the signal bands, like the ideal strategy. If the visual system uses probability summation across a set of filters which include frequencies outside the signal bands, then we should observe positive weights in bands close to the signal bands. In general, we will not be able to distinguish between rival pooling rules using the same set of filters on the basis of the weights alone.

[2]These bands actually represent the frequencies consistent with a small range of speeds and directions due to the fact that the bands include many planes other than the black plane. To visualize this range, imagine the set of planes passing through the origin which only intersect bands 1-6.

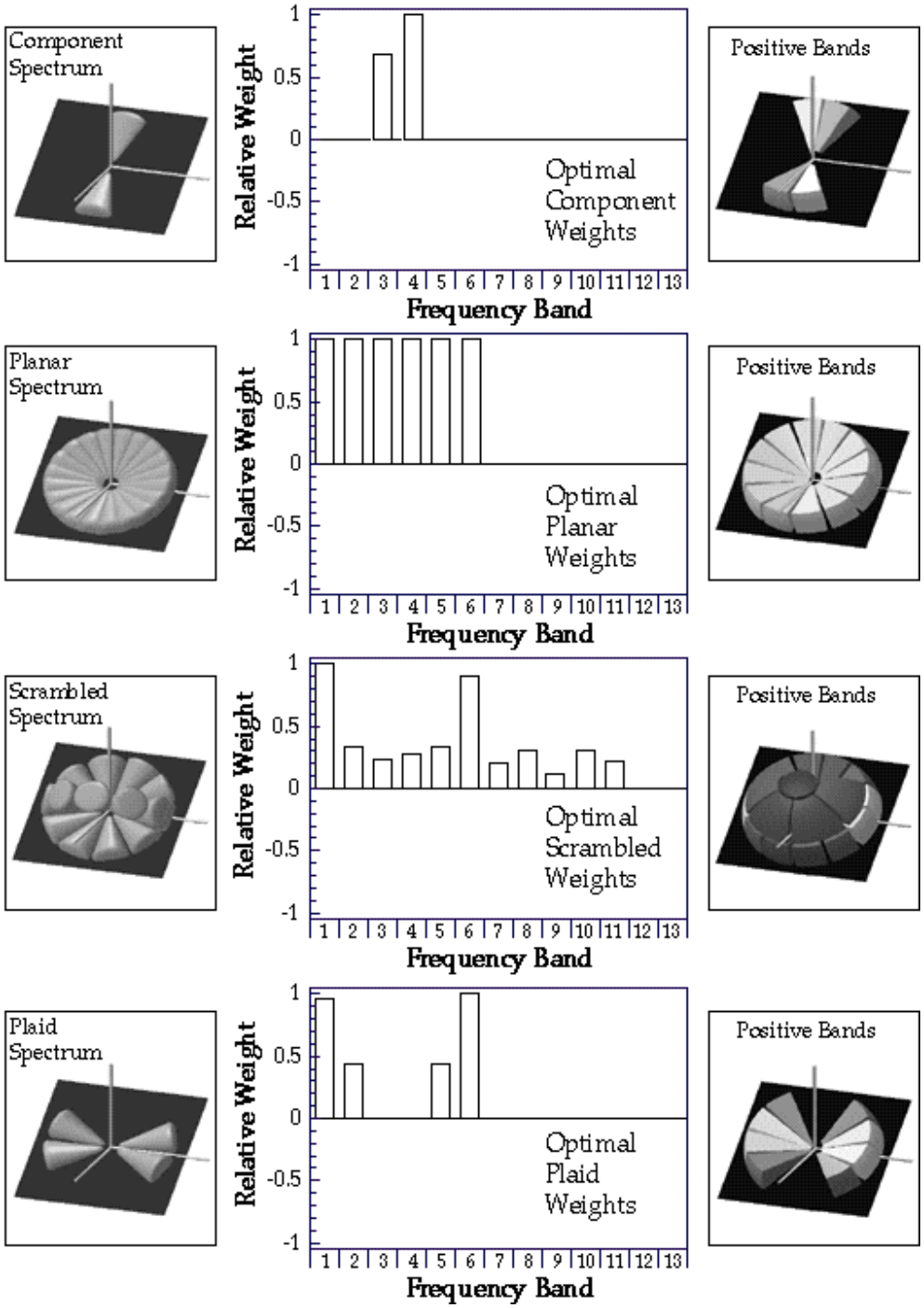

Figure 3.2: Ideal weights for detecting the four different signal types. **Left column:** The pictures depict the 65% level sets of the average power spectral density of the four different stimuli. **Center column:** Bar graphs depict the optimal weights computed using the ideal observer for each stimulus. Because the ideal observers are matched power detectors, the weights reflect the overlap between the frequency bands in figure 3.1 and the signal spectra. **Right column:** The frequency bands corresponding to the non-zero weights are displayed using grayscale intensity to encode the magnitude of the weights.

Spatial structure detectors

Since all the stimuli have a bandpass spatial appearance, it is possible to detect these stimuli solely on the basis of the difference in spatial appearance between the two intervals, ignoring the phenomenal motion. An observer detecting the stimuli on the basis of the spatial appearance would essentially be comparing the average spatial appearance across the two intervals. This amounts to saying that the observer applies a filter which has the same spatial frequency structure as the time averaged stimulus at each instant in time. In the spatio-temporal frequency domain, these filters can be constructed by projecting the signal spectrum onto the spatial frequency plane, and then copying this projection across temporal frequency. Evidence for this strategy would be the same positive weights attached to all the frequencies with the same spatial orientation as the signal. For instance if a band oriented near 0 deg receives a positive band, then all the bands near that orientation should also receive positive weights (e.g. if band 9 receives positive weight, then so should bands 3,4,7, & 11).

Temporal structure detectors

The stimuli also have a bandpass temporal structure, which means that the stimuli could be detected on the basis of the difference in flicker between the two intervals. A fourier domain description of this strategy is to combine across all spatial frequencies within the temporal frequency range covered by the signal. Evidence for this strategy would be positive weights attached to all the bands except the high temporal frequency bands 12 and 13.

Discrimination strategies and Motion Opponency

It is possible that the motion system is not optimized for detection, but rather is optimized for discrimination. Optimal discriminators subtract the signals which are to be discriminated across the two intervals, such that these models predict negative weights. Two plausible discrimination models are motion opponent and velocity discriminators. Motion opponency has a long history as an integral part of many motion processing models [104, 127, 1, 62]. The characteristic feature of these models is that the outputs of spatio-temporal filters tuned for gratings drifting in opposite directions are subtractively combined. In terms of the spherical frequency decomposition, this sort of model predicts that negative weights will be attached to the bands which are reflections around the temporal frequency axis of bands receiving positive weights. For example if band 3 receives a positive weight, we would expect negative weights on bands 7 and/or 9.

There is a great deal of evidence which supports some kind of inhibitory interaction between motion in opposite directions, for instance the lack of motion of counterphase gratings [1], certain aftereffects of motion[131], and the result that leftward and rightward moving gratings can be rendered undetectable in the presence of a suprathreshold mask [62]. However, none of these effects require subtractive interaction, and using an extensive set of measurements Lubin [78] showed that the detectability of contrast increments to sums of gratings moving in opposite directions was better fit by a divisive interaction. The current methodology is not designed to distinguish between these possibilities, and divisive interactions are discussed below.

Another possibility is that the motion system is naturally optimized to discriminate between motions in the image. This hypothesis is motivated by the observation that thresholds for detection and discrimination are comparable [20]. In the context of a motion detection experiment, the natural discrimination would be between images coherently moving and those which are essentially stationary. This discrimination could be implemented by subtracting a 'stationary signal', i.e. all the frequencies which have temporal frequencies close to zero. In terms of the decomposition, negative weights would be expected for bands 1,6,7 & 11.

Normalization, Denoising, and Predictive Coding

This subsection describes a collection of possibilities involving divisive interactions between the frequency bands which are suggested by several different processing demands. A general non-linear model which appears in several contexts divides the contrast energy in each band by a weighted sum of the contrast energies in the other bands:

$$E_{norm_i} = \frac{E_{BP_i}}{\sum_j w_j E_{BP_j}} \tag{3.1}$$

The most common name for this kind of model is contrast normalization, and is motivated by an appeal to the limited response range of cortical neurons, which requires a gain control mechanism to maintain the visual system's responsiveness. Contrast normalization has been used to model psychophysical results [78] and the contrast responsiveness of cortical neurons [57]. Although normalization is typically motivated by an appeal to the processing constraints imposed by the response properties of neurons, divisive interactions can also be motivated from a purely information processing perspective. Contrast normalization can be shown to be a particular instance of a more general strategy for signal whitening and denoiseing. In addition, a model which essentially computes contrast normalization has been presented in the context of predictive coding [33], in which the divisive step is used to compute the predictive carrier band.

Typically in normalization models $w_j = 1$ for all j, so that the energies in each band are normalized by an estimate of the total energy in the stimulus. Since division will produce negative weights in our analysis, normalization should produce equal negative weights for all the bands excluding the signal. However, when some of the bands are more reliable estimators of the background noise or better predictors of the carrier location than other bands, then we should not expect all the weights to be equal. If we allow for unequal w_j, then the only steadfast prediction is the presence of *some* negative weights outside the signal bands.

3.3 Perturbation Analysis

The typical method used to infer task strategy is to perform a large number of threshold comparisons between different stimuli, each of which can support only broad qualitative distinctions. The goal of the present analysis is to improve upon this situation. What we seek is a method which is more direct, makes better use of the subject's responses, and supports more quantitative inferences about task strategy. Several previous investigators have approached this problem using a kind of perturbation analysis [4, 48, 105, 67]. In this paper we use a similar method to estimate linear weights for each band in which we correlate the subject's responses with the perturbations in the energies within a set of frequency bands on a trial by trial basis.

The method relies on the stochastic nature of the stimuli. Recall that signal stimuli are filtered noises which have mean power spectra given by the filter which produced them, while the backgrounds are white noise samples which have expected flat power spectra. Because the stimuli are noises, their spectral power fluctuates around the mean. In figure 3.3 we show the total power (energy) in a set of seven different nonoverlapping frequency bands plotted above the observer's binary response as a function of trial number. Depending on how an observer weights frequency space, the fluctuations in power within each band will cause different patterns of correct and incorrect decisions. Thus there exists a correlation between subject responses and the fluctuations in stimulus energies. This correlation suggests we can estimate the weights by back-correlating an observer's trial by trial performance with the spectral power actually present in the bands on each trial. The simplest method of estimating the weights is to directly correlate the power in the bands with the observer's responses[105]. However, this approach has limitations which make it less

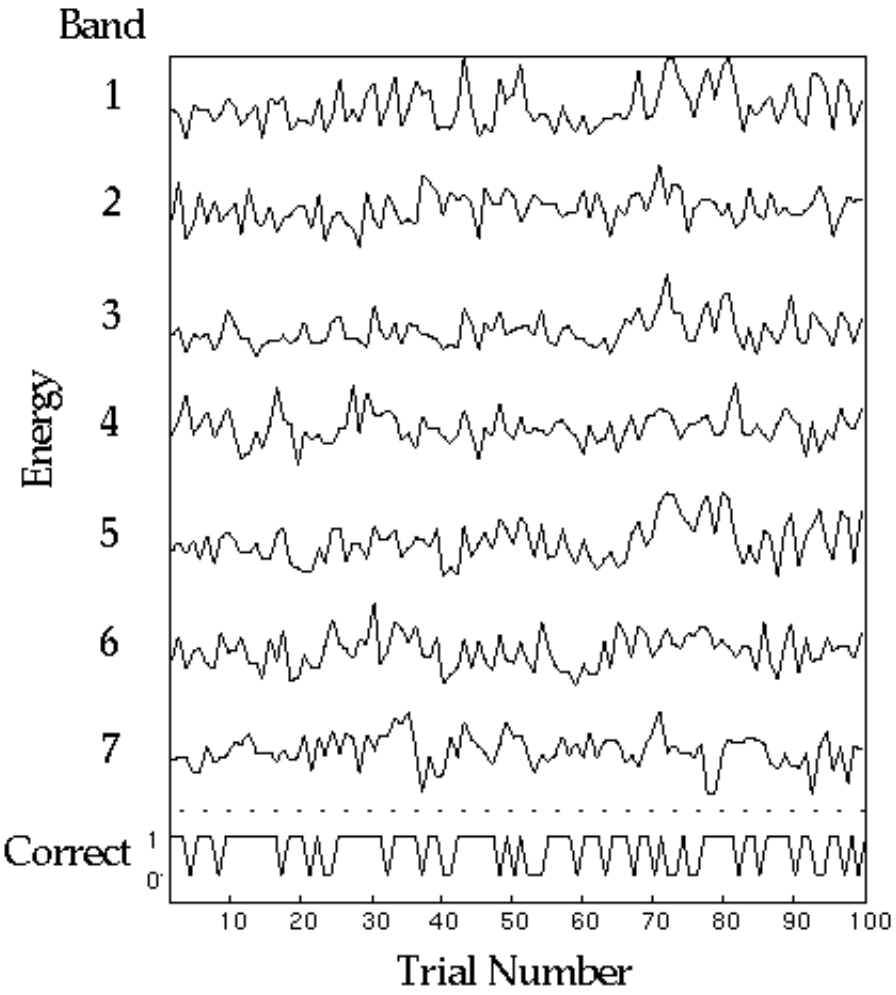

Figure 3.3: Total power in seven different nonoverlapping frequency bands plotted as a function of trial number. Because the stimuli are stochastic, total power within each band is a random function which fluctuates around a mean value. The bottom trace is the observer's responses, 0=incorrect, 1=correct.

applicable to the current study, the principle of which is that the weights estimated from correlations are known to be biased when the data is collected at several mean energy levels. Because of the limited amount of data collected, we were interested in using an unbiased method which combined all of the collected data[3]. Similar to Knill [67], we estimate weights from the maximum likelihood fit of a signal detection model to the data.

The signal detection model is illustrated in figure 3.4. In this model the observer is assumed to compute the energies e_i within each band for both signal plus noise and noise alone intervals. The energy computation is depicted as the integrals in the first boxes. Each of the energy estimates are corrupted by an independent additive noise N_i. Subsequently the energies are weighted by the scalars w_i and passed through an unknown function g. The weighted, transformed energies are then summed within each interval, and finally the difference of these sums is used as a decision variable dv, which is corrupted by central noise $N_{central}$.

[3]Note that unbiased here refers to the theory, given the assumptions of the analysis are correct. If the assumptions of an analysis are incorrect, then of course the method will not produce the correct weights. In Appendix A we explore some likely violations of the assumptions and how this might affect the weights.

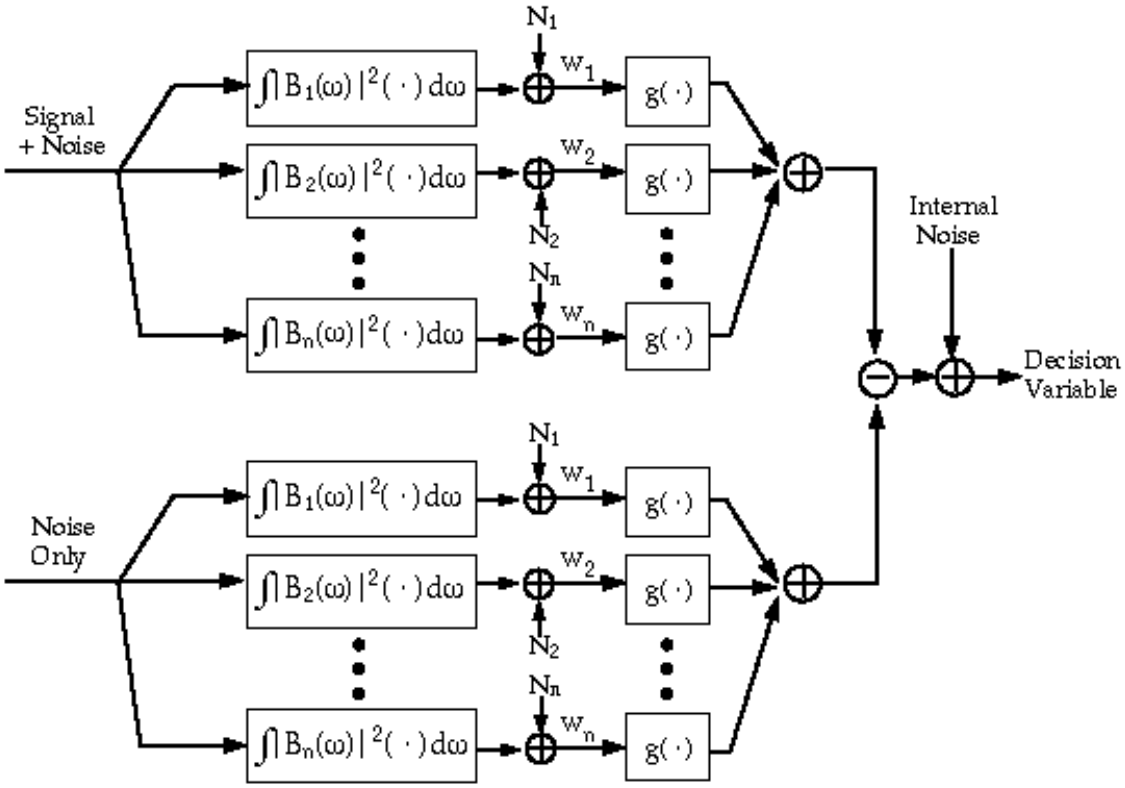

Figure 3.4: Signal detection model used in our analysis. See text for details.

$$dv = \sum_i \Delta g(w_i \cdot (e_i + N_i)) + N_{central} \tag{3.2}$$

Subject's responses are determined by the decision variable dv, with $dv > 0$ producing a correct response and $dv < 0$ an incorrect response. The unknown function g is used to represent the effects of pointwise non-linearities and/or non-gaussian noise to the pooling process[4].

This model can be linearized, as described in Appendix A, to yield the simpler expression:

$$dv = \sum_i w_i' \cdot \Delta e_i + N_{total} \tag{3.3}$$

where w_i' represent the linearized weights and N_{total} is the total noise at the decision variable. This model essentially lumps the effects of nonlinearities and non-gaussian noises into the noise term, whose distribution then ultimately determines the shape of the psychometric function in the model. If the noise were gaussian and independent of the energies, then the psychometric function should be well described as a cumulative gaussian. Let the subject's response on trial j be labeled by:

$$R_j = \begin{cases} 1 & : \quad dv > 0 \\ 0 & : \quad dv < 0 \end{cases} \tag{3.4}$$

[4]Note that the weight estimation procedure is only guaranteed to be unbiased if g is linear. For non-linear g, the linearization may be different at different signal levels, causing the estimated weights to be compromises across signal level.

then the probability of a correct response is given by:

$$p(R_j = 1) = p(dv > 0) = 1 - \Phi(0, \mu, \sigma^2_{N_{total}}) \tag{3.5}$$

where Φ is the cumulative gaussian function with the first argument giving the upper integrand. In the expression μ is a linear function of the energies and σ is assumed to be constant.

$$\begin{aligned} \mu &= \mu_{bias} + s(w_i' \cdot \Delta e_i) \\ \sigma^2_{N_{total}} &= c \end{aligned} \tag{3.6}$$

The free parameters in this model are the weights w_i', the mean bias term μ_{bias}, and the variance constant. The scalar s is simply absorbed into the magnitude of the weights.

In the case in which there are non-linearities or non-gaussian noise sources which are summed to form the decision variable, a central limit theorem argument suggests that N_{total} might still be described as approximately gaussian. However, the parameters of this gaussian distribution are likely to be functions of the energies. For instance, the variance of N_{total} may increase with increased signal levels, which manifests itself as a skewed psychometric function. The psychometric functions fit in Chap 1 all had significant skew, suggesting the variance is a function of the signal level. The simplest choice for modeling this possibility assumes N_{total} is a gaussian random variable with mean and variance functions linear in the input energies. The resulting model is the same as before but allows the variance to linearly change with the input energies.

$$\begin{aligned} \mu &= \mu_{bias} + (w_i' \cdot \Delta e_i) \\ \sigma^2_{N_{total}} &= c + d(w_i' \cdot (e_{i_{signal}} + e_{i_{noiseonly}})) \end{aligned} \tag{3.7}$$

The separate scale factor d is needed since the mean and variance functions may grow at different rates. The use of the sum rather than the difference in energies between the intervals for the variance term stems from the fact that the difference of random variables produces a new random whose variance is the sum of the previous two. Fits of this model to the data could not be rejected at the 0.05 level, which shows that the addition of the linear variance term was sufficient to account for the skew in the psychometric function. We tested the psychometric model against a null model in which the probability correct at each mean energy is allowed to take its maximum likelihood (mean) value [135]. A likelihood ratio test was used, in which the relative likelihoods of the psychometric model and the null model, $2\log(L_{psy}/L_{null})$, is approximately χ^2 distributed with $n - 13 - 3$ degrees of freedom. The weights are computed from the number of data samples n minus the total number of parameters: 3 psychometric plus the 13 weight parameters.

The likelihood of the weights and parameters given the data can be computed assuming that the set of subject's n responses R_j are independent bernoulli random variables:

$$L(w_i, \mu_{bias}, c, d|\{R_j\}) = \prod_{j=1}^{n} R_j p(R_j = 1) + (1 - R_j)p(R_j = 0) \tag{3.8}$$

$$\begin{aligned} L(w_i, \mu_{bias}, c, d|\{R_j\}) = \prod_{j=1}^{n} & R_j(1 - \Phi(0, \mu_{bias} + w_i' \cdot \Delta e_i, c + d(w_i' \cdot \Sigma e_i) \\ & +(1 - R_j)\Phi(0, \mu_{bias} + w_i' \cdot \Delta e_i, c + d(w_i' \cdot \Sigma e_i) \end{aligned} \tag{3.9}$$

where $\Sigma e_i = (e_{i_{signal}} + e_{i_{noiseonly}})$.

The weights and the three parameters were simultaneously fit to the data by computing the maximum likelihood parameters using a numerical search routine. To better assess whether the actual maximum was discovered, ten different random initial values of the parameters were used for each fit. For all the fits except the Planar data for subject PS, fits for each of the initial values found the same maximum. In the exception, several distinct maximum were discovered, maximum likelihood of this set was taken as the best fit. Each of the maxima was also corroborated by a Monte Carlo estimation of the mean of the likelihood function (described below), which showed that the estimated maxima were within the error of the estimated means for all the fits.

The standard errors and covariances of the weights and parameters estimates were computed by two methods, from the inverse of the hessian matrix of the likelihood function and by a Monte Carlo integration. The hessian matrix has elements given by

$$H_{ij} = \frac{\partial^2 L(\vec{\theta})}{\partial\theta_i \partial\theta_j} \tag{3.10}$$

where $\vec{\theta}$ is the combined vector of parameters and weights, and the inverse of the hessian matrix is a measure of the covariance matrix of the parameters and weights. The hessian was computed numerically using a modified finite difference algorithm, and the standard errors of the weights were computed from the square root of the diagonal elements of the inverse Hessian matrix after integrating out the three free parameters. Due to the large number of free parameters, the hessian approximation was sometimes unstable[5]. To ameliorate this problem, we used a second method to estimate the covariance matrix. The mean and covariance were estimated by computing the first and second moment of the likelihood function using Monte Carlo integration over 100,000 samples. The covariances computed in this manner were extremely similar for the two methods when the estimate of the Hessian was stable.

The standard errors in the estimates of the covariance matrix was assessed by performing a parametric bootstrap estimation. In this procedure, 10,000 data sets were simulated assuming that the model of the decision variable distribution was correct. The value of the decision variable for each trial was generated by sampling from a gaussian distribution whose mean and standard deviation were given by equation 3.8, using maximum likelihood weights and parameters and the energies for that trial. Correct responses were generated by finding trials for which the decision variable was greater than zero. New maximum likelihood weights were then generated by applying the fitting procedure to the simulated data. The results of this showed that the estimates of standard error were quite repeatable, with average deviations of roughly 5% of the standard errors. However, the estimates of the correlations between bands were quite noisy with an average standard error of 0.44. Most of the correlations were quite low, with the highest absolute value of the correlation coefficients being 0.52. Thus none of the correlations were significantly different from zero at the 0.05 level, however the noise in these estimates precludes many conclusions to be made about the correlations (e.g. we can't conclude from this that the correlations are equivalent to zero).

The estimated mean and covariance of the weights were used to compute a gaussian approximation to the likelihood function. T-tests of the difference of the weights from zero were computed using the estimated mean and covariance as the parameters of the sample distribution. Where multiple T-tests were used, the T-statistic was adjusted by the Tukey-Kramer correction.

3.4 Results

The results of the weight fitting procedure for the four stimuli are presented in figs 3.5- 3.8.

[5] i.e. the matrix inversion was nearly singular

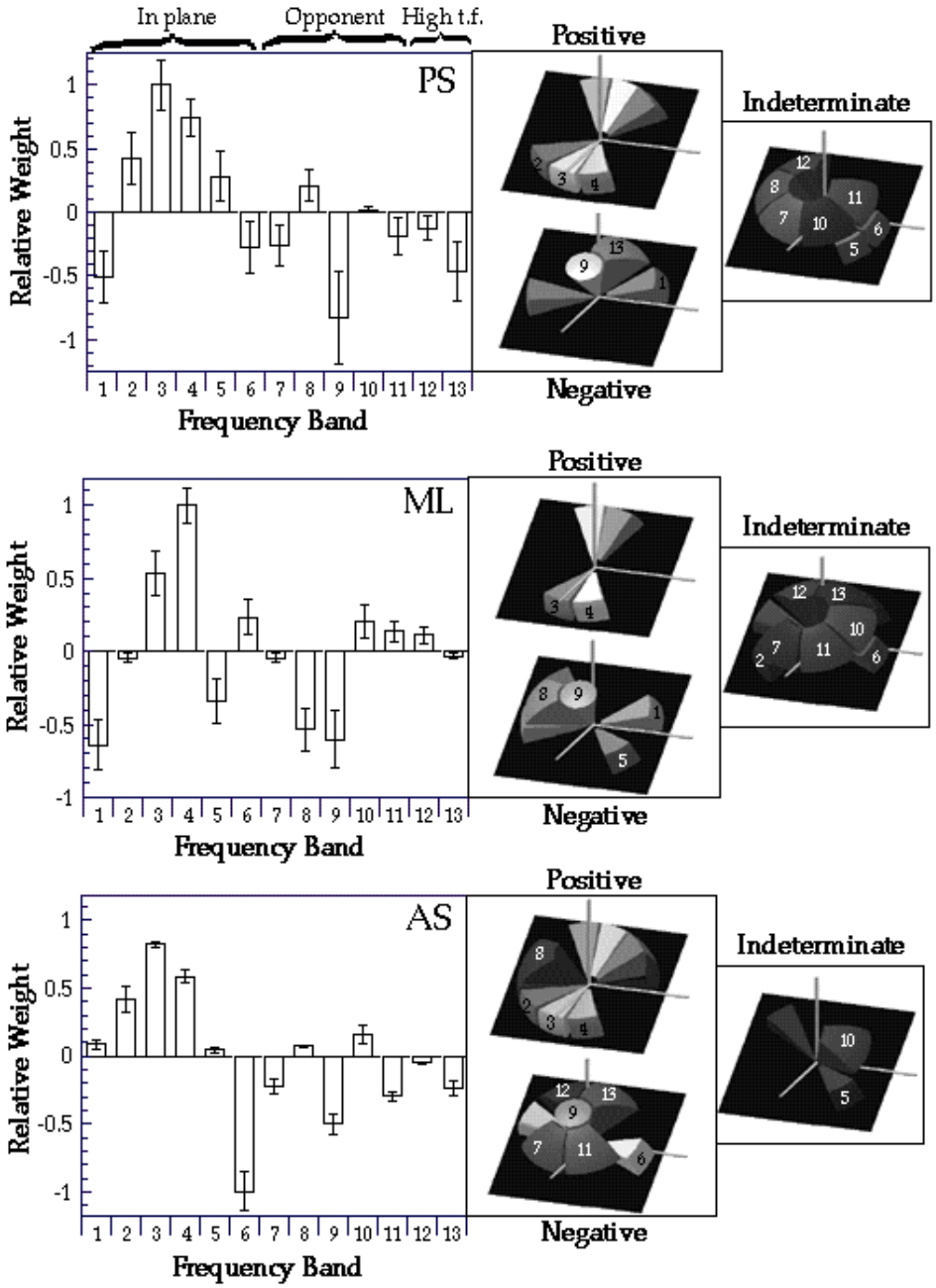

Figure 3.5: Resulting weights for Component stimuli for three subjects. Relative weight is plotted against the 13 frequency bands shown in fig 3.1. The bands corresponding to the weights are depicted are depicted as in fig 3.2 and are split into three categories: the bands receiving positive (top left) weights which are significantly different from zero at the 0.05 level, the significant negatively weighted bands (bottom left), and the bands which were not significantly different from zero (middle right). Because of the large number of comparisons and in some cases the large standard error on the weight we cannot infer from the lack of significance of the weights for these bands their equivalence to zero. A more valid inference is that we cannot reliably determine what the actual value of the weight is. Because of this inferential problem we chose to label these bands as Indeterminate.

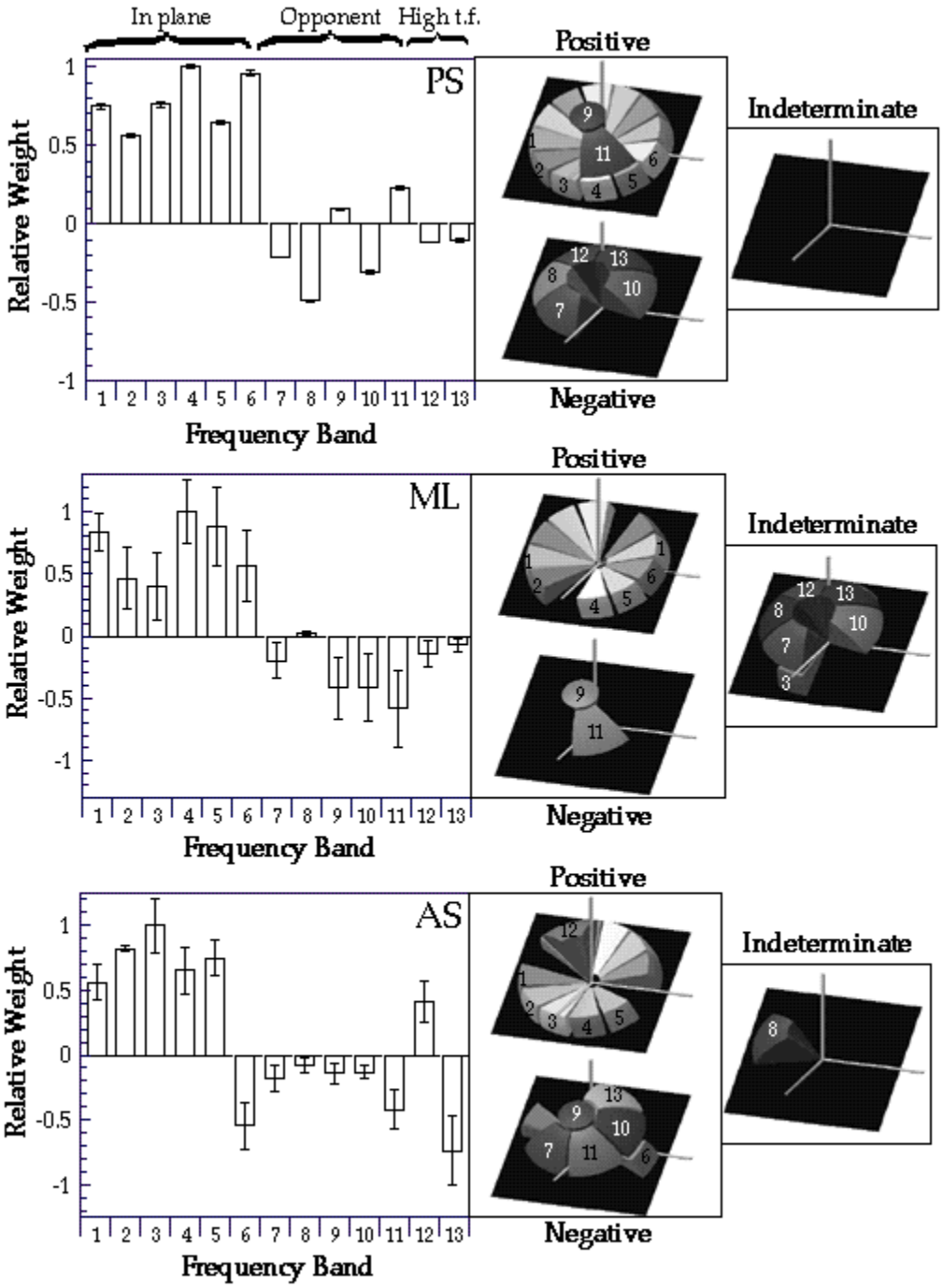

Figure 3.6: Resulting weights for Planar stimuli for three subjects.

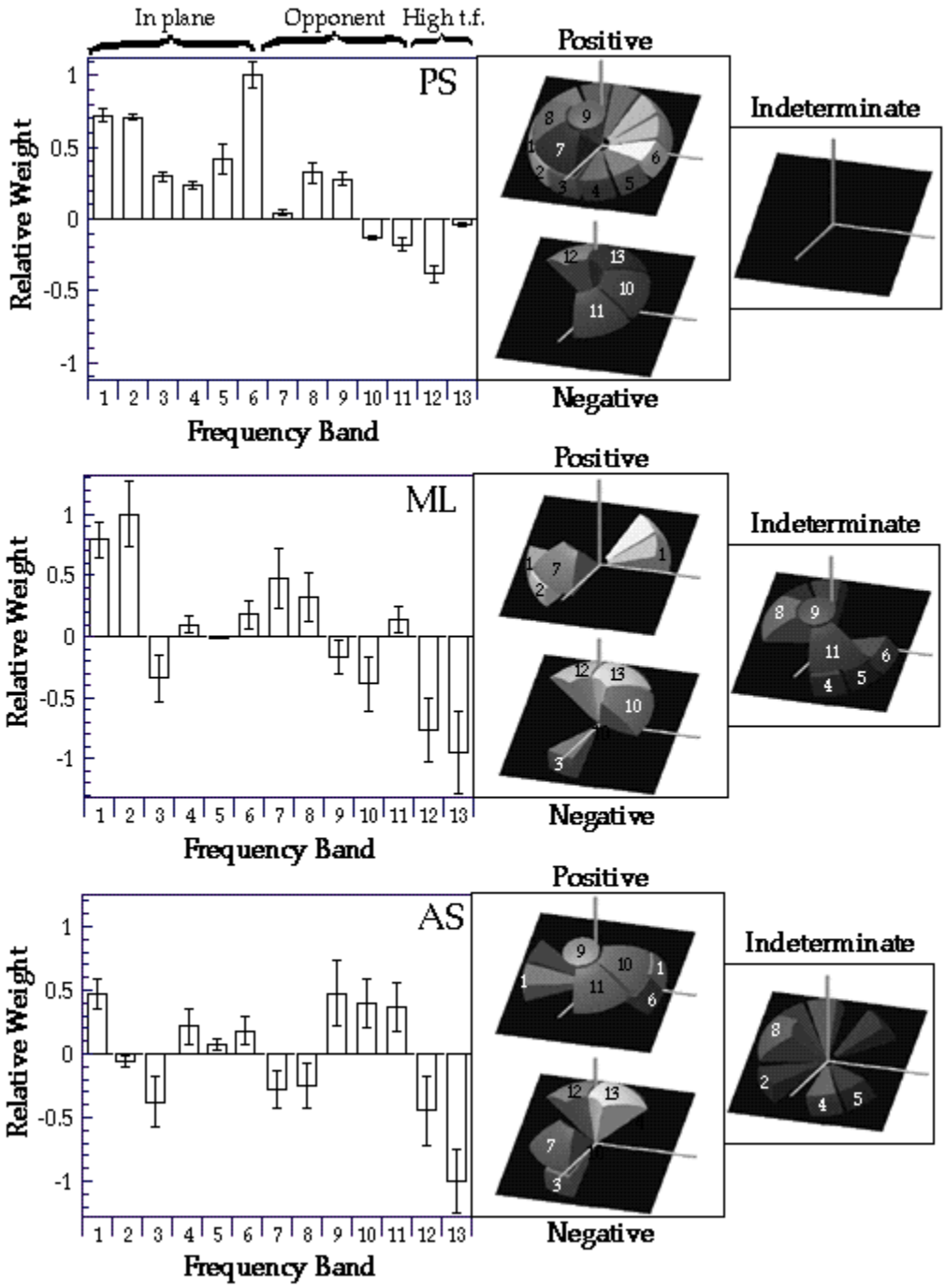

Figure 3.7: Resulting weights for Scrambled stimuli for three subjects.

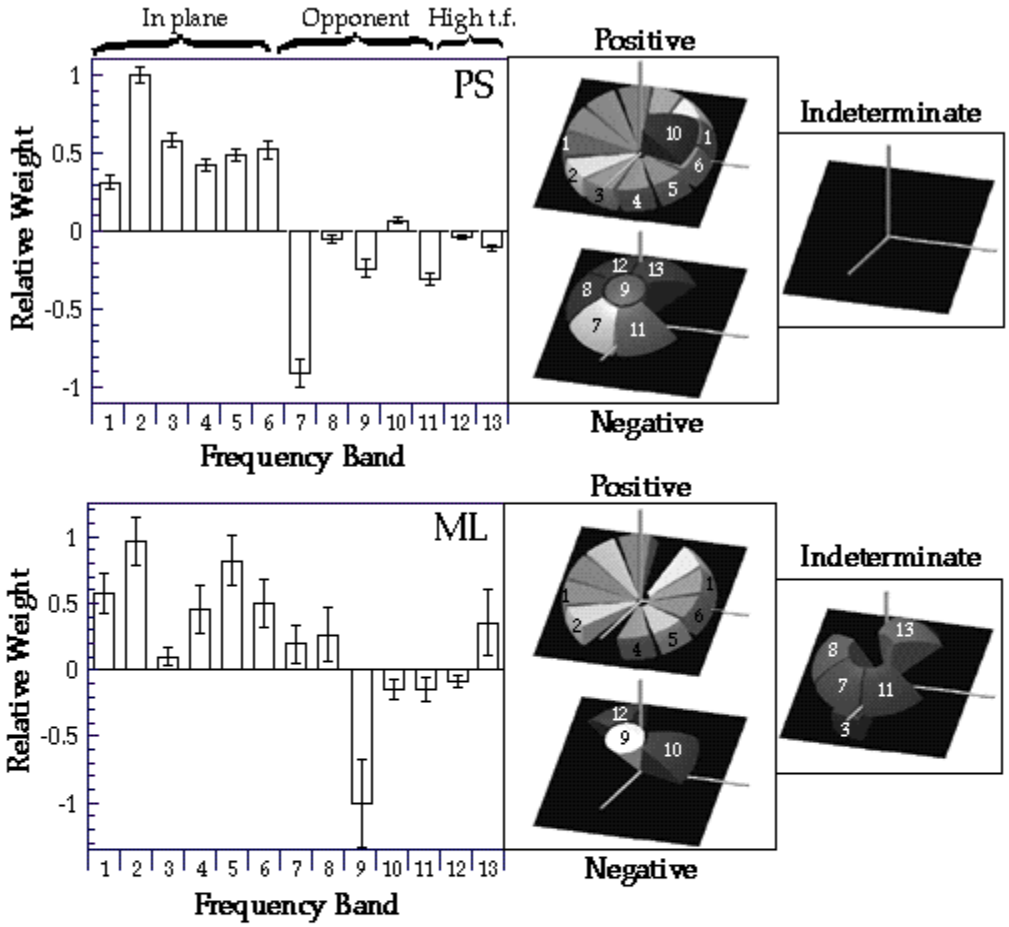

Figure 3.8: Resulting weights for Plaid stimuli for two subjects.

3.4.1 Results for Component stimuli

The results of the analysis applied to the Component stimuli are shown in fig. 3.5 for three subjects. The left hand side of the figure shows the estimated relative weights and their standard errors for each of the 13 bands. The right hand side depicts the bands corresponding to the weights, divided into three sets: the positive weights significantly greater than zero at the 0.05 level, the significant negative weights, and those weights which were not significantly different from zero. These insignificant weights were labeled 'Indeterminate' to guard against the inference that these weights are either unimportant or equivalent to zero. Weights close to zero could also be the result of subject's using a non-stationary weighting strategy, in which the bands are sometimes positively and sometimes negatively weighted during the course of the experiment. There are two prominent trends in the data. The first is that for all three subjects the bands which contain the signal (3 & 4) are the dominant positively weighted bands. For subjects PS and AS, a band which does not contain the signal (band 2) also receives a significant positive weight. The second trend is in the set of negative weights. All three subjects show significant negative weights given to band 9, which represents energy in the opposite direction as the signal band, and either band 1 or band 6, which represent fourier energy from frequencies which are orthogonally spatially oriented.

3.4.2 Discussion for Component stimuli

The weighting results have two interesting properties. The positive weights show that the subjects use a filtering strategy which is close to the properties of the signal. In fact, the significant positive weights for subject ML are within the error bounds of the ideal weighting. Subjects PS and AS include irrelevant information, which are most likely due to either i) a mismatch between the signal spectrum and the visual filter most sensitive to the signal, or ii) probability summation across all the visual filters which intersect the signal band so that the weights represent the envelope of these filters. The negative weights are in some ways more interesting, since they are not predicted from ideal weighting behavior. The common weighting of band 9 across subjects suggests something like motion opponency is occurring. This opponency could be subtractive [127, 1], or it could be divisive [78]. Divisive opponency was found by Lubin[78] in a set of contrast increment detection experiments involving the addition of small contrast increments to one of the components of counterphase gratings. Subtractive opponency may indicate that the visually system is more naturally a narrow band frequency discriminator than a narrow band detector. Divisive opponency, on the other hand could represent an attempt to 'whiten' the signal, or in other words the visual system may try to discount an expected structure of the background which includes frequencies moving in the opposite direction. In either case, the results show that some sort of opponency is occurring, which accords well with many previous models of narrowband signal detection.

The negative weights in bands 1 or 6 was more surprising. These bands represent frequencies whose spatial orientation is nearly orthogonal to the signal's orientation and which are temporally slow. This could represent the use of a spatial orientation discrimination strategy to detect the stimuli.

3.4.3 Results of Planar Data

The results for the Planar stimuli are presented in fig. 3.6. The significant positive weights show that the visual system was able to successfully pool information across the bands in the plane, corroborating the results of the qualitative analysis presented in the previous chapter. For all three subjects the bands which include the signal (1-6) are positively weighted with the exception of band 6 for subject AS. The magnitude of these weights vary, however, and thus are not equivalent to the ideal strategy of equally weighting bands 1-6. Subject PS's and AS's data also include small positive weights on bands which do not contain the

signal (PS: 9 and 11, AS: 12). In addition, the weights across bands 1-6 are clearly different for Planar stimuli than for Component stimuli.

No general trend exists for the negative weights across the three subjects, with the exception that significant weights are present and the majority of the off planar weights are negative for subjects PS and AS.

3.4.4 Discussion of Planar Data

The weights indicate that subjects are able to selectively pool frequency information across the planar bands. In addition, the summed difference between the subject' s weights on the planar and the ideal weights is a valid predictor of the ordering of subject efficiencies on the task PS > ML > AS. A likely source of the positive off planar contributions for subjects AS and PS is internal filter mismatch. If for instance the interval filter used by the visual system has a broader temporal passband which is larger than the temporal frequency range of the Planar signal, then the filter might systematically include the off planar regions which contain lower (bands 7 and 11) and higher (bands 12 and 13) temporal frequencies.

Because of the lack of agreement in the negative weights across subjects, inferences about the role of these weights are more difficult to make, and an interpretation should take into account this variability. The data are not consistent with subjects using an opponent velocity strategy, which would produce negative weights across 8,9,& 10, or a discrimination of moving vs. non-moving which would produce negative weights for bands 7 &11. The negative weights could be the result of an unequally weighted divisive normalization process or a subtractive opponency which is not general across subjects. One possible source of the variability in the weights is that observers are are not stationary in their weighting strategy. For instance, the subjects may be negatively weighting all the bands outside of the expected signals bands, but the subject's estimate of the expected signal may drift during the course of the experiment, causing bands outside of the signal to be positively weighted some of the time.

3.4.5 Results for Scrambled Stimuli

The weights estimated for detecting the Scrambled stimuli are shown in figure 3.7. Comparing the weights for the three subjects to the ideal weights shows that none of the subjects was able to learn the Scrambled signal spectrum completely. Subject PS comes the closest to the ideal weights with the correct sign on all the bands except 10 and 11. The significant positive bands for Subject ML, 1,2, and 7, are all adjacent and concentrated around a single frequency region low in temporal frequency and spatial orientations close to horizontal. The positive weights for subject AS are primarily concentrated around the bands intersecting a plane consistent with upward movement (the direction opposite to the Planar stimuli).

Common significant negative bands are 12 and 13 for all three subjects, exactly those bands which do not contain the Scrambled signal spectrum.

3.4.6 Discussion of Scrambled Data

Each of the three subjects appear to be using a different strategy for detecting the stimuli. The conclusion that none of the subjects appear to have learned the signal structure is corroborated by the low efficiencies on Scrambled stimuli[6] for all three subjects. In the positive weights two trends seem apparent. Subject PS appears to appropriately weight bands 1-6, suggesting that the subject may have been primarily looking for the downward moving component of this signal. The other two subjects positive weights are concentrated

[6]Recall that Scrambled efficiencies lied in between the predicted efficiency for using only a single component band and the predicted efficiency for probability summation across the Scrambled components.

around the horizontally oriented low temporal frequency bands. Since these bands are adjacent it is plausible that both these subjects primarily used horizontally oriented bandpass filters similar to those suggested by the Component data to detect the signal. The consistent negative weights on the high temporal frequency bands could be the result of subjects using a strategy of discriminating the signal from high temporal frequencies.

3.4.7 Results for Plaid Stimuli

Results were collected for two subjects on a stimulus type not used in the last chapter. The construction of the stimulus filter is easily described. Two filters with the same spectra as the Component filter were summed after being rotated to have spatial orientations of +/-70 deg and to lie in the common plane used for the other stimuli.

The results for Plaid stimuli for two subjects are presented in figure 3.8. For neither subject are the positive weights close to ideal. The main deviation is the presence of significant positive weights in bands 3 and/or 4, which are ideally zero, and that the magnitude of bands 2 and 5 is greater than expected. The common significant negative weights are the opponent direction bands 9 and the high temporal frequency band 12.

3.4.8 Discussion of Plaid Data

The plaid data is interesting in that the subjects were not able to learn to concentrate the weighting around the signal bands, and instead included bands from all spatial orientations (excepting band 3 for ML). This is true despite the fact that the weights for the component signal, and the weights for the Scrambled signal for subject ML, show that the subjects are able to selectively weight signals which are narrow band in spatial orientation. In fact the weights for the plaid stimuli have a qualitative resemblance to the weights for the Planar stimuli for both subjects. This suggests that the visual system may use a planar filter to detect these stimuli. Comparing the magnitudes of the Plaid and Planar weights by inspection and by t-test show that the weights are not significantly different for subject ML ($p < 0.05$), but are all significantly different ($p < 0.01$) for subject PS. Thus even if a planar pooling strategy is being used to detect both stimuli, there may be some adjustment of the weight gains into the pooling process for subject PS. This sort of gain adjustment could be the result of the different inhibitory influences the planar and the plaid signals have on the different regions of frequency space or represent partial adaptability to the properties of the signal. Another prominent possibility is that the positive weights are simply due to subjects relying on filters with large spatial orientation bandwidths which intersect both the signal and bands 3 and 4. There are two pieces of evidence which argue that this interpretation is less likely. For subject PS, the weights on bands 4 and 5 are nearly identical. If the weight on band 4 was due to the visual system using a filter which overlaps both band 4 and 5, then we would expect the two bands to be nearly perfectly correlated ($\sim$0.95). The actual correlation is 0.09 +/- 0.3, which is nearly three standard deviations less than the prediction. The other correlations are also low, but given the large standard error on the correlation estimates, we cannot say much about them. The other piece of evidence comes from the additivity experiments which are the topic of the next chapter. There we show that a plaid is additively combined with a single passband centered at 90 deg across all ratios of plaid and passband energy. This suggests that plaid stimuli may be always detected by a planar power detector. If this is true, then it may be that the visual system uses two basic strategies for detecting motion, one being narrowband in spatial orientation, and the other being broadband.

3.5 Discussion

The main results of the weighting analysis can be summarized as follows. For Planar and Component stimuli, the positive weights are reasonably matched to the signal bands. The positive weights for Plaid stimuli inappropriately include planar bands which do not contain the signal, while the positive weights for Scrambled stimuli inappropriately exclude many of the signal bands. For all signals subjects showed negative weighting of bands outside the signal. From these results we can make some basic conclusions about the kinds of strategies the subjects use in detecting the stimuli.

First, for none of the stimuli was there evidence for subjects using the spatial or temporal structure strategies. Thus the observers were able to use some of the spatio-temporal correlations in all the stimulus types. Second, although different weighting strategies were used for each stimulus, the results across the four stimuli show that the visual system is not able to adapt its weighting to arbitrary signals, particularly in the case of the Scrambled stimuli, and less so in the case of the Plaid stimuli. This lack of flexibility shows that the visual system does not perform like a generic power detector. Rather the visual system is better optimized to process certain signals. Although the number of different signals used in the experiments is quite small, we can conclude that the visual system is relatively optimized for stimuli with spectra concentrated around a point (Component) and a plane (Planar) in frequency space.

The relative optimization for Planar and Component stimuli show that the visual system can use at least two distinct modes of operation when confronted with signals which have fourier components lying on a common plane. The first mode of operation is a detector which is narrow band in orientation, while the second involves a detector which pools power across planar regions of fourier space and is broadband in orientation. The broadband detectors are readily identified with the planar power detectors which have been the focus of the thesis. The resulting weights augment the qualitative arguments made in the last chapter for the existence of such detectors. Thus it seems likely that the visual system does use special detectors for processing local velocity. The weighting result for the Component stimuli, however, demonstrate that the visual system has access to spatial structure information as well as the velocity (planar) information at the decision stage for these stimuli. Information from passbands may be used in conjunction with local velocity estimates to analyze the properties of moving textures. The possibility that both orientation pooled and orientation narrow band signals are preserved at a relatively late stage in visual processing is supported by the existence of both component and pattern type cells in area MT of simian visual cortex[90].

As previously noted, bands which are adjacent to the signal bands are included for both the Component and Plaid stimuli. The orientation bandwidths estimated from the weights are larger than would be expected from previous psychophysical and electrophysiological studies. This overlap could be the effect of using a population of similar narrow band power detectors and 'attending' to the set of detectors which intersect the spectra. In this case the bandwidths actually represent the envelope of the bandwidths of the detectors which intersect the signal.

3.5.1 Negative Weights

One of the most striking results of the analysis is the ubiquitous presence of negative weights outside the signal bands across all observers and stimuli. The negative weights are not predicted by a simple power detector model and must be accounted for by some additional aspect of motion processing. As noted before this could indicate one of several possibilities, which include optimization for certain discrimination tasks, and normalization. While the current analysis could not decide between the two, discrimination and normalization have different properties which are readily testable. A simple experimental paradigm is to use mixtures of the signal constructed from the positively weighted bands with a signal constructed from

the negative bands. By observing the change in performance induced by varying the relative energies in the mixtures, it is possible to infer whether a divisive, subtractive, or some other kind of law governs the interactions of the bands.

3.5.2 Generality of the Analysis

In performing the analysis we made particular choices of frequency decomposition and psychometric model. In this section we discuss the generality and potential limitations of these choices. We first discuss the psychometric model, then the choices of restricting the frequency decomposition to a sphere, and finally how the number of frequency bands was chosen.

The limitations of the psychometric function model are extensively dealt with in Appendix A & B, so a summary of those results follows. If subjects use an additive summation rule to produce a decision variable which is gaussian distributed and subject's weights are stationary, then the estimation procedure is unbiased. We relax the distribution and additive rule assumptions and show that a large class of non-gaussian distributions and non-linearities can be dealt with by adding a variable to the psychometric model. However, the procedure uses a linearization around a given signal level, which means the procedure is not guaranteed to produce unbiased weights when multiple signal levels are used. The potential biases are a compromise between the different linearized weights at each signal level. Since we have no reason to expect sudden changes in sign or magnitude of the subject's weights as a function of signal level, we expect the procedure is fairly robust to violations of the assumptions, particularly in the signs of the weight estimates. The assumption that subject's weights are stationary is implicit in any psychophysical paradigm similar to ours. Our principal support for this assumption is the extensive training observers had on all the stimulus types, which was continued until performance asymptoted.

In the frequency decomposition, choices were made to limit the analysis to a sphere in frequency space, and to limit the number of frequency bands. The analysis was limited to the spherical region after determining that subjects were not significantly weighting the frequencies outside this region. We performed an initial decomposition of frequency space into a set of four bands, two of which included the frequency sphere and two of which did not. Computing weights across these four bands showed that for all subjects and all conditions, the two bands which did not include the sphere received essentially zero weight. In addition, for none of the subjects or conditions did the addition of bands outside the sphere lower the log likelihood of the fit more than 3 tenths of a log unit, which is not significant difference in the fit. Thus we concluded that the frequencies outside the sphere were not used by the subjects. This agrees with previous results which suggest that there exists a 'Window of visibility', which is a nearly arithmetical trade-off between spatial and temporal frequency such that when spatial frequency is increased, temporal frequency must be decreased by a similar amount[140].

The number of bands was limited by two considerations. The first was the number of trials collected. Simulations of the fitting process showed that reasonable fits, in terms of the accuracy and the error of the estimate were obtained when the number of trials exceeded the number of bands by about two orders of magnitude. Since the total number of trials varied between 750-1300, we wanted to keep the total number of bands close to ten or less. In addition, the multidimensional minimization procedure is more likely to get stuck in local minima for large numbers of weights. Both argue for using a small number of frequency bands. However, simulations also showed that lumping differently weighted frequency regions into one frequency band results in estimated weights which are approximately averages of the true weights. Given the presence of both positive and negative weights in the results, lumping could conceivably result in estimated weights close to zero. Thus, to get an accurate picture of the subject's weights we should use as many bands as is feasible. The simulations suggested that using 10 to 20 different bands constituted a reasonable compromise. We can consider the constraint on the number of bands to be a *sampling* constraint.

Like all sampling methods, our method will work best when the observer's weighting function is smooth and slow-varying across frequency. Since the 13 bands we used constitutes a fairly coarse sampling of frequency space, our analysis is insensitive to any variations in the subject's weighting function which are finer grained than our sampling.

3.6 Conclusions

Of the pooling strategies discussed in the introduction, detection using only spatial or temporal structure can be eliminated outright for all the stimuli. Thus, although we used a task which does not explicitly require a judgement based on the perception of motion, we show that subjects are using motion information to detect the stimuli.

The weights for Planar stimuli suggest the visual system uses planar power detectors to process local motion information. The fact that observers can correctly weight narrowband stimuli may indicate the visual system needs image texture information (given by the collection of narrowband detectors) as well as image velocity information (given by planar detectors) to interpret local motions.

The inability of subjects to learn the signal model for Scrambled and Plaid stimuli suggest that observers do not use generic pooling strategies like probability summation to detect the stimuli.

The presence of negative weights means the basic power detector model is not complete, hence the visual system is not optimized for a pure translation detection strategy. Negative weights for Planar stimuli did not fit the two most obvious motion discrimination strategies, but could be the result of contrast normalization, pre-whitening, or predictive coding strategies. This suggests that the goal of the visual system is local translation *encoding* rather than local translation detection. However, negative weights for Component stimuli did fit a simple discrimination model, with weights consistent with opponent motion discrimination and orientation discrimination, but are also consistent with directed normalization. Further research will be necessary to determine the functional role of the negative weights.

3.7 Appendix A

In this appendix we show that estimating linear weights is a kind of first order approximation to the class of decision models which use a one-dimensional decision variable.

We formalize the decision model by assuming the visual system computes the quantities X_s and X_n on the signal plus noise (s) and noise alone (n) intervals, which are scalar functions of the energies $\mathbf{E}_i$ for bands $i \in \{1, \ldots, M\}$. The difference between X_s and X_n is used as a decision variable d to produce a binary response R.

$$\begin{aligned} X_s &= f(\mathbf{E}_1^s, \mathbf{E}_2^s, \ldots, \mathbf{E}_M^s) \\ X_n &= f(\mathbf{E}_1^n, \mathbf{E}_2^n, \ldots, \mathbf{E}_M^n) \\ d &= X_s - X_n \\ R &= \{_0^1 \text{ for } d \gtrless 0 \end{aligned} \tag{3.11}$$

If we assume that f can be expanded as a Taylor series in $\mathbf{E}_i$ around the energy vector $\vec{\mathbf{E}}_0$, then we may write down the decision variable as a linear combination of the energies by ignoring the second and higher order terms.

$$X = \sum_i f(\mathbf{E}_{i_0}) + \sum_i \frac{\partial f(\mathbf{E}_{i_0})}{\partial \mathbf{E}_i} \cdot \mathbf{E}_i + \sum_i \frac{\partial^2 f(\mathbf{E}_{i_0})}{\partial \mathbf{E}_i^2} \cdot \mathbf{E}_i^2 + \cdots \tag{3.12}$$

If we designate $\frac{\partial f(\mathbf{E}_{i_0})}{\partial \mathbf{E}_i}$ by $w_i(\mathbf{E}e_{i_0})$ and ignore the second and higher order derivatives, we have:

$$X = b + \sum_i w_i(\mathbf{E}_{i_0}) \cdot \mathbf{E}_i \tag{3.13}$$

where $b = \sum_i f(\mathbf{E}_{i_0})$ is a mean bias term. Then the decision variable d becomes:

$$\begin{aligned} d &= X_s - X_n \\ d &= (b_s - b_n) + \sum_i w_i \cdot (\mathbf{E}_i^s - \mathbf{E}_i^n) \\ d &= b' + \sum_i w_i \cdot \Delta\mathbf{E}_i + (\text{ higher order terms}) \end{aligned} \tag{3.14}$$

Thus, as long as the linear part of f dominates, then the analysis will capture the dominant behavior of the decision system. This means that the weight fitting procedure does not rely on an additive combination rule for the energies. However, there are two prominent cases in which the estimated weights will not exactly correspond to 'actual' weights.

The approximation is best when the perturbations in the energies are small and centered around a fixed set of $\mathbf{E}_{i_0}$, so that the $w_i(\mathbf{E}_{i_0})$ are approximately constant. In the experiments we used 5 or 6 different signal energies, and hence 5 or 6 different mean values of the $\mathbf{E}_{i_0}$. When the $w_i(\mathbf{E}_{i_0})$ do not remain reasonably constant across the different values of the energies, the fitted weights will represent a compromise between the linearizations at the different signal energies. The other salient source of error is if the function f has significant second or higher order terms. Here again we are performing a best linear fit to a curve which is non-linear, and the fitted weights will represent the compromises that went into the fit.

An important example of the linearization is where the energy in each of the bands multiplicatively interacts with the other bands. For example, in a normalization model, the energy in each of the bands is divided by a weighted sum of the other bands.

$$\mathbf{E}_{out} = \frac{\mathbf{E}_1}{\sum_i n_i \mathbf{E}_i} \tag{3.15}$$

We are essentially finding the best linear fit to this expression:

$$\frac{\mathbf{E}_1}{\sum_i n_i \mathbf{E}_i} \simeq \sum_i w_i \mathbf{E}_i \tag{3.16}$$

For a particular value of the energies, this is given by the leading order term of a Taylor series expansion:

$$\sum_i \frac{\partial}{\partial \mathbf{E}_i} \left(\frac{\mathbf{E}_1}{\sum_i n_i \mathbf{E}_i}\right) = \sum_i -n_i E_1 \left(\frac{1}{(\sum_i n_i \mathbf{E}_i)^2}\right)\Big|_{\vec{\mathbf{E}}=\vec{\mathbf{E}}_0} \tag{3.17}$$

which can be written:

$$\sum_i w_i(\vec{\mathbf{E}}_0)\mathbf{E}_1 \tag{3.18}$$

where the w_i are negative and functions of the energies. Since the energies vary from trial to trial, the actual weights will be averages of the best linear weights for a single energy. Notice that the coefficients w_i are not equal to the actual multiplicative weights n_i, however, the relative weights $w_i / \max_i(|w_i|)$ are equal to the relative multiplicative weights.

3.8 Appendix B

In this appendix we investigate the robustness of the fitting procedure to violations of the underlying assumptions. First we discuss the basic model and show that the parameter estimates are consistent.

Recall that the basic model is a weighted sum of energies within the set of spatio-temporal passbands plus some internal noises.

$$d = \vec{\alpha} \cdot (\vec{\mathbf{E}}_{BP} + \vec{\mathbf{n}}_{BP})_{signal} - \vec{\alpha} \cdot (\vec{\mathbf{E}}_{BP} + \vec{\mathbf{n}}_{BP})_{noisealone} + \mathbf{n}_{central} \tag{3.19}$$

where $\vec{\alpha}$ is the set of weights, $\vec{\mathbf{E}}_{BP}$ is the vector of energies across the set of passbands, and the $\mathbf{n}$ are internal noises, either associated with the measurement of each energy ($\vec{\mathbf{n}}_{BP}$), or present centrally in the decision variable ($\mathbf{n}_{central}$). If the noises are independent of the value of $\vec{\mathbf{E}}$, then d can be expressed as a noise process added to the weighted difference in energies between intervals:

$$\begin{aligned} d &= \vec{\alpha} \cdot \Delta\vec{\mathbf{E}}_{BP} + \mathbf{n}_{total} \\ \text{where } \mathbf{n}_{total} &= \vec{\alpha} \cdot \Delta\vec{\mathbf{n}}_{BP} + \mathbf{n}_{central} \end{aligned} \tag{3.20}$$

The simplest case is when $\mathbf{n}_{total}$ can be well approximated by as a gaussian random variable with mean $\mu_{\mathbf{n}_{total}}$ and variance $\sigma^2_{\mathbf{n}_{total}}$. In this case we can write down the probability of a correct response from the distribution of the decision variable d. We will assume a correct response occurs whenever $d > 0$. Then the probability of a correct response $R_i = 1$ is given by

$$p(R_i = 1) = 1 - \mathbf{\Phi}(0, \mu_{\mathbf{n}_{total}} + \vec{\alpha}_0 \cdot \Delta\vec{\mathbf{E}}_{BP}, \sigma^2_{\mathbf{n}_{total}}) \tag{3.21}$$

where $\mathbf{\Phi}$ is the cumulative normal function where the first argument is the upper limit of integration, the second and third arguments are the mean and variance of the gaussian distribution.

We are essentially fitting the weights by modeling the distribution which underlies the psychometric function as a normal distribution with constant variance and a mean which is a linear function of the weighted difference in energies,

$$\begin{aligned} \mu_d &= \mathrm{a}x + \mathrm{b} \\ \sigma^2_d &= \mathrm{c} \end{aligned} \tag{3.22}$$

where $x = \vec{\alpha} \cdot \Delta\vec{\mathbf{E}}_{BP}$ and a is a scalar which is absorbed into the magnitude of the weights. Thus, there are $M + 2$ free parameters which we bundle into a vector $\vec{\beta} = [\vec{\alpha}, b, c]$, where $\vec{\alpha}$ has M parameters.

To estimate the weights we maximize the likelihood function over $\vec{\beta}$, which is formed by assuming the responses are bernoulli random variables:

$$L(R_i|\vec{\alpha}) = \prod_i (p_i(\vec{\beta}))^{R_i} \cdot (1 - p_i(\vec{\beta}))^{1-R_i} \tag{3.23}$$

It is equivalent to maximize the log likelihood:

$$\begin{aligned} \log L(R_i|\vec{\beta}) &= \sum_i \log(p_i(\vec{\beta})) \cdot R_i + \log(1 - p_i(\vec{\beta})) \cdot (1 - R_i) \\ \vec{\beta}^* &= \arg\max_{\vec{\beta}}(\log L(R_i|\vec{\beta})) \end{aligned} \tag{3.24}$$

The equation is nonlinear in $\vec{\beta}$ and we could not solve it in closed form, hence we performed the maximization numerically.

Assuming that the model is true, we can show that the procedure produces unbiased estimates of the weights, called *consistency* in statistical parlance. Consistency follows if we can show that the expected likelihood function across all data sets of size N has a maximum at the true value of the parameter as N grows large. A simple argument follows. We compute the expected log likelihood function over the set of all R_i. Since the log likelihood function is linear in the R_i, the expected log likelihood function is simply the weighted sum of the expectations of the R_i. Since the R_i are bernoulli random variables, $E[R_i] = p_i$. Thus

$$E[\log L(R_i|\vec{\beta})] = \sum_i \log(p_i(\vec{\beta})) \cdot p_i + \log(1 - p_i(\vec{\beta})) \cdot (1 - p_i) \tag{3.25}$$

This equation has a well known maximum at $p_i = 1/2$. Equation 3.21 shows that this value of p_i occurs when $\vec{\beta} = \vec{\beta}_0$. Thus the procedure is consistent. We also verified this argument by performing a large set of Monte Carlo simulations of the fitting process. The fitted psychometric models for the Planar stimuli for subjects PS and AS were used to generate 20 different artificial data sets. The trial by trial energies had gaussian noises added to each band, were then weighted by a random set of weights of the same magnitude as those estimated for the subject and summed together. Simulated correct responses were generated if the signal interval had the larger sum. The noises added to the energies had randomly chosen variances which were constrained to sum to the estimate of the subject's decision variable variance. A thousand different fits to each of these simulated data sets were performed. In all cases the estimated weights were very close to the true values, even though the variances in each band could be quite different.

While this model is quite simplistic, a slight modification captures the leading behavior of a wide range of models. The modification is to allow the variance of the decision variable distribution to change with the signal level. We will briefly show how a variance term which is a function of signal level naturally arises in two situations, non-gaussian additive noises and non-linear transformations of the energy variables after internal noises have been added.

3.8.1 Nongaussian internal noise sources

Consider the basic model:

$$\begin{aligned} d &= \vec{\alpha} \cdot \Delta\vec{\mathbf{E}}_{BP} + \mathbf{n}_{total} \\ \text{where } \mathbf{n}_{total} &= \vec{\alpha} \cdot \Delta\vec{\mathbf{n}}_{BP} + \mathbf{n}_{central} \end{aligned} \tag{3.26}$$

Assume that $\vec{\mathbf{n}}_{BP}$ are non-gaussian noises (e.g. Poisson). If the dimensionality of $\vec{\mathbf{n}}_{BP}$ is reasonably large then we expect $\mathbf{n}_{total}$ will be approximately gaussian by a central limit theorem argument. The mean and variance of $\mathbf{n}_{total}$ will be a linear combination of the first and second moments of the distributions of $\vec{\mathbf{n}}_{BP}$.

Since $\mathbf{n}_{BP_i}$ is added to $\alpha_i \mathbf{E}_{BP_i}$, the density function can be written $g_i(\, x \, - \, \alpha_i \mathbf{E}_{BP_i}\,)$, where x denotes a generic energy variable. Assuming that the first two moments of this distribution exist, the first moment will be given by $\mu_i = \kappa_i + \alpha_i \mathbf{E}_{BP_i}$, where κ_i is the mean of the distribution when centered on zero. In general, the variance will be a constant plus some function of the mean:

$$\begin{aligned} \sigma_i^2 &= \lambda_{1_i} + h(\mu_i) \\ \sigma_i^2 &= \lambda_{1_i} + h(\kappa_i + \alpha_i \mathbf{E}_{BP_i}) \end{aligned} \tag{3.27}$$

Expanding this equation in a Taylor series in $\mathbf{E}_{BP_i}$ and only keeping the first order term yields:

$$\sigma_i^2 = \lambda'_{1_i} + \lambda_{2_i}(\alpha_i \mathbf{E}_{BP_i}) \tag{3.28}$$

where λ'_{1_i} represents λ_{1_i} lumped together with the constant from the Taylor series. Using these expressions for the mean and variance of the $\mathbf{n}_{BP_i}$, we can write down the mean and variance of $\mathbf{n}_{total}$:

$$\begin{aligned} \mu_{total} &= \sum_i \kappa_i + \sum_i \alpha_i (\mathbf{E}^s_{BP_i} - \mathbf{E}^n_{BP_i}) \\ \mu_{total} &= \kappa + \vec{\alpha} \cdot \Delta \vec{\mathbf{E}}_{BP} \end{aligned} \tag{3.29}$$

$$\begin{aligned} \sigma^2_{total} &= \sum_i \lambda'_{1_i} + \sum_i \lambda_{2_i} \alpha_i (\mathbf{E}^s_{BP_i} + \mathbf{E}^n_{BP_i}) \\ \sigma^2_{total} &= \gamma_1 + \gamma_2 \sum_i \alpha'_i (\mathbf{E}^s_{BP_i} + \mathbf{E}^n_{BP_i}) \\ \sigma^2_{total} &= \gamma_1 + \gamma_2 \left(\vec{\alpha}' \cdot \vec{\mathbf{E}}^{\mathbf{S}}_{BP} \right) \end{aligned} \tag{3.30}$$

where γ_1 lumped set of constants, γ_2 represents the common factors in the λ_{2_i}, the α'_i represent the weights lumped with the non-common factors of the λ_{2_i}, and $\vec{\mathbf{E}}^{\mathbf{S}}_{BP}$ represents the sum of the energies in the signal plus noise and noise alone intervals.

Since it is reasonable to assume that all the energy bands are processed similarly, we can assume that the density functions $g_i(x)$ will also be quite similar. Hence it is likely that the $\lambda_{2_i} \simeq \gamma_2$ for all i, and so $\vec{\alpha}' \simeq \vec{\alpha}$. Equations 3.30 and 3.31 specify a psychometric model in which there are $M+3$ parameters: $\vec{\alpha}$, κ, γ_1, and γ_2.

3.8.2 Nonlinear combinations of bands

Consider the non-linear model:

$$d = f(\vec{\mathbf{E}}^s_{BP} + \vec{\mathbf{n}}^s_{BP}, \vec{\mathbf{E}}^n_{BP} + \vec{\mathbf{n}}^n_{BP}) \tag{3.31}$$

If we approximate the function $f()$ by its Taylor series and truncate to the first term, then we have the situation sketched in Appendix A. The additive noise sources will then be transformed by the function into something unlikely to be gaussian. At this point we have arrived at the same situation as detailed in the section above. The simplest approximation to this general case is the same as the model given above, a cumulative normal psychometric model with $M+3$ parameters $\vec{\alpha}$, κ, γ_1, and γ_2.

3.9 Appendix C: Analysis Filters

The simplest way of describing the analysis is in terms of a spherical coordinate system $(\omega_\theta, \omega_\phi, \omega_r)$ which is rotated away from the $(\omega_x, \omega_y, \omega_t)$ axes such that the equator of the spherical coordinates lies within the plane in figure 3.1. Given this coordinate system, the bands can be described by the bounds on these spherical variables. The bounds on spatial frequency magnitude were (0.49,10.2) cyc/deg along the spatial frequency plane. The bounds on temporal frequency along the ω_t axis were (1.4, 21.8) Hz. The frequency radius of the sphere is given by $\omega_r = \sqrt{\omega_x^2 + \omega_y^2 + (\omega_t/2.1)^2}$, for which the bounds are (0.49,10.2). The angular bounds are summarized in the following table:

Bands	ω_θ range	ω_ϕ range
{1,2,3,4,5,6 }	(0, 30) + j(30,30) deg	(-30, 30) deg
{10,13,12,8,7,11 }	(-30, 30) + j(60,60) deg	(30, 75) deg
9	(0, 360) deg	(75, 90) deg

where $j \in \{1, \ldots, 6\}$ corresponds to the jth band in the set.

Since the discrete fourier transform was used, some of the frequency voxels intersected the boundaries between the bands. These frequencies were assigned to the band which included the majority of the voxel.

Chapter 4

Additivity Experiments

4.1 Introduction

In the last two chapters we have shown evidence for subject's ability to efficiently pool Fourier power across planar regions of frequency space. In this chapter we test one of the prominent predictions of the planar power detector model: additive pooling of energy on the plane. The primary reason for testing additivity is that the existence of an additive law suggests observers are using specialized pooling mechanisms to detect stimuli, since contrast pooling across frequency bands is typically subadditive [51, 50]. Conversely, a non-additive pooling rule would not require the use of specialized mechanisms, and parsimony of strategy would argue that the pooling observed in detecting the stimuli is due to a general rule.

Another reason to test additivity is that the ideal pooling strategy for this task is additive. By measuring the observer's conformance to an additive law, we can infer what proportion of the subject's inefficiency is due to an inappropriate pooling rule. Finally, the results of the analysis in the last chapter could be substantially improved by having knowledge of the pooling rule, either in the strength of the conclusions if the rule is additive, or by suggesting an improved psychometric model for a re-analysis.

In the rest of the paper, we explain why the planar power detector model predicts additivity, and describe the stimuli and the experimental procedure. We then explain the data analysis and show the results. We discuss the results in terms of models of motion processing, possible relations to physiology, and the possibility of partial adaptability of orientation pooling.

4.1.1 Additivity predictions

The planar power detector additively pools power concentrated around a common plane to measure the spectral energy E. If we split the signal into several non-overlapping bands, then the detector can be described as summing the energies within each signal band, weighted by the detector's spectral sensitivity. If $|S(\vec{\omega})|^2 = \sum_i |S_{b_i}(\vec{\omega})|^2$ denotes the signal spectrum and $|H(\vec{\omega})|^2$ denotes the power spectrum of the detector, then the output E can be expressed as:

$$\begin{aligned} E &= \int_{\vec{\omega}} |H(\vec{\omega})|^2 |S(\vec{\omega})|^2 d\vec{\omega} \\ E &= \sum_i w_i E_{b_i} \end{aligned} \tag{4.1}$$

Where E_{b_i} is the signal energy in band b_i, and the w_i are weights which represent the average effect of the detector's sensitivity within band b_i on the signal energy in the band. We show in the Appendix that the performance of a planar power detector is well described by a psychometric function which only depends

on the energies in the signal bands and the background noise power level **N**. If we denote the psychometric function by Ψ, then the probability of a correct response R_i for an observer relying on a planar power detector is given by:

$$p(R_i = 1) = \Psi\left(\sum_i w_i E_{b_i}, \mathbf{N}\right) \tag{4.2}$$

If we fix **N**, then for any fixed probability correct the weighted sum of the signal energies must equal a constant: $\sum_i w_i E_{b_i} = c$. Assuming that the observer's performance is entirely based on planar power detectors, then we should observe the same performance for any combination of the energies whose weighted sum is equal to the same constant c.

Prediction 1: In a task in which an observer relies on planar power detectors, the observer should additively combine the energies from bands which intersect a common plane in frequency space.

Prediction 2: If planar configurations of power are specially processed by the visual system, we would expect subadditive combination of non-planar configurations of power.

To test these predictions we used an experimental paradigm similar to the one used in the previous experiments.

4.1.2 Experimental Logic

The basic idea is to combine two sets of bandpass signals in different ratios of signal energy. If the signal sets are being additively combined, then performance should be determined by the total signal energy independent of the ratio. We used two different types of bandpass signal sets, *plaid* signals and *bandpass* signals. The *bandpass* signals are produced by passing spatio-temporal white noise through a filter which is bandpass in spatial orientation, spatial frequency magnitude, and temporal frequency. These signals can be described as noisy 'gratings', since the stimuli have dominant spatial and temporal frequencies. The *plaid* signals are formed by summing two equally weighted bandpass components which have different peak spatial orientations but which lie on a common plane. Since each of the components can be described as a noisy 'grating', it is natural to call these combinations 'plaids'.

The plane in frequency space associated with each velocity can be specified by two frequencies on the plane, since the plane is constrained to pass through the origin. Thus each plaid signal determines a unique plane through the component's peak frequencies. The minimum number of bands which can be used to test for pooling on and off a common plane is three: a plaid which defines the common plane and a third 'test' band which can be added to the plaid in various energy ratios to test additivity. Two different plaids signals and three different bandpass signals were used in three different experimental conditions. The expected signal spectra for the three conditions is shown in figure 4.1.

Two of the conditions, designated *In-Plane* and *Asymmetric* involve combining plaid and bandpass signals which lie on a common plane in various energy ratios. In the In-Plane condition the plaid signal has bandpass components which are symmetrically oriented around the direction of motion of the translation specified by the common plane. The bandpass signal has an orientation which lies between the plaid orientations, and thus is in direction of the plaid pattern motion. Since the additivity prediction does not depend on which bands are traded off in the plane, we chose a different combination of the same three bands for the second condition. In the Asymmetric condition, the plaid is asymmetric and is formed from the signals in the In Plane condition by combining one of the plaid components with the bandpass signal. The bandpass signal in the Asymmetric condition is just the other plaid component from the In Plane condition. In this condition the combination signal has the same perceived direction of motion as the In Plane condition, however, the plaid and bandpass signals alone are perceived to move in directions different from the common motion.

The third condition was designated *Off Plane*, and involved combining the symmetric plaid from the In Plane condition with the bandpass signal from this condition rotated out of the common plane to lie near the zero velocity. This off plane bandpass signal was chosen because the spatial frequency spectra of the components in the In Plane and Off Plane conditions is nearly identical, and it allows us to separate orientation specific pooling from motion specific pooling without the problems potentially caused by motion opponency.

Two aspects of the experimental design promoted the reliance of the observer on planar power detectors. The first aspect was that the combination signal had a large range of spatial orientations. Using a large range of orientations encourages the observer to use an internal filter which is broadband in orientation, like a planar power detector. The second aspect was that all the combinations of the plaid and bandpass signals were randomly intermixed. Random intermixing encourages the observer to use a strategy which combines across the ensemble of signals presented. For the experimental setup, additivity is not the optimal combination rule. The ideal observer in this task uses power detectors which are matched to the plaid spectrum and the bandpass spectrum, computes the energies in each signal band for a large set of possible plaid and bandpass power levels, and averages the exponentiated energies across the set of signal power levels. However, if the observer's most sensitive detectors for the task are planar power detectors, the best the observer can do is to sum together the bands in the plane weighted by the expected power for the band.

4.1.3 Data Presentation

The natural way to present the data from the additivity experiments is to plot the constant % correct thresholds in (E_b, E_{pl}) space, illustrated in figure 4.2. Additivity shows up in this type of plot as a straight line with negative slope which connects the plaid alone and bandpass alone thresholds. To assess deviations from additivity we used a generalized summation equation:

$$1 = (E_b/T_{E_b})^\alpha + (E_{pl}/T_{E_{pl}})^\alpha \tag{4.3}$$

$$c = E_b^\alpha + s \cdot E_{pl}^\alpha \tag{4.4}$$

Additivity plots [51, 117] and analogues to equation 4.4 have been previously used to assess additivity. Equation 4.4 occurs naturally in the context of probability summation [50] and as the solution to a functional equation for pooling [132].

For $\alpha > 1$, equation 4.4 leads to subadditive combinations, i.e. to thresholds which are larger than the sum of the component thresholds. Also for integral $\alpha > 1$ this expression can represent probability summation among α different independent bands. For $\alpha < 1$, this equation leads to superadditive combinations, i.e. to thresholds which are smaller than the sum of the component thresholds. The form of the equation was chosen as a simple way to parametrize additive vs. non-additive combinations. By fitting this equation to the resulting thresholds, we can determine the summation rule from the value of α.

4.2 Methods

4.2.1 Stimuli

The method for producing the stimuli was the same as chapter 2. Stimuli were produced by passing spatio-temporal gaussian white noise through the set of filters described below.

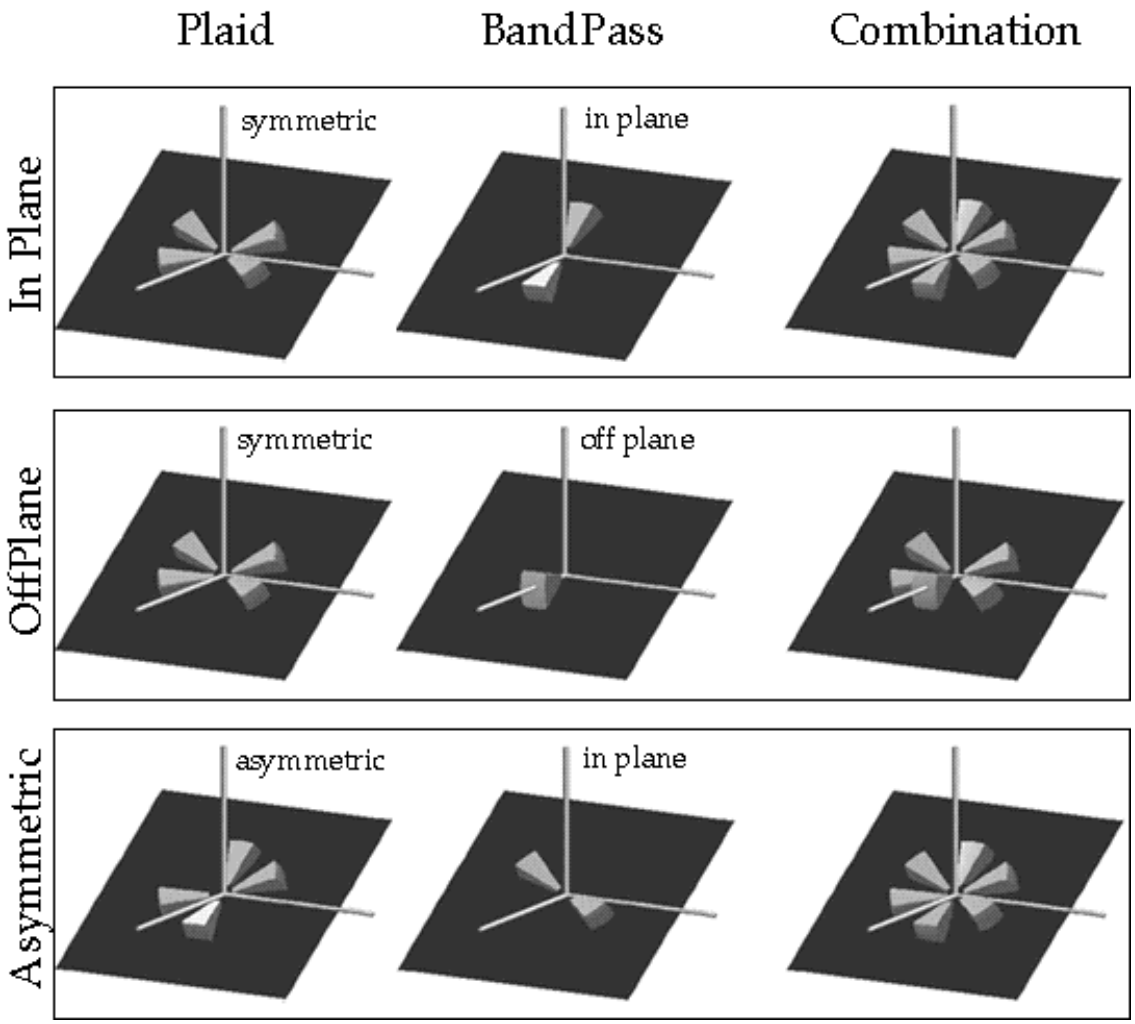

Figure 4.1: Filters used in additivity experiments. The illustrations are presented as a table with the rows corresponding to different additivity conditions and the columns to the bandpass filters and their combination. The three different additivity conditions, In Plane, Off Plane, and Asymmetric designate properties of the stimuli.

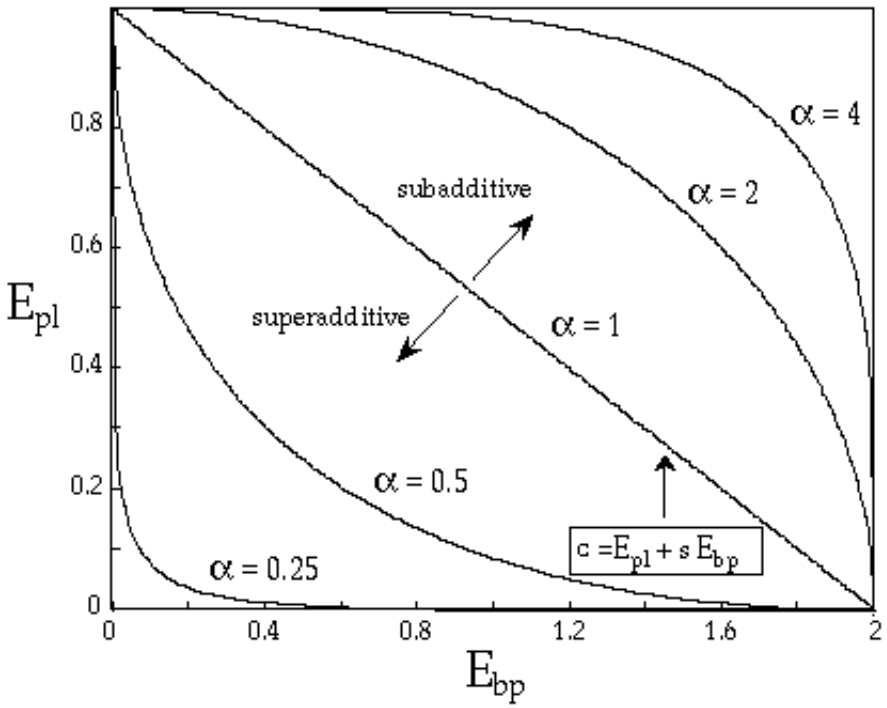

Figure 4.2: Plotting format for additivity experiments.

4.2.2 Filters

All of the filters used were rotated copies of a single bandpass filter. This filter has the following functional form:

$$BP(\omega_r, \omega_\theta, \omega_\phi) = W_r(\omega_r) W_\theta(\omega_\theta) W_\phi(\omega_\phi) \tag{4.5}$$

where W_x is a smooth box function on the variable x (see methods section Chap 2). Smooth box functions were used because they allow fine control over the placement and smoothness of the spectral boundaries. W_r had a transition region width of 1.45, and low-high frequency cutoffs of (0.49,7.6), where the frequency radius of the sphere is given by $\omega_r = \sqrt{\omega_x^2 + \omega_y^2 + (\omega_t/2.1)^2}$. W_θ and W_ϕ had a transition widths of 8 degrees, and the high low cutoffs which spanned 36 degrees.

All of the other filters were simply combinations of 3-D rotated copies of this base filter. If we choose the base position of the BP filter to be centered around the ω_x axis, then we can describe the positions of the other filters by the composition of two rotations of BP. Let $\vec{\omega}$ represent the vector $[\omega_x, \omega_y, \omega_t]$, and $\mathbf{R_x}(\phi_0)$ and $\mathbf{R_t}(\theta_0)$ denote the 3-D rotation matrices which leave the ω_x and ω_t axes fixed respectively. The composition $R_x R_t$ rotates the filter away from the ω_x axis by θ_0 degrees and then away from the spatial frequency plane by ϕ_0 degrees. Most of the filters were rotated up to lie in a common plane which specified a downward motion with a speed of 1.93 deg/sec.

In Plane Condition The symmetric plaid filter is formed by summing together two copies of the BP filter. The rotations angles for R_x are $\theta_0 = 28\,\text{deg}$ and $\theta_0 = 152\,\text{deg}$, which makes the orientations symmetric around the ω_y axis. Each of the filters have R_t rotation angles of $\phi_0 = 36.9\,\text{deg}$ so that the bands lie in a common plane.

The bandpass filter for the Symmetric condition has rotations angles of $\theta_0 = 90\,\text{deg}$ and $\phi_0 = 36.9\,\text{deg}$ so that the plaid and bandpass filters lie on a common plane, with the bandpass spatial orientation orthogonal to the direction of motion specified by the plane.

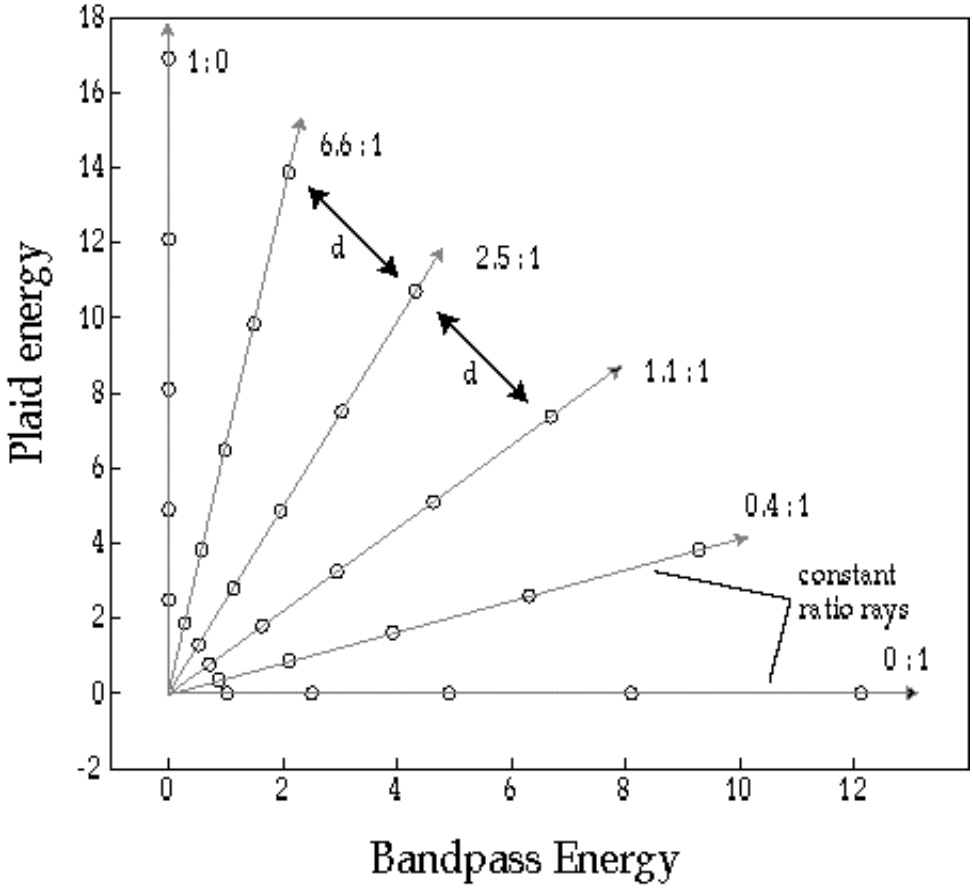

Figure 4.3: Diagram illustrating the data collection method. Data was collected for 6 different constant E_{pl}/E_d ratios shown as gray arrows. The ratios are presented to the right of the arrows. The gray circles represent the points the data was collected at. The ratios were chosen so that the distances d, between points along the diagonal are equal on normalized energy axes. The data was analyzed by fitting Weibull functions to the data along each constant ratio ray.

Asymmetric Condition The asymmetric plaid filter is also formed by summing together two copies of the BP filter. The rotations angles for R_x are $\theta_0 = 28\,\text{deg}$ and $\theta_0 = 90\,\text{deg}$. The bandpass filter for this condition has $\theta_0 = 152\,\text{deg}$. Each of the filters have R_t rotation angles of $\phi_0 = 36.9\,\text{deg}$ so that the bands lie in a common plane.

Off Plane Condition In this condition the plaid filter is the same as in the In Plane condition. The bandpass filter has the same $\theta_0 = 90\,\text{deg}$, but it does not lie in the plane. $\phi_0 = 3\,\text{deg}$ for this condition, hence the filter is nearly centered around the spatial frequency plane.

4.2.3 Procedures

Data were collected using a 2IFC task, in which subjects discriminated signal plus noise and noise alone intervals. Signal stimuli in each condition were additive mixtures of one of the plaid stimuli with one of the bandpass stimuli. Data from each condition was collected in separate sessions, but the sessions were intermixed. Subjects were provided with knowledge of which stimuli were to be detected at the beginning of each session. Subjects were also given two hours practice on each condition prior to data collection.

Figure 4.3 illustrates the data collection method. Signal energy was varied using the method of constant

Subject	*Condition*	$\mathbf{E}_{pl}/\mathbf{E}_{bp}$					
PS	In Plane	1:0	6.6:1	2.5:1	1.1:1	0.4:1	0:1
	Off Plane	1:0	5.8:1	2.2:1	0.96:1	0.36:1	0:1
	Asymmetric	1:0	6.0:1	2.3:1	1.0:1	0.38:1	0:1
ML	In Plane	1:0	6.4:1	2.4:1	1.1:1	0.4:1	0:1
	Off Plane	1:0	4.5:1	1.7:1	0.7:1	0.3:1	0:1
	Asymmetric	1:0	5.9:1	2.2:1	1.0:1	0.38:1	0:1

Table 4.1: Table of constant ($\mathbf{E}_{pl}/\mathbf{E}_{bp}$)

stimuli for six different constant plaid-bandpass energy ratios ($\mathbf{E}_{pl}/\mathbf{E}_{bp}$), shown in the figure as grey rays. Five different combination energies were used to estimate the psychometric function along each ray, shown as open circles in the figure, for a total of 30 different combinations. At each combination, 100-120 trials were collected. It required $3\frac{1}{2}$-4 hours to collect all the trials for a condition. To avoid subject fatigue the data collection for each condition was split into one hour sessions.

To insure that the constant ratio rays were evenly distributed, we used estimates of the subject's thresholds for plaid and bandpass stimuli alone to distribute the measurements across the $(\mathbf{E}_{pl}, \mathbf{E}_{bp})$ plane. We measured the subject's thresholds for each of the plaid and component stimuli alone using the method of constant stimuli. We then determined the ratios which caused the constant ratio rays to divide the line connecting the 80% correct $\mathbf{E}_{pl}$ and $\mathbf{E}_{bp}$ estimates into equal length segments. Thus the ratios were different for each subject and condition. The energy ratios used are presented in table 4.1.

Thresholds were determined by fitting Weibull functions to the detection data along each constant ratio ray using a maximum likelihood procedure. Error bars for the thresholds were computed from the inverse numerical Hessian of the likelihood function for threshold, which were cross-validated using a parametric bootstrap procedure. In the bootstrap procedure, 1000 data sets were simulated by sampling from the binomial distribution with the parameter p given by the measured probability correct. Maximum likelihood fits of the parameters were then generated for each data set. The resulting distributions of fitted parameters were used to estimate the standard error on the parameters.

4.2.4 Data Analysis

Additivity was assessed by fitting the following equation to constant %correct threshold points along each of the constant ratio rays:

$$\mathbf{E}_{bp}^{\alpha} + (s\mathbf{E}_{pl})^{\alpha} = c^{\alpha} \tag{4.6}$$

This equation represents a Minkowski metric model of pooling [50] which frequently arises in the context of probability summation models. In the present context it provides a simple means for parametrizing additivity through the exponent α. $\alpha < 1$: superadditive, $\alpha = 1$: additive, $\alpha > 1$: subadditive. The 'slope' s, gives us a measure of the relative weights given to the plaid and bandpass energies, when additivity holds.

The model was fit to the data using non-linear least squares minimization. The squared distances along the constant ratio rays between the measured threshold energies and the curve described by the equation were inversely weighted by the variances of the threshold estimates. The sum of these weighted distances were minimized over three parameters, α, s, and c using the Broyden-Fletcher-Goldfarb-Shanno variable metric multidimensional minimization method [102].

Statistics on the fits were generated using a parametric bootstrap procedure. In the procedure, bootstrap fits of the psychometric functions were used to generate 1000 estimates of each of the energy thresholds. Least squares fits for each of the estimates was performed, generating distributions for α, s, and c.

Variances for the parameter estimates were computed from these distributions. One way ANOVAs and T-tests were performed on the parameter estimates using these variance estimates as the within-condition variances. Since we could use as many bootstrap samples as desired, the within-condition degrees of freedom were effectively infinite. We used a large positive number 10^5, instead of infinity for the number of degrees of freedom.

4.3 Results

4.3.1 In Plane Results

The results for the In Plane condition are shown in figure 4.4, plotted in plaid-bandpass threshold energy space. Each data point represents the threshold energy along a constant ratio ray for one of four different % correct values: 60%, 70%, 80% and 90%. The error bars are oriented along the constant ratio rays, and were estimated from the psychometric function fit.

The dashed lines represent approximate constant performance contours, while the solid lines represent the curves generated by the best fitting parameters of the pooling equation. When additivity holds the performance contours should lie along straight lines, or equivalently, the best fitting additivity equation exponents α should be 1. We see that for both subjects the best fitting curves are essentially linear. The best fitting αs for each constant performance curve are gathered into a table on the right side of the figures. Inspection shows that all of the αs are very close to 1. Subject PS shows a small but consistent trend for α to increase with increasing % correct, however, none of the alpha are significantly different from 1 (T-test, 0.05 level). The alpha for Subject ML do not show an increased trend, and are clustered more tightly around 1. Thus the visual system can be described as additively pooling the bands in this condition.

4.3.2 Asymmetric Results

The results for the Asymmetric condition are similar to the On Plane condition, as predicted by the planar power detector model. The α estimates for subject PS are 0.2-0.3 higher than in the On Plane condition, however, none of the α are significantly different from 1 (T-test, 0.05 level). Thus the visual system is able to additively pool power as long as the components lie on a common plane. It also shows that subjects are not using a detection approach which requires the phenomenal motion of the plaid and bandpass components to be the same for additive pooling to occur.

4.3.3 Off Plane Results

The results for the Off Plane condition are very different. Notice that the best fitting curves are curved in the subadditive direction. T-tests show that all of the α for this condition are significantly greater than 1 at the 0.01 level. Thus subjects are not able to additively pool bandpass power off the common plane specified by the symmetric plaid. This result agrees with the results from Chapter2, which showed that subjects are less efficient at detecting non-planar stimuli.

There is a highly significant trend for α to decrease as % correct increases for both subjects, which is discussed below.

4.3.4 Additivity exponents

The fitted additivity exponents α are summarized in figure 4.7. As noted before, the salient feature of the data is that exponents are clustered around 1 for the In Plane and Asymmetric conditions, but are

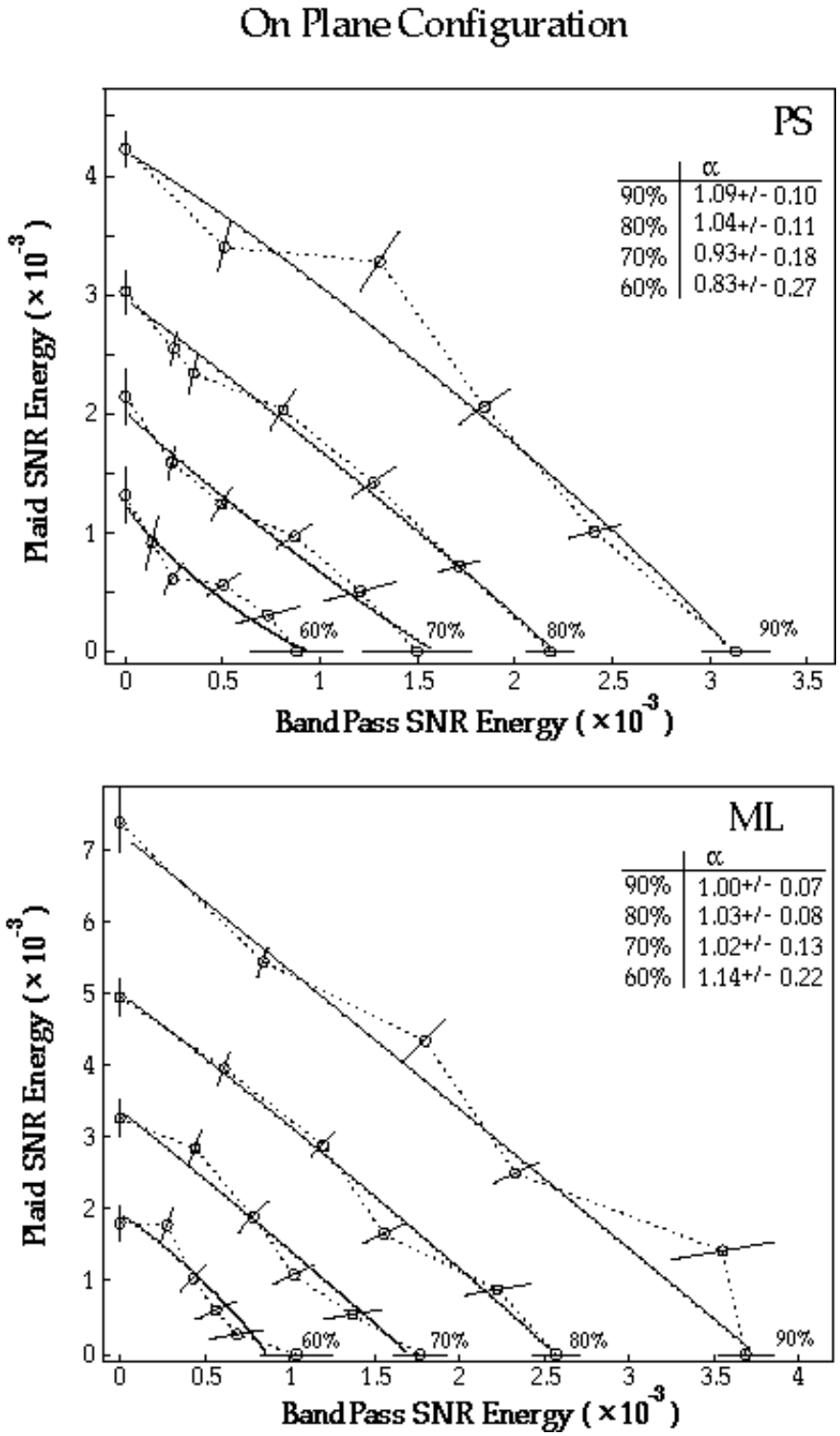

Figure 4.4: In Plane additivity data. None of the estimated α are different from 1 at the 0.05 level using a T-test.

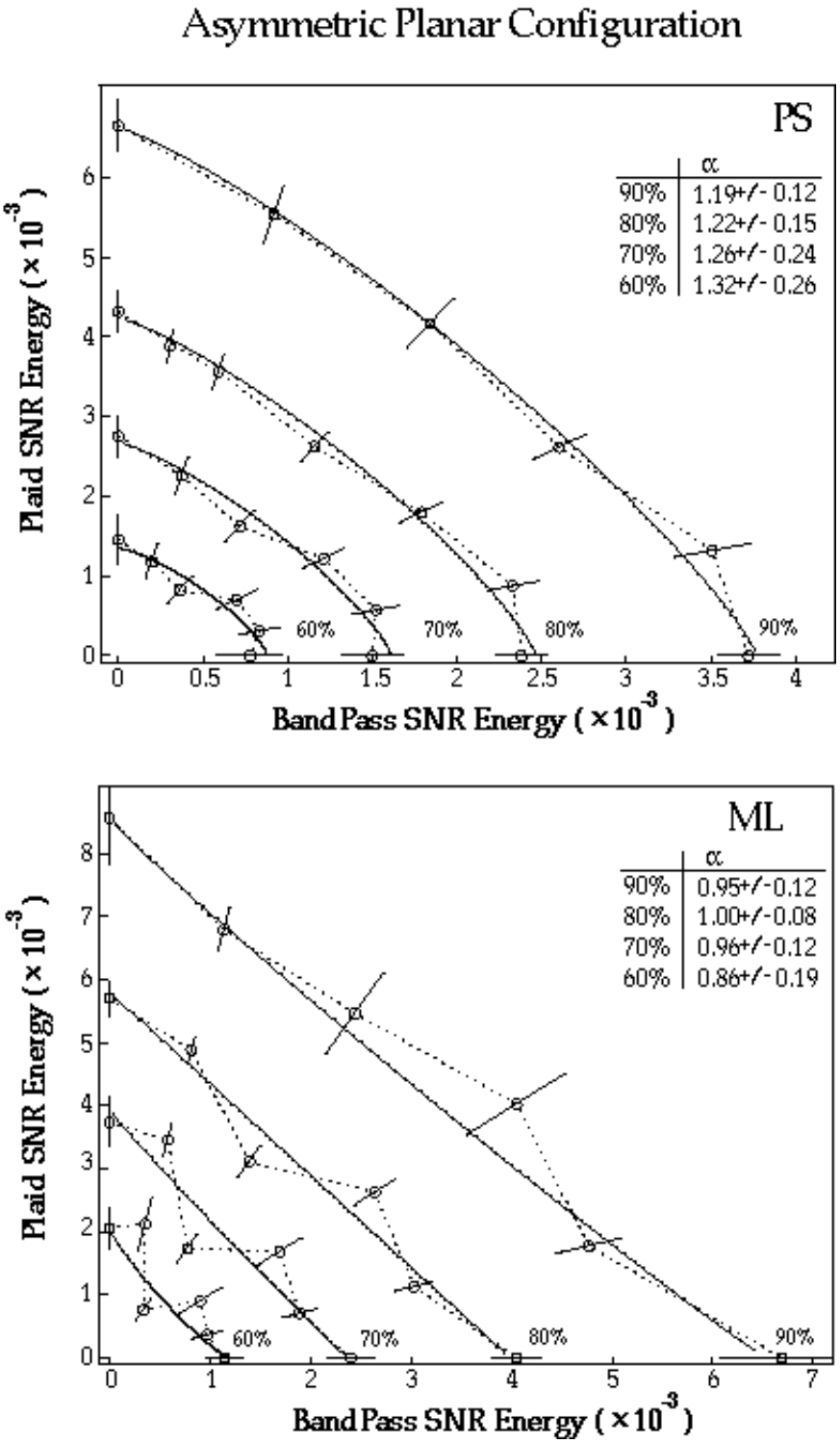

Figure 4.5: Asymmetric additivity data. None of the estimated α are different from 1 at the 0.05 level using a T-test.

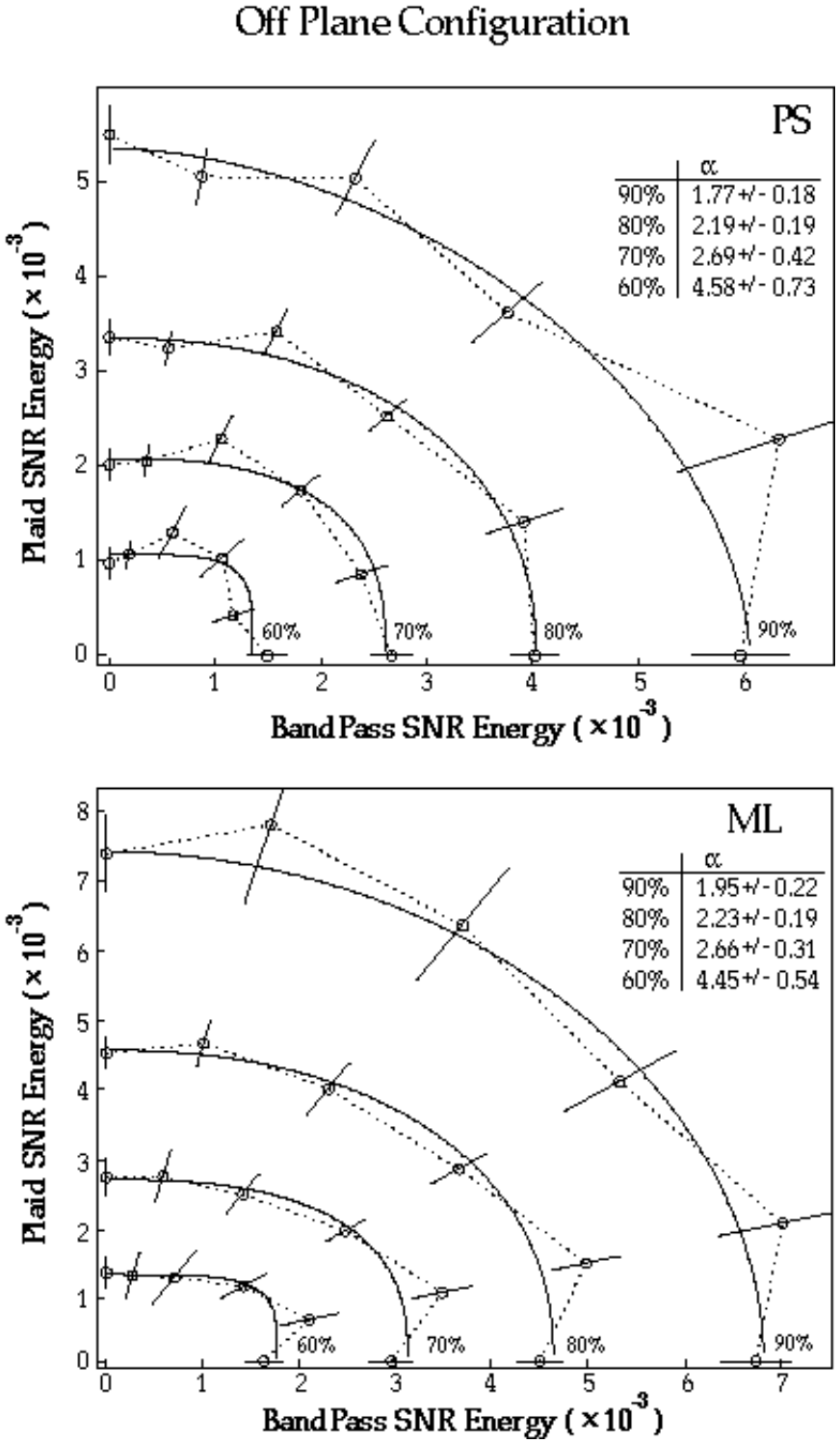

Figure 4.6: Off Plane additivity data. All of the estimated α are significantly different from 1 at the 0.001 level using a T-test.

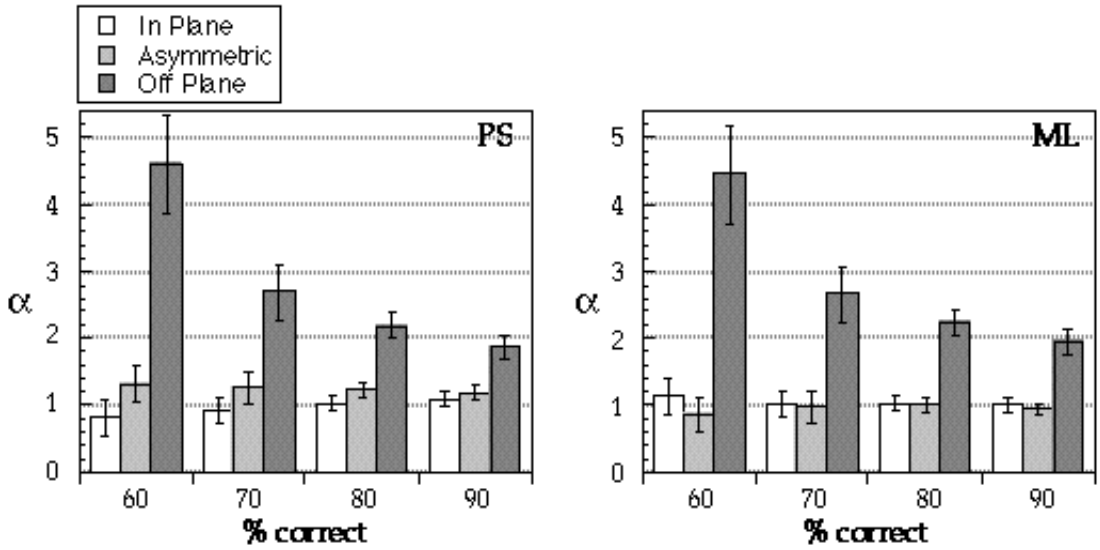

Figure 4.7: Additivity exponents for the three conditions.

Subject	*Comparison*	p: 60%	p: 70%	p: 80%	p: 90%
PS	**In Plane** vs. **Asymmetric**	0.19	0.29	0.26	0.48
	In Plane vs. **Off Plane**	< 0.001	0.003	< 0.001	0.001
	Asymmetric vs. **Off Plane**	< 0.001	0.008	< 0.001	0.006
ML	**In Plane** vs. **Asymmetric**	0.33	0.75	0.77	0.71
	In Plane vs. **Off Plane**	< 0.001	< 0.001	< 0.001	< 0.001
	Asymmetric vs. **Off Plane**	< 0.001	< 0.001	< 0.001	< 0.001

Table 4.2: Table of α comparisons across condition at each probability correct, showing the probability that the compared αs are drawn from the same distribution using the T-test statistic. The Off Plane condition is significantly different from the other two conditions at better than the 0.01 level for both subjects.

significantly greater than 1 for the Off Plane condition. The planar power detector model predicts that the pooling rule will be additive and hence equivalent for conditions In Plane and Asymmetric, but will be significantly non-additive for the Off Plane condition. We performed a set of T-tests of the α estimates between the conditions at each % correct level, the results of which are presented in table 4.2. The analysis shows that the two conditions in which the bands are coplanar, In Plane and Asymmetric, are not significantly different (max $p = 0.19$). In contrast, both of these conditions are significantly different ($p < 0.01$) from the Off Plane condition.

Invariance across %correct

In our analysis of the ideal weighting strategy we predicted that the additivity result should hold equally for any % correct slice we choose to analyze. We tested for invariance of the fitted pooling equation parameters across the four %correct levels by performing a one-way ANOVA. The results of the analysis are gathered in table 4.3. The estimates of α across % correct are not significantly different at the 0.05 level for the In Plane and Asymmetric conditions for either subject. Thus for the conditions in which all of the bands lie in a single plane, the assumption of additivity independent of the level of performance cannot be falsified.

There is a significant trend in the Off Plane data which is nearly identical for both subjects. The trend is toward decreasing α with increasing % correct. The trend could be the result of pooling efficiency

Subject	*Condition*	F(3,∞)	*p*
PS	In Plane	0.14	0.94
	Asymmetric	0.03	0.99
	Off Plane	2.72	0.043*
ML	In Plane	0.07	0.98
	Asymmetric	0.07	0.98
	Off Plane	4.98	0.002*

Table 4.3: Table of ANOVA results for the estimates of α across % correct. p gives the probability that each of the α are drawn from the same distribution, i.e. that there is no significant change in alpha across the % correct slices. The Off Plane condition is significant at the 0.05 level for both subjects, shown by the starred p values.

changing with % correct, such that subjects pool more efficiently at higher signal levels. If this hypothesis were true we would expect the zero mixture Plaid thresholds for the In Plane and Off Plane conditions to be roughly equivalent and for the mixture thresholds in the Off Plane condition to improve with % correct. Expressed in terms of the slopes of psychometric functions, it predicts that the zero mixture plaid psychometric slopes will be equivalent while the non-zero mixture will be shallower in the Off Plane condition.

Instead we find that the Weibull slope parameters β along the plaid only axis are steeper for the On Plane than Off Plane conditions (PS: $\beta_{OnPlane} = 1.69$, $\beta_{OffPlane} = 1.14$, ML: $\beta_{OnPlane} = 1.4$, $\beta_{OffPlane} = 1.13$), while the average mixture slopes are essentially equivalent (PS: $\beta_{OnPlane} = 1.57$, $\beta_{OffPlane} = 1.52$, ML: $\beta_{OnPlane} = 1.4$, $\beta_{OffPlane} = 1.49$). The decreased slope of the plaid energy psychometric function in the Off Plane condition is the result of lower plaid thresholds at the low signal level/low % correct end in the Off Plane condition than the In Plane condition, and slightly higher thresholds at the high signal level/high % correct end. This suggests that the plaid stimuli are being processed differently in the presence of the Off Plane bandpass stimuli than the In Plane bandpass stimuli. The result suggests that the source of the variation across alpha is in the processing of the plaid component stimuli. To test this idea, we refit the Off Plane data using the plaid only mixture data from the On Plane condition. The resulting α estimates showed no trend across % correct, PS: $\alpha_{mod} = (2.7, 2.6, 2.6, 2.7)$, ML: $\alpha_{mod} = (2.1, 1.9, 1.9, 2.0)$.

4.3.5 Weighting across bands and the fitted slopes

Given the result of additivity, the slopes for the In Plane and Asymmetric conditions can tell us something about the relative weighting across the three bandpass components which comprised the mixtures in the two conditions. The slopes are given by the parameter s in the pooling equation 4.4, and the fitted slopes are shown in figure 4.8.

We can use the slopes in the In Plane and Asymmetric conditions to derive estimates of the weights across the three stimulus bands. Let w_a denote the weight for the Asymmetric bandpass component, w_b the weight for the In Plane bandpass component, and w_c the weight for the remaining band. An equivalent rule for naming the bands is to label them $\{c, b, a\}$ going counterclockwise in spatial orientation from the ω_x axis. Then in the In Plane condition, the slope s_{in} in $s_{in}\mathbf{E}_{bp} + \mathbf{E}_{pl} = c$ is equal to

$$s_{in} = \frac{w_b}{(w_a + w_c)/2} \tag{4.7}$$

Thus $(w_a + w_c) = s_{in}/2w_b$. In the case $w_a = w_c$, then these weights are 0.74 for subject PS and 0.48 for

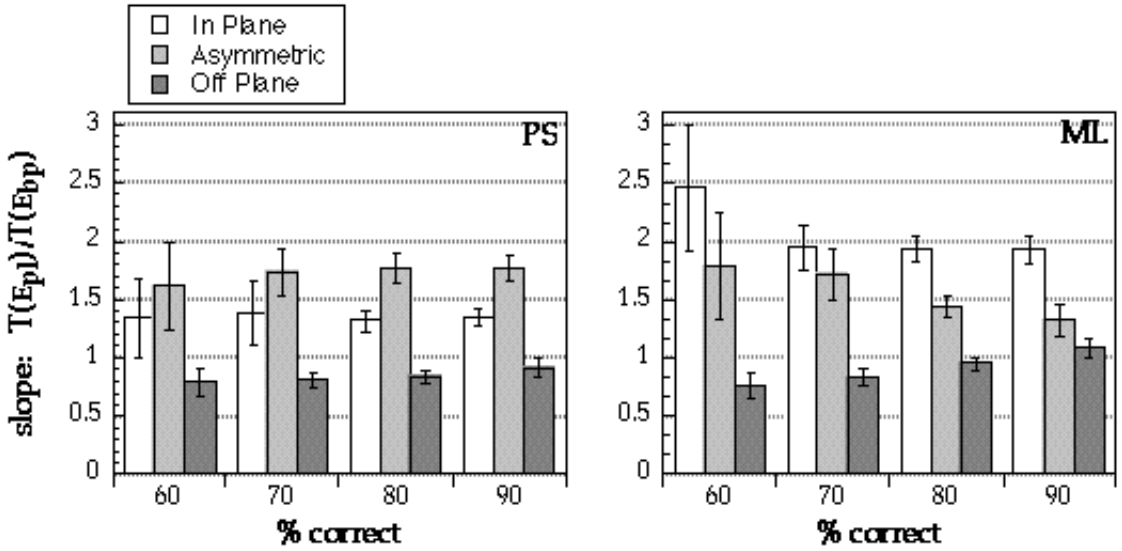

Figure 4.8: Slopes of the fits the three conditions. A one-way ANOVA of the slopes across % correct shows no significant effect for either subject in any condition (largest probability is 0.43).

subject ML, relative to $w_b = 1$. In the Asymmetric condition the slope is given by:

$$s_{asym} = \frac{w_a}{(w_b + w_c)/2} \tag{4.8}$$

If we assume that the weighting is identical in the two experiments and set $w_b = 1$, then we can use equations (4.7) & (4.8) to solve for the weights for w_a & w_c relative to w_b. The resulting weights for both subjects are gathered in table 4.4.

Instead of an equal split between the weighting as expected by symmetry, assuming the invariance of weights across conditions leads to a large bias in the weighting of bands corresponding to leftward and downward moving gratings over those moving rightward and downward. However, to derive the weights, we only considered the slopes and not the magnitude of the thresholds. If both In Plane and Asymmetric weights are equal, then the threshold energies at a given %correct should also be equal, since the conditions obey equation 4.2. Thus, we can use the common weights to compute a predicted Asymmetric bandpass threshold from the In Plane threshold energy at a given % correct by multiplying w_a by the In Plane plaid threshold. We find that the actual thresholds in the Asymmetric condition are 27% & 34% higher on average than the predictions from the In Plane condition for subjects PS and ML respectively. Using the variance of the Asymmetric bandpass thresholds to construct a T-test, we found that these thresholds are significantly different from the predictions at the $p < 0.02$ level. Thus, it seems likely that the assumption that the weights are constant across the two conditions is false.

The other possibility is that subjects are able to adjust their weighting across bands, using different sets of weights for the In Plane and Asymmetric conditions. Subjects will perform better in the experimental conditions if they weight the three frequency bands by scalars proportional to the expected energy in each band. Because the additivity experiments involved intermixing different sets of stimuli in the In Plane and Asymmetric conditions, the expected energies, and hence the optimal weights are different for the two conditions. We computed the optimal weights for each subject (see Appendix B) in each condition, shown in table 4.4. The weights show that subjects can do better on average if they weight the bandpass band more heavily than the plaid bands. If we assume that subjects equally weighted the plaid bands in each condition (i.e. In Plane:$w_a = w_c$, Asymmetric:$w_b = w_c$), we can compare the optimal weights with

Subject	*Condition*	*Hypothesis*	w_a	w_b	w_c
PS	In Plane	Optimal weights	0.65	1	0.65
		Adaptable weights	0.74	1	0.74
		Fixed weights	1.15	1	0.33
	Off Plane	Optimal weights	1	0.61	0.61
		Adaptable weights	1	0.58	0.58
		Fixed weights	1.15	1	0.33
ML	In Plane	Optimal weights	0.63	1	0.63
		Adaptable weights	0.52	1	0.52
		Fixed weights	0.86	1	0.11
	Off Plane	Optimal weights	1	0.61	0.61
		Adaptable weights	1	0.67	0.67
		Fixed weights	0.86	1	0.11

Table 4.4: Estimates of the relative weights across the bands in the plane under two different hypotheses, the weights are fixed across the In Plane and Asymmetric conditions or the weights are adaptable to the demands of each condition. The weight estimates are presented below the optimal weights for the condition (see Appendix B). w_a and w_c correspond to the bands in the plane oriented rightward and leftward of the direction of motion (downward) respectively, while w_b corresponds to the band whose orientation coincides with the movement direction. Assuming that both the weights are adaptable and equal weighting of bands within the plaids leads to weight estimates which are not significantly different from the optimal weights. Assuming the weights are fixed across the conditions leads to a strong weight bias toward leftward moving bands, which we conclude is less likely. See text for details.

the weights estimated from the slopes. The estimated weights are shown in table 4.4 in the rows labeled 'Adaptable weights'. None the Adaptable weights are significantly different from the optimal weights (T-test, 0.05 level), which suggests that subjects may be adapting their weights to optimize performance in the two different conditions.

4.4 Discussion and Conclusions

We have shown that planar configurations of power are additively pooled. This result constitutes the strongest evidence we are aware of for the existence of planar power detectors. The three conditions together verify that planar configurations are special, while the additive law suggests that the detectors are indeed *power* detectors. In terms of contrast, additivity in power (or energy) is a quadratic summation rule. It is possible that the results are actually due to a nonadditive summation rule acting on contrast which has been subjected to a non-linearity other than squaring, such that the conjunction of the summation rule and the non-linearity conspire to produce the apparent additivity. However, the exact cancellation that this would require seems implausible. Regardless, the net result is that configurations of power lying on a common plane obey a summation rule which is equivalent to the rule used by planar power detectors, while configurations of power which do not lie on a common plane are detected using a suboptimal rule.

For the Off Plane condition, the exponents indicated subadditivity. A simple hypothesis is that the plaid and bandpass stimuli are independently processed and pooled by probability summation.

4.4.1 Context sensitive weighting across the plane

The results suggested that observers are adapting their weighting strategy across bands which lie in a common plane to the demands of a task. These adaptations are modest, involving only 25-30% changes

in the values of the weights, and hence do not contradict the conclusion that the visual system has a limited capacity to adapt the weights across bands. One way to account for the adaptability of weighting is to postulate the existence of a large set of fixed planar power detectors, each with different orientation sensitivities. The adaptability of performance could then be attributed to the observer relying on detectors with different spatial sensitivities in the two conditions. This possibility leads to the confounding of spatial structure and velocity within a detector, which must then be disambiguated by making separate estimates of the spatial structure to avoid image structure dependent biases in the encoded velocities.

Another way to account for adaptability is to postulate planar power detectors in which the weighting across spatial orientation can be modified by perceptual learning. This possibility would be a way of implementing optimal velocity estimators for the stochastic stimuli. Optimal velocity estimation can be implemented using a population of detectors tuned to different planes in frequency space. The spectrum around which the detector pools will be given by the expected spatial frequency spectrum projected unto the plane specified by the velocity. If the visual system uses takes this expectation over a window of time on the order of the experiments, then the adaptability could be accounted for by an attempt at optimality (i.e. the development of perceptual expertise).

Neurally, this possibility could involve the feedback from some site which modifies the weighting based on the pattern of correct and incorrect decisions. The idea that the visual mechanisms are flexible and can be modified by perceptual learning has been recently suggested by several researchers as a means of accounting for the stimulus and task specificity of several kinds of perceptual expertise (e.g. spatial hyperacuity)[25, 3, 30, 36, 41]. Physiological evidence for receptive field structure being modified by feedback also exists. Recently, McLean & Palmer[84] have shown that the phase selectivity of neurons in primary visual cortex can be modified by associative learning, and were able to change the direction selectivity in one cell in ten. In audition, Weinberger and colleagues [34, 144, 39] have presented evidence that the peak frequency of auditory neurons in the guinea pig and cat could be modified by classical conditioning. In visual area MT in monkeys, Treue (1996)[118] showed that neuronal firing rates could be modulated by attention, suggesting that MT neurons have the potential to be modified by feedback from higher stages of visual processing.

In addition to the difference in On Plane and Asymmetric weighting, we found evidence that the same plaid stimuli are processed differently in the In Plane and Off Plane conditions. This difference cannot be explained by an instantaneous interaction between the plaid and the bandpass stimuli, since the trials we are considering did not include the bandpass stimuli. Due to the subadditive combination in the Off Plane condition, we might expect some elevation of the plaid thresholds over the In Plane condition. For instance, the bandpass stimuli in the Off Plane condition may increase the number of irrelevant detectors the subject attends to in detecting the plaid stimuli (increased signal uncertainty)[96], or it might cause the visual system to use a detector which is more poorly matched to the plaid signal than in the In Plane condition. The latter possibility is less plausible given the fact that plaid thresholds are actually lower for smaller %correct values in the Off Plane condition. Instead, the lower thresholds suggest that subjects may actually be using detectors which are better matched to the plaid stimuli in the Off Plane condition, but are attending to many irrelevant detectors. This possibility could produce the pattern of results, because at low signal values the irrelevant detector responses will increase the decision variable variance less than the positive weighting of the middle band would in the In Plane condition. We plan to investigate the possibility that the frequency weighting for plaid stimuli differs in the two conditions using the perturbation analysis from last chapter.

4.4.2 Relations to physiology

There are a number of studies which suggest that the most probable location of planar power detectors in the brain is in the human analogue to simian visual cortical area MT. Simoncelli and Heeger (1998) [113] have recently modeled a great deal of electrophysiological data recorded from this area with a modified planar power detector model. A subset of the neurons in MT, designated either 'pattern motion selective'[90] or 'Type II'[8, 107] have several of the properties expected for planar power detectors. This set of MT neurons are tuned to speed and direction [80, 106, 73] and are relatively insensitive to the spatial pattern characteristics[90, 9]. Movshon et al. has also shown that the spatial and temporal frequency tuning in some of these cells covary, which is required for pattern invariant speed tuning[91]. Lesion studies show that MT neurons are critically involved in both the computation of direction of motion[93] and speed [95], while electrical stimulation can produce directional biases in a perceptual task[109].

In addition, motion opponency for MT neurons similar to that found in the perturbation analysis has been reported by several investigators[86, 106, 114, 103, 149, 150]. The opponency is manifested as a suppression of neuronal firing rates for motions in the direction opposite to the cell's preferred direction. This inhibition has a component coextensive with the classical receptive, best demonstrated by experiments of Qian and Anderson [103], who showed that superimposing random dot patterns suppressed firing rates, especially when the opposite moving dots were locally paired in space. Snowden (1991) showed that this inhibition is well modeled by a divisive interaction and Mikami et al.[86] showed that the inhibition can be maximal for opponent motions which have speeds different than the cells preferred speed. Finally, several researchers have shown that there is a strong inhibitory interaction for motions outside of the cells classical receptive field[86], and that this inhibition is structured but spatially inhomogeneous[149, 150, 148].

The existence of spatially inhomogeneous interactions outside the classical receptive has been reported by many researchers[10, 116, 17, 149, 150]. On the basis of these findings all of these researchers have suggested that area MT may be involved in computations more complicated than simple image velocity estimation, such as figure/ground segmentation, motion in depth, and computing surface curvature and orientation in depth[148]. In addition, it has been shown that MT neurons are also tuned to binocular disparity[80], which has led to the speculation of a role for MT in motion in depth computations [10]. Finally, there are suggestions that MT neurons may play a role in determining targets for tracking eye movements [42]. These additional properties of MT neurons could be used to better assess experimentally whether the psychophysically defined planar power detectors have their basis in visual area MT.

4.4.3 Model for velocity estimation

The preceding experiments have suggested the outlines of the properties of velocity detectors. However, the visual system is more often faced with the problem of *estimating* image velocities than *detecting* image movement. In this section we discuss how the results fit into a velocity estimation scheme, and outline some possibilities for further testing.

Velocity estimation is a simple extension of translation detection. We propose a model based on the experimental evidence and a set of assumptions motivated by optimal velocity estimation in conditions of uncertainty. We assume: 1) The visual system makes local estimates of velocity. 2) The visual system makes use of the expected spatial structure of the signal in making velocity estimates.

An interesting property of the translation detectors we have been discussing is that a population of these detectors can be used to construct a likelihood function for image velocity. Thus the detectors could be part of a general system for estimating local image velocity. The visual system can compute optimal local estimates of velocity by using the expected spatial structure of the image, and the prior distribution of image velocities.

Let $\hat{S}_e(\omega_x, \omega_y)$ denote the expected spatial spectrum of the signal, and $W(\omega_x, \omega_y, \omega_t)$ represent the spectrum of the localization window. Then if the texture is moving with velocity $\vec{v}$, the expected signal spectrum is given by:

$$S(\vec{\omega}, \vec{v}) = \left(\hat{S}_e(\omega_x, \omega_y)\delta(\omega_t + \vec{v} \cdot \vec{\omega}_{sp})\right) \bigotimes W(\omega_x, \omega_y, \omega_t) \tag{4.9}$$

This equation represents the expected signal spectrum projected onto the plane given by the velocity and blurred (convolved) by the localization function. Given $S(\vec{\omega}, \vec{v})$ we can construct a family of detectors tuned to different values of $\vec{v}$. The likelihood function over velocity can be approximated by computing the inner product of the input signal power spectrum $X(\vec{\omega})$ with a set of detectors each with a different fixed values of $\vec{v}$.

$$\log L(X|\vec{v}_i) \simeq \int_{\vec{\omega}} |S(\vec{\omega}, \vec{v}_i)|^2 |X(\vec{\omega})|^2 d\vec{\omega} \tag{4.10}$$

where we have dropped out the scaling factors for simplicity, and used the small signal approximation.

A maximum (or mean) a posteriori estimate of the velocity can be made from this sampled likelihood function by introducing a prior distribution on velocity.

$$p(\vec{v}|X) = \frac{L(X|\vec{v})p(\vec{v})}{\int L(X|\vec{v})p(\vec{v})d\vec{v}} \tag{4.11}$$

A likely prior on velocity would be a bias for slower speeds. As pointed out by Simoncelli (1993), a bias for slower speeds could explain several perceptual phenomena, including the wagon wheel effect and the fact that one-dimensional signals are typically seen to move in the direction orthogonal to their spatial orientations.

One of the interesting predictions of this sort of model is that speed and direction discrimination based on these filters should obey a Weber's law in speed, which has been previously shown[20].

4.5 Appendix

The planar power detector additively pools power lying on a common plane to measure the spectral energy E around the plane. If we split the plane into several bands, then the planar power detector can be described as adding the output energies within each of the bands.

$$\begin{aligned} E &= \int_{\vec{\omega}} |P(\vec{\omega})|^2 |S(\vec{\omega})|^2 d\vec{\omega} \\ E &= \int_{\vec{\omega}} |P(\vec{\omega})|^2 \sum_i |S_{b_i}(\vec{\omega})|^2 d\vec{\omega} \\ E &= \sum_i \int_{\vec{\omega}} |P(\vec{\omega})|^2 |S_{b_i}(\vec{\omega})|^2 d\vec{\omega} \\ E &= \sum_i w_i E_{b_i} \end{aligned} \tag{4.12}$$

Where E is the output energy, $|P(\vec{\omega})|^2$ is the planar power detector spectrum, $|S(\vec{\omega})|^2$ is the signal spectrum, and $|S_{b_i}(\vec{\omega})|^2$ are the band pass components of the signal spectrum. The weights w_i represent the effect the Planar filter has on the energy within each signal band i.

Next we show that the performance of the ideal observer has the form

$$p(R_i = 1) = \Psi(\sum_i w_i E_{b_i}) \tag{4.13}$$

If we fix $p(R_i = 1) = p_0$, then $\sum_i w_i E_{b_i}$ must also be a constant c. Thus the equal performance contours will lie on hyperplanes in the space of energies within the signal bands, E_{b_i}.

In Chapter 1 we gave an expression for the ideal performance,

$$p(X_i = 1) = \Phi(0, \mu_{H1} - \mu_{H0}, \sigma_{H1}^2 + \sigma_{H0}^2) \tag{4.14}$$

where the mean and variance depend on the signal and receiver filter spectra, and the signal and background noise power levels:

$$\begin{aligned} \mu &= \mu_{H1} - \mu_{H0} = 2s \int_{\vec{\omega}} |P(\vec{\omega})|^2 |S_n(\vec{\omega})|^2 d\vec{\omega} \\ \sigma^2 &= \sigma_{H1}^2 + \sigma_{H0}^2 \\ &= 8(s^2 \int_{\vec{\omega}} |P(\vec{\omega})|^4 |S_n(\vec{\omega})|^4 d\vec{\omega} + 2s\mathbf{N} \int_{\vec{\omega}} |P(\vec{\omega})|^4 |S_n(\vec{\omega})|^2 d\vec{\omega} + 2\mathbf{N}^2 \int_{\vec{\omega}} |P(\vec{\omega})|^4 d\vec{\omega}) \end{aligned} \tag{4.15}$$

where $s|S_n(\vec{\omega})|^2 = |S(\vec{\omega})|^2$, i.e. $|S_n(\vec{\omega})|^2$ is the normalized signal spectrum and s is the signal power level. The mean is simply proportional to the energy, and hence is linear in the energies in the bands. The variance can be shown to be linear to an extremely good approximation. In the experiments, $s \ll \mathbf{N}$ and hence the first term in the variance can safely be dropped. $\int_{\vec{\omega}} |P(\vec{\omega})|^4 d\vec{\omega} = k_p$ evaluates to a constant k_p. To evaluate the middle term we notice that when $|S_n(\vec{\omega})|^2 = \sum_i |S_{b_i}(\vec{\omega})|^2$, the term $\int_{\vec{\omega}} |P(\vec{\omega})|^4 |S_n(\vec{\omega})|^2 d\vec{\omega}$ evaluates to $\sum_i a_i w_i E_{b_i}$, i.e. weighting the signal spectrum by the square of the planar filter spectrum simply scales the energy by a fixed amount a_i.

Thus the mean and variance can be written:

$$\begin{aligned} \mu &= 2 \sum_i w_i E_{b_i} \\ \sigma^2 &= 16(\sum_i a_i w_i E_{b_i} \mathbf{N} + k_p \mathbf{N}^2) \end{aligned} \tag{4.16}$$

When the filter overlap on each of the signal bands is identical, the inner products $\int_{\vec{\omega}} |P(\vec{\omega})|^4 |S_{b_i}(\vec{\omega})|^2 d\vec{\omega}$ are the same for all the bands, and hence all the $a_i = a$ are the same. This is true for two of the conditions in the experiment, the In Plane condition and the Asymmetric condition.

Thus for the conditions in this experiment, the ideal performance can be written:

$$\begin{aligned} p(X_i = 1) &= \Phi\left(0, 2 \sum_i w_i E_{b_i}, 16(a \sum_i w_i E_{b_i} \mathbf{N} + k_p \mathbf{N}^2)\right) \\ &= \Psi(\sum_i w_i E_{b_i}, \mathbf{N}) \end{aligned} \tag{4.17}$$

For fixed $\mathbf{N}$ and any fixed probability correct $p(X_i = 1)$, the ideal observer's performance is linear in the energies.

It is important to point out that the condition of identical overlap between signal bands is not very important, since the third term dominates the variance expression.

The previous discussion can be easily adapted to the human observer. We assume that the visual system uses an unknown internal filter which is roughly selective for planar regions of spectral power. In addition, the visual system is subject to additional sensory and internal noise. If we assume that the additional noise is equally distributed across all the bands, then the subject's decision variable variance should be dominated by the effects of the background noise and the sensory and internal noises. In this case both of the signal dependent terms in the variance will be nearly insignificant, and the subjects performance should be approximately linear in the energies in the signal bands.

4.6 Appendix B: Ideal observers for the task

The stimuli in the task can be described by the equations:

$$\begin{aligned} H_1 = \text{signal present:} \quad & \mathbf{r}(x,y,t) = a_1 \cdot \mathbf{s}_{pl}(x,y,t) + a_2 \cdot \mathbf{s}_b(x,y,t) + \mathbf{n}(x,y,t) \\ H_0 = \text{noise alone:} \quad & \mathbf{r}(x,y,t) = \mathbf{n}(x,y,t) \end{aligned}$$

where $\mathbf{s}_{pl}$ denotes the plaid signal waveform and $\mathbf{s}_b$ denotes the bandpass signal waveform. The constants a_1 and a_2 determine the contrast of the signal noises $\mathbf{s}$, hence a_1^2 and a_2^2 are proportional to the signal energies. We will compute the ideal for the case in which the signal energies are much lower than the background noise energy. In the equations, a_1 and a_2 form a random vector since the data are collected by randomly selecting from a set of $[a_1, a_2]$ pairs.

The Bayes decision for the 2AFC task is to choose the interval i with the larger likelihood ratio $L(\mathbf{r})_i$ averaged over the $\vec{a} = [a_1, a_2]$ pairs:

$$\mathrm{E}\left[L(\mathbf{r}|\vec{a})_1\right] \underset{2}{\overset{1}{\gtrless}} \mathrm{E}\left[L(\mathbf{r}|\vec{a})_2\right] \tag{4.18}$$

where the likelihood ratio is the ratio of the conditional probabilities of the the waveform $\mathbf{r}$ given signal present and noise alone conditions:

$$L(\mathbf{r}|\vec{a}) = \frac{p(\mathbf{r}|H_1, \vec{a})}{p(\mathbf{r}|H_0, \vec{a})} \tag{4.19}$$

Let $|\mathbf{S}_{pl}(\vec{\omega_i})|^2$ and $|\mathbf{S}_{pl}(\vec{\omega_i})|^2$ denote the plaid signal and bandpass signal normalized power spectra respectively, and let $\vec{S}(\vec{\omega_i}) = \left[|\mathbf{S}_{pl}(\vec{\omega_i})|^2, |\mathbf{S}_{pl}(\vec{\omega_i})|^2\right]$.

The likelihood ratio is given by:

$$\begin{aligned} \Lambda(\mathbf{R}|\vec{a}) &= \frac{\prod_{i=1}^{M} \frac{1}{[2\pi(\vec{a}^2 \cdot \vec{S}(\vec{\omega_i}) + \mathbf{N})]^{0.5}} \exp(-0.5 \sum_{i=1}^{M} \frac{\mathbf{R_i R_i^*}}{(\vec{a}^2 \cdot \vec{S}(\vec{\omega_i}) + \mathbf{N})})}{\prod_{i=1}^{M} \frac{1}{[2\pi \mathbf{N}]^{0.5}} \exp(-0.5 \sum_{i=1}^{M} \frac{\mathbf{R_i R_i^*}}{\mathbf{N}})} \\ \Lambda(\mathbf{R}|\vec{a}) &= \left(\prod_{i=1}^{M} \frac{\mathbf{N}^{0.5}}{(\vec{a}^2 \cdot \vec{S}(\vec{\omega_i}) + \mathbf{N})^{0.5}} \right) \exp\left(-\frac{0.5}{\mathbf{N}} \sum_{i=1}^{M} \frac{\left(\vec{a}^2 \cdot \vec{S}(\vec{\omega_i})\right) \mathbf{R_i R_i^*}}{(a^2 |\mathbf{S}_{pl}(\vec{\omega_i})|^2 + b^2 |\mathbf{S}_b(\vec{\omega_i})|^2 + \mathbf{N})} \right) \end{aligned} \tag{4.20}$$

When the background noise power is much greater than the total signal power the likelihood ratio reduces to:

$$\Lambda(\mathbf{R}|\vec{a}) = \exp\left(-\frac{0.5}{\mathbf{N}^2} \sum_{i=1}^{M} \left(\vec{a}^2 \cdot \vec{S}(\vec{\omega_i})\right) \mathbf{R_i R_i^*} \right) \tag{4.21}$$

The likelihood averaged across $\vec{a}^2$ is given by:

$$\mathrm{E}\left[L(\mathbf{r}|\vec{a})\right] = \sum_{\vec{a}} p(\vec{a}) \Lambda(\mathbf{R}|\vec{a}) \tag{4.22}$$

which follows because $p(\vec{a}) = p(\vec{a}^2)$, due to the fact that $\vec{a}$ is constrained to be positive. This expression does not simplify but can be simulated.

If the visual system computes the decision based on the sum of energies in the bands, but can vary the weighting within the bands, then we can compute the optimal weights based on the set of stimulus mixtures used. We can derive the optimal additive rule by using the log likelihood functions.

$$\log \Lambda(\mathbf{R}|\vec{a}) = -\frac{0.5}{\mathbf{N}^2} \sum_{i=1}^{M} \left(\vec{a}^2 \cdot \vec{S}(\vec{\omega_i})\right) \mathbf{R_i} \mathbf{R_i^*} \tag{4.23}$$

The expectation of the log likelihood function over $\vec{a}^2$ is given by:

$$\begin{aligned} \mathrm{E}\left[\log \Lambda(\mathbf{R}|\vec{a})\right] &= -\frac{0.5}{\mathbf{N}^2} \sum_{\vec{a}} p(\vec{a}) \sum_{i=1}^{M} \left(\vec{a}^2 \cdot \vec{S}(\vec{\omega_i})\right) \mathbf{R_i} \mathbf{R_i^*} d\vec{a} \\ &= -\frac{0.5}{\mathbf{N}^2} \sum_{i=1}^{M} \left(\vec{a}^{*\,2} \cdot \vec{S}(\vec{\omega_i})\right) \mathbf{R_i} \mathbf{R_i^*} \end{aligned} \tag{4.24}$$

where $\vec{a}^{*\,2}$ is the mean stimulus power vector. Thus in this case the decision is to compare the energies in the filter $\vec{a}^{*\,2} \cdot \vec{S}(\vec{\omega_i})$ on both intervals and choose the interval with the larger energy.

From the set of (E_{pl}, E_{bp}) energy vectors we used we can compute the expected energy and hence the expected weights on the energy bands. These in turn produce expected slopes which we can compare against the data. For subject PS the expected weights for the In Plane condition are (a, b, c) = (1,1.55,1), while for the Asymmetric condition is (a, b, c) = (1.63,1,1). For subject ML, the expected weights for the In Plane condition are (a, b, c) = (1,1.58,1), and for the Asymmetric condition are (a, b, c) = (1.65,1,1).

Chapter 5

Summary

We have provided substantial evidence for the existence of specialized mechanisms in the human visual system which detect local image translations using a strategy of pooling spectral power across planes in frequency space.

In the first chapter we motivated the problem of measuring local image displacements, and discussed how the planar power detector model arises naturally when the measurements are made by a system which is uncertain of the spatial profile of the moving signals.

In the second chapter, we designed a set of novel motion stimuli, and derived ideal observers to provide an absolute measure of performance. We found that observers efficiently detected stimuli matched to planar power detectors, relative to a set of control stimuli.

In the third chapter we reanalyzed the data from the second chapter using an ideal observer perturbation analysis to estimate subject's weighting function across spatio-temporal frequency. We found that subject's weighting functions supported the existence of planar power pooling in the visual system. However, the results also pointed to the existence of ubiquitous negative weights outside the signal bands, suggesting the planar power detector model must be modified to include some inhibitory interactions from frequency bands outside the plane.

The fourth chapter tested for additive pooling of power on the plane. The results showed that additivity held for bands intersecting a common plane, but performance was subadditive for bands not lying in a common plane. These results strongly argue for the existence of planar power detectors in the visual system. In addition we found evidence suggesting the visual system may be able to modify its spectral weighting function across the plane to better match the expected spectral properties of the signal. These results were discussed in terms of potential perceptual learning, possible relations between planar power detectors and neurons analogous to those in simian visual area MT, and the possibility the visual system uses a population of planar power detectors to perform image velocity estimation.

Bibliography

[1] E. H. Adelson and J. R. Bergen. Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A*, 2(2):284–299, 1985.

[2] E. H. Adelson and J. A. Movshon. Phenomenal coherence of moving visual patterns. *Nature*, 300(5892):523–525, 1982.

[3] M. Ahissar and S. Hochstein. Task difficulty and the specificity of perceptual learning. *Nature*, 387:401–406, 1997.

[4] A. Ahumada and J. Lovell. Stimulus features in signal detection. *J. Acoust. Soc. Am.*, 49:1751–1756, 1970.

[5] A. J. Ahumada and A. B. Watson. Equivalent-noise model for contrast detection and discrimination. *J. Opt. Soc. Am. A*, 2(7):1133–1139, 1985.

[6] D. Alais, P. Wenderoth, and D. Burke. The size and number of plaid blobs mediate the misperception of type-ii plaid direction. *Vis. Res.*, 37:143–150, 1997.

[7] D. G. Albrecht and W. S. Geisler. Motion selectivity and the contrast response function of simple cells in the visual cortex. *Visual Neuroscience*, 7:531–546, 1991.

[8] T. D. Albright. Direction and orientation selectivity of neurons in visual area MT of the macaque. *J. Neurophysiology*, 52:1106–1130, 1984.

[9] T. D. Albright. Form-cue invariant motion processing in primate visual cortex. *Science*, 255:1141–1143, 1992.

[10] J. Allman, F. Miezin, and E. McGuinness. Direction- and velocity-specific responses from beyond the classical receptive field in the middle temporal visual area. *Perception*, 14:105–126, 1985.

[11] S. J. Anderson and D. C. Burr. Spatial and temporal selectivity of the human motion detection system. *Vis. Res.*, 15:1147–1154, 1985.

[12] S. J. Anderson and D. C. Burr. Receptive field properties of human motion detector units inferred from spatial frequency masking. *Vis. Res.*, 29:1343–1358, 1987.

[13] S. J. Anderson and D. C. Burr. Receptive field size of human motion detecting units. *Vis. Res.*, 27:621–635, 1987.

[14] H. B. Barlow. A method of determining the overall quantum efficiency of visual discrimination. *J. Physiol., Lond.*, 160:155–168, 1962.

[15] H. B. Barlow. The efficiency of detecting changes of density in random dot patterns. *Vis. Res.*, 18:637–650, 1978.

[16] H. B. Barlow. The ferrier lecture: critical limiting factors in the design of the eye and visual cortex. *Proc. R. Soc. Lond. B*, 212:1–34, 1981.

[17] R. T. Born and R. B. Tootell. Segregation of global and local motion processing in middle temporal visual area. *Nature*, 357:497–499, 1992.

[18] L. Bowns. Evidence for a feature tracking explanation of why type ii plaids move in the vector sum direction at short durations. *Vis. Res.*, 36:3685–3694, 1996.

[19] M. J. Bravo and S. N. Watamaniuk. Evidence for two speed signals: a coarse local signal for segregation and a precise global signal for discrimination. *Vis. Res.*, 35:1691–1697, 1995.

[20] B. D. Bruyn and G. A. Orban. Human velocity and direction discrimination measured with random dot patterns. *Vis. Res.*, 28:1323–1335, 1988.

[21] A. E. Burgess and H. B. Barlow. The efficiency of numerosity discrimination in random dot images. *Vis. Res.*, 23:811–829, 1983.

[22] A. E. Burgess, X. Li, and C. K. Abbey. Visual signal detectability with two noise components: anomalous masking effects. *J. Opt. Soc. Am. A*, 14:2420–2442, 1997.

[23] A. E. Burgess, R. F. Wagner, R. J. Jennings, and H. B. Barlow. Efficiency of human visual signal discrimination. *Science*, 214:93–94, 1981.

[24] D. C. Burr, J. Ross, and M. C. Morrone. Seeing objects in motion. *Proc. R. Soc. Lond. B*, 227:249–265, 1986.

[25] T. Caelli and M. N. Oguztoreli. Some task and signal dependent rules for spatial vision. *Spat Vis*, 2:295–315, 1987.

[26] E. Castet. Apparent speed of type i symmetrical plaids. *Vis. Res.*, 36:223–232, 1996.

[27] P. Cavanagh. Attention-based motion perception. *Science*, 257:1563–1565, 1992.

[28] C. Chubb and G. Sperling. Drift-balanced radom stimuli: A general basis for studying non-fourier motion perception. *J. Opt. Soc. Am. A*, 5:1986–2007, 1988.

[29] C. Chubb and G. Sperling. Texture quilts: Basic tools for studying motion-from-texture. *Journal of Mathematical Psychology*, 35:411–442, 1991.

[30] R. E. Crist, M. K. Kapadia, G. Westheimer, and C. D. Gilbert. Perceptual learning of spatial localization: specificity for orientation, position, and context. *J Neurophysiol*, 78:2889–2894, 1997.

[31] J. G. Daugman. Two-dimensional spectral analysis of cortical receptive field profiles. *Vis. Res.*, 20:847–856, 1980.

[32] J. G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Opt. Soc. Am. A*, 2:1160–1169, 1985.

[33] J. G. Daugman and C. J. Downing. Demodulation, predictive coding, and spatial vision. *J. Opt. Soc. Am. A*, 12:641–660, 1995.

[34] D. M. Diamond and N. M. Weinberger. Classical conditioning rapidly induces specific changes in frequency receptive fields of single neurons in secondary and ventral ectosylvian auditory cortical fields. *Brain Res*, 372:357–360, 1986.

[35] T. Dijkstra. *Visual control of posture and visual perception of shape*. PhD thesis, Katholieke Universiteit Nijmegen, Nijmegen, Netherlands, 1994.

[36] A. Dorais and D. Sagi. Contrast masking effects change with practice. *Vis. Res.*, 37:1725–1733, 1997.

[37] M. P. Eckert and G. Buchsbaum. Efficient coding of natural time varying images in the early visual system. *Phil. Trans. R. Soc. Lond. B*, 339:385–395, 1993.

[38] M. P. Eckstein, A. J. Ahumada, and A. B. Watson. Visual signal detection in structured backgrounds. ii. effects of contrast gain control, background variations, and white noise. *J. Opt. Soc. Am. A*, 14:2406–2419, 1997.

[39] J. M. Edeline, P. Pham, and N. M. Weinberger. Rapid development of learning-induced receptive field plasticity in the auditory cortex. *Behav Neurosci*, 107:539–551, 1993.

[40] B. Efron and R. J. Tibshirani. *An introduction to the bootstrap*. Chapman & Hall, New York, NY, 1993.

[41] M. Fahle. Specificity of learning curvature, orientation, and vernier discriminations. *Vis. Res.*, 37:1885–1895, 1997.

[42] V. P. Ferrera and S. G. Lisberger. Neuronal responses in visual areas MT and MST during smooth pursuit target selection. *J Neurophysiol*, 78:1433–1446, 1997.

[43] V. P. Ferrera and H. R. Wilson. Direction specific masking and the analysis of motion in two dimensions. *Vis. Res.*, 27:1783–1796, 1987.

[44] V. P. Ferrera and H. R. Wilson. Perceived speed of moving two-dimensional patterns. *Vis. Res.*, 31:877–893, 1991.

[45] D. J. Field. Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A*, 4:2379–2394, 1987.

[46] D. J. Fleet and A. D. Jepson. Hierarchical construction of orientation and velocity selective filters. *IEEE Pat. Anal. Mach. Intell.*, 11:315–325, 1986.

[47] D. J. Fleet and K. Langley. Computational analysis of non-fourier motion. *Vis. Res.*, 34:3057–3079, 1994.

[48] R. H. Gilkey and D. Robinson. Models of auditory masking: A molecular psychophysical approach. *J. Acoust. Soc. Am.*, 79:1499–1510, 1986.

[49] H. Glunder. Correlative velocity estimation: visual motion analysis, independent of object form, in arrays of velocity tuned bilocal detectors. *J. Opt. Soc. Am. A*, 7:255–263, 1990.

[50] N. Graham. *Visual Pattern Analyzers*. Oxford University Press, New York, NY, 1989.

[51] N. Graham, J. G. Robson, and J. Nachmias. Grating summation in fovea and periphery. *Vis. Res.*, 18:815–825, 1978.

[52] D. M. Green and J. Swets. *Signal detection theory and psychophysics.* Wiley, New York, 1974.

[53] N. M. Grzywacz and A. L. Yuille. A model for the estimation of local image velocity by cells in the visual cortex. *Proc. R. Soc. Lond. A*, 239:129–161, 1990.

[54] D. B. Hamilton, D. G. Albrecht, and W. S. Geisler. Visual cortical receptive fields in monkey and cat: Spatial and temporal phase transfer function. *Vis. Res.*, 29:1285–1308, 1989.

[55] M. G. Harris. The perception of moving stimuli: a model of spatiotemporal coding in human vision. *Vis. Res.*, 26:1281–1287, 1986.

[56] D. J. Heeger. Model for the extraction of image flow. *J. Opt. Soc. Am. A*, 4(8):1455–1471, August 1987.

[57] D. J. Heeger. Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, 29:1285–1308, 1992.

[58] C. W. Helstrom. *Statistical theory of signal detection.* Oxford, New York, 1968.

[59] E. C. Hildreth, N. M. Grzywacz, E. H. Adelson, and V. K. Inada. The perceptual buildup of three-dimensional structure from motion. *Percept Psychophys*, 48:19–36, 1990.

[60] E. Hiris and R. Blake. Direction repulsion in motion transparency. *Vis. Neurosci.*, 13:187–197, 1996.

[61] B. K. P. Horn and B. G. Schunck. Determining optic flow. *Artif. Intell.*, 17:185–203, 1981.

[62] C. F. S. III, R. E. Kronauer, J. C. Madsen, and S. A. Klein. Opponent-movement mechanisms in human vision. *J. Opt. Soc. Am. A*, 1:876–884, 1984.

[63] J. H. T. Jamar and J. J. Koenderink. Contrast detection and detection of contrast modulation for noise gratings. *Vis. Res.*, 25:511–521, 1985.

[64] D. H. Kelly. Visual processing of moving stimuli. *J. Opt. Soc. Am. A*, 2:216–225, 1985.

[65] D. Kersten. Statistical efficiency for the detection of visual noise. *Vis. Res.*, 27:1029–1043, 1987.

[66] J. Kim and H. R. Wilson. Motion integration over space: interaction of the center and surround motion. *Vis. Res.*, 37:991–1005, 1997.

[67] D. C. Knill. Ideal observer perturbation analysis reveals human strategies for inferring surface orientation from texture. *Vis. Res.*, in press, 1997.

[68] J. J. Koenderink. Optic flow. *Vis. Res.*, 26:161–179, 1986.

[69] J. J. Koenderink and A. J. van Doorn. Spatial noise for visual research. *Vis. Res.*, 14:721–723, 1974.

[70] J. J. Koenderink and A. J. van Doorn. How an ambulant observer can construct a model of the enviornment from the geometrical structure of the visual inflow. In G. Hauske and E. Butenandt, editors, *Kybernetic.* Oldenburg, Muenchen, 1978.

[71] J. J. Koenderink and A. J. van Dorn. Invariant properties of the motion parallax field due to the movement of rigid bodies relative to an observer. *Optica Acta*, 22:773–791, 1975.

[72] J. Krauskopf and B. Farrell. Influence of color on the perception of coherent motion. *Nature*, 348:328–331, 1990.

[73] L. Lagae, S. Raiguel, and G. A. Orban. Speed and direction selectivity of macaque middle temporal neurons. *J Neurophys*, 69:19–39, 1993.

[74] M. S. Landy and J. R. Bergen. Texture segregation and orientation gradient. *Vis. Res.*, 31:679–691, 1991.

[75] J. S. Lappin, J. F. Norman, and L. Mowafy. The detectability of geometric structure in rapidly changing optical patterns. *Perception*, 20:513–528, 1991.

[76] G. E. Legge, D. Kersten, and A. E. Burgess. Contrast discrimination in noise. *J. Opt. Soc. Am. A*, 4:391–404, 1987.

[77] H. Levitt. Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, 49:1458–1471, 1971.

[78] J. X. Lubin. *Interactions among motion sensitive mechanisms in human vision.* PhD thesis, University ofPennsylvania, Department of Psychology, Philadelphia, PA, 1992.

[79] D. Marr and S. Ullman. Directional selectivity and its use in early visual processing. *Proc. R. Soc. Lond. B*, 211:151–180, 1981.

[80] J. H. R. Maunsell and D. C. V. Essen. Functional properties of neurons in middle temporal visual area of the macaque monkey I. Selectivity for stimulus direction, speed, and orientation. *Journal of Neurophysiology*, 49:1127–1147, 1983.

[81] J. E. W. Mayhew and J. P. Frisby. Suprathreshold contrast perception and complex random textures. *Vis. Res.*, 18:895–897, 1978.

[82] S. P. McKee, G. H. Silverman, and K. Nakayama. Precise velocity discrimination despite random variations in temporal frequency and contrast. *Vis. Res.*, 26:609–619, 1986.

[83] J. McLean and L. A. Palmer. Organization of simple cell responses in the three dimensional (3 d) frequency domain. *Visual Neuroscience*, 11:295–306, 1994.

[84] J. McLean and L. A. Palmer. Plasticity of neuronal response properties in adult cat striate cortex. *Vis. Neurosci.*, 15:177–196, 1998.

[85] J. McLean, S. Raab, and L. A. Palmer. Contribution of linear mechanisms to the specification of local motion by simple cells in areas 17 and 18 of the cat. *Visual Neuroscience*, 11:271–294, 1994.

[86] A. Mikami, W. T. Newsome, and R. H. Wurtz. Motion selectivity in macaque visual cortex. I. mechanisms of direction and speed selectivity in extrastriate area MT. *J Neurophysiol*, 55:1308–1327, 1986.

[87] O. R. Mitchell. Effect of spatial frequency on the visibility of unstructured patterns. *J. Opt. Soc. Am.*, 66:327–332, 1976.

[88] M. J. Morgan. Spatial filtering precedes motion detection. *Nature*, 355:344–346, 1992.

[89] M. J. Morgan and G. Mather. Motion discrimination in two-frame sequences with differing spatial frequency content. *Vis. Res.*, 34:197–208, 1994.

[90] J. A. Movshon, E. H. Adelson, M. S. Gizzi, and W. T. Newsome. The analysis of moving visual patterns. In C. Chagas, R. Gattass, and C. Gross, editors, *Experimental Brain Research Supplementum II: Pattern Recognition Mechanisms*, pages 117–151. Springer-Verlag, New York, 1986.

[91] J. A. Movshon, W. T. Newsome, M. S. Gizzi, and J. B. Levitt. Spatio-temporal tuning and speed sensitivity in macaque visual cortical neurons. *Investigative Ophthalmology and Visual Science (Suppl.)*, 29:327, 1988.

[92] K. Nakayama. Biological image motion processing: a review. *Vis. Res.*, 25:625–660, 1985.

[93] W. T. Newsome and E. B. Pare. A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *Journal of Neuroscience*, 8:2201–2211, 1988.

[94] H. P. Norman, J. F. Norman, J. T. Todd, and D. T. Lindsey. Spatial interactions in perceived speed. *Perception*, 25:815–830, 1996.

[95] T. Pasternak and W. H. Merigan. Motion perception following lesions of the superior temporal sulcus in the monkey. *Cerebral Cortex*, 4:247–259, 1994.

[96] D. G. Pelli. Uncertainty explains many aspects of visual contrast detection and dis crimination. *J. Opt. Soc. Am. A*, 2:1508–1532, 1985.

[97] D. G. Pelli. Noise in the visual system may be early. In M. S. Landy and J. A. Movshon, editors, *Computational Models of Visual Processing.*, pages 147–151. MIT press, Cambridge, MA, 1991.

[98] D. G. Pelli and L. Zhang. Accurate control of contrast on microcomputer displays. *Vis. Res.*, 31:1337–1350, 1991.

[99] A. Pentland. Photometric motion. *IEEE Pat. Anal. Mach. Intell.*, 13:879–890, 1991.

[100] T. Poggio and W. Reichardt. Considerations on models of movement detection. *Kybernetic*, 13:223–227, 1973.

[101] K. Prazdny. On the information in optical flows. *Comp. Vis. Graphics Image Proc.*, 22:239–259, 1982.

[102] W. H. Press, S. A. Teukolsky, W. T. Vettering, and B. P. Flannery. *Numerical Recipes in C, 2nd ed.* Cambridge, New York, NY, 1988.

[103] N. Qian, R. A. Andersen, and E. H. Adelson. Transparent motion perception as detection of unbalanced motion signals. iii. modeling. *J. Neurosci.*, 14:7381–7392, 1994.

[104] W. Reichardt. Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. In W. A. Rosenblith, editor, *Sensory Communication*, pages 303–317. John Wiley and Sons, New York, 1961.

[105] V. M. Richards and S. Zhu. Relative estimates of combination weights, decision criteria, and internal noise based on correlation coefficients. *J. Acoust. Soc. Am.*, 95:423–434, 1994.

[106] H. R. Rodman and T. D. Albright. Coding of stimulus velocity in area MT of the macaque. *Vis. Res.*, 27:2035–2048, 1987.

[107] H. R. Rodman and T. D. Albright. Single-unit analysis of pattern-motion selective properties in the middle temporal visual area (MT). *Experimental Brain Research*, 75:53–64, 1989.

[108] B. Rogers and M. Graham. Motion parallax as an independent cue for depth perception. *Perception*, 8:125–134, 1979.

[109] C. D. Salzman, K. H. Britten, and W. T. Newsome. Cortical stimulation influences perceptual judgements of motion direction. *Nature*, 346:174–177, 1990.

[110] T. D. Sanger. Stereo disparity computation using gabor filters. *Biol. Cybern.*, 59:405–418, 1988.

[111] P. R. Schrater and E. P. Simoncelli. Velocity representation: Evidence from motion adaptation. *Vis. Res.*, in press, 1997.

[112] E. P. Simoncelli. *Distributed Analysis and Representation of Visual Motion.* PhD thesis, Massachusetts Institute of Technology, Dept ofElectrical Engineering and Computer Science, Cambridge, MA, January 1993.

[113] E. P. Simoncelli and D. J. Heeger. A model of neural responses in visual area MT. *Vis. Res.*, in press, 1998.

[114] R. J. Snowden and O. J. Braddick. The temporal integration and resolution of velocity signals. *Vis. Res.*, 31:907–914, 1991.

[115] L. S. Stone, A. B. Watson, and J. B. Mulligan. Effect of contrast on the perceived direction of a moving plaid. *Vis. Res.*, 30:1049–1067, 1990.

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

[117] J. E. Thornton and E. N. Pugh. Red/green color opponency at detection threshold. *Science*, 219:191–193, 1983.

[118] S. Treue and J. H. Maunsell. Attentional modulation of visual motion processing in cortical areas MT and MST. *Nature*, 382:539–541, 1996.

[119] W. A. van de Grind, J. J. Koenderink, A. J. van Doorn, M. V. Milders, and H. Voerman. Inhomogeneity and anisotropies for motion detection in the monocular visual field of human observers. *Vis. Res.*, 33:1089–1107, 1993.

[120] W. A. van de Grind, A. J. van Doorn, and J. J. Koenderink. Detection of coherent movement in peripherally viewed random-dot patterns. *J. Opt. Soc. Am. A*, 73:1674–1683, 1983.

[121] A. J. van Doorn and J. J. Koenderink. Spatial properties of the visual detectability of moving spatial white noise. *Exp. Brain Research*, 45:189–195, 1982.

[122] A. J. van Doorn and J. J. Koenderink. Temporal properties of the visual detectability of moving spatial white noise. *Exp. Brain Research*, 45:179–188, 1982.

[123] A. J. van Doorn and J. J. Koenderink. Visibility of movement gradients. *Biol. Cybernet.*, 44:167–175, 1982.

[124] A. J. van Doorn and J. J. Koenderink. Detectability of velocity gradients in moving random-dot patterns. *Vis. Res.*, 23:799–804, 1983.

[125] A. J. van Doorn and J. J. Koenderink. The structure of the human motion detection system. *IEEE Trans. Systems Man and Cybernetics*, 13:916–922, 1983.

[126] A. van Meeteren and H. B. Barlow. The statistical efficiency for detecting sinusoidal modulation of average dot density in random figures. *Vis. Res.*, 21:765–789, 1981.

[127] J. P. H. van Santen and G. Sperling. Temporal covariance models of human motion perception. *J. Opt. Soc. Am. A*, 1:451–473, 1984.

[128] H. L. van Trees. *Detection, Estimation, and Modulation Theory. Part I.* Wiley, New York, 1968.

[129] H. L. van Trees. *Detection, Estimation, and Modulation Theory. Part III.* Wiley, New York, 1971.

[130] A. Verri and T. Poggio. Motion field and optical flow: Qualitative properties. *IEEE Pat. Anal. Mach. Intell.*, 11:490–498, 1989.

[131] N. J. Wade. A selective history of the study of visual motion aftereffects. *Perception*, 23:1111–1134, 1994.

[132] B. A. Wandell and E. N. Pugh. A field-additive pathway detects brief-duration, long-wavelength incremental flashes. *Vis. Res.*, 20:613–624, 1980.

[133] W. H. Warren and D. J. Hannon. Direction of self-motion is perceived from optical flow. *Nature*, 336:162–163, 1988.

[134] S. N. Watamaniuk, S. P. McKee, and N. M. Grzywacz. Detecting a trajectory embedded in random-direction motion noise. *Vis. Res.*, 35:65–77, 1995.

[135] A. B. Watson. Probability summation over time. *Vis. Res.*, 19:515–522, 1979.

[136] A. B. Watson. Optimal displacement in apparent motion and quadrature models of motion sensing. *Vis. Res.*, 30:1389–1393, 1990.

[137] A. B. Watson. Perceptual-components architecture for digital video. *J. Opt. Soc. Am. A*, 2:322–341, 1990.

[138] A. B. Watson and A. J. Ahumada. A look at motion in the frequency domain. In J. K. Tsotsos, editor, *Motion: Perception and representation*, pages 1–10. Association for Computing Machinery, New York, 1983.

[139] A. B. Watson and A. J. Ahumada. Model of human visual-motion sensing. *J. Opt. Soc. Am. A*, 2:322–342, 1985.

[140] A. B. Watson, A. J. Ahumada, and J. E. Farrell. The window of visibility: A psychophysical theory of fidelity in time-sampled motion displays. *J. Opt. Soc. Am. A*, 3:300–307, 1986.

[141] A. B. Watson and M. P. Eckert. Motion contrast sensitivity: Visibility of motion gradients of various spatial frequencies. *J. Opt. Soc. Am. A*, 11:496–505, 1994.

[142] A. B. Watson and D. G. Pelli. QUEST: A Bayesian adaptive psychophysical method. *Perception & Psychophysics*, 33:113–120, 1983.

[143] A. B. Watson and K. Turano. The optimal motion stimulus. *Vis. Res.*, 35:325–336, 1994.

[144] N. M. Weinberger. Learning-induced changes of auditory receptive fields. *Curr Opin Neurobiol*, 3:570–577, 1993.

[145] P. Wenderoth, D. Alais, D. Burke, and R. van der Zwan. The role of the blobs in determining the perception of drifting plaids and their motion aftereffects. *Perception*, 23:1163–1169, 1994.

[146] P. Werkhoven and J. J. Koenderink. Extraction of motion parallax structure in the visual system. *Biol. Cybern.*, 63:185–191, 1990.

[147] H. R. Wilson and J. Kim. Perceived motion in the vector sum direction. *Vis. Res.*, 34:1835–1842, 1994.

[148] D. K. Xiao, V. L. Marcar, S. E. Raiguel, and G. A. Orban. Selectivity of macaque MT/V5 neurons for surface orientation in depth specified by motion. *Eur J Neurosci*, 9:956–964, 1997.

[149] D. K. Xiao, S. Raiguel, V. Marcar, J. J. Koenderink, and G. A. Orban. Spatial heterogeneity of inhibitory surrounds in the middle temporal visual area. *Proc Natl Acad Sci*, 92:11303–11306, 1995.

[150] D. K. Xiao, S. Raiguel, V. Marcar, and G. A. Orban. The spatial distribution of the antagonistic surround of MT/V5 neurons. *Cereb Cortex*, 7:662–677, 1997.

[151] M. Young, M. Landy, and L. Maloney. A perturbation analysis of depth perception from combinations of texture and motion cues. *Vis. Res.*, 33:2685–2696, 1993.