②

REPORT DOCUMENTATION PAGE

1a REPORT SECURITY CLASSIFICATION	1b RESTRICTIVE MARKINGS
UNCLASSIFIED	
2a SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION/AVAILABILITY OF REPORT
	Approved for public release; distribution unlimited.

AD-A212 670

2b ...	5 MONITORING ORGANIZATION REPORT NUMBER(S)
	AFOSR-TR- 89-1131

6a. NAME OF PERFORMING ORGANIZATION	6b OFFICE SYMBOL (If applicable)	7a NAME OF MONITORING ORGANIZATION
New York University		Air Force Office of Scientific Research/NL

6c. ADDRESS (City, State, and ZIP Code)	7b ADDRESS (City, State, and ZIP Code)
Psychology Department New York, NY 10003	Building 410 Bolling AFB, DC 20332-6448

8a. NAME OF FUNDING/SPONSORING ORGANIZATION	8b OFFICE SYMBOL (If applicable)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
AFOSR	NL	AFOSR-88-0140

8c. ADDRESS (City, State, and ZIP Code)	10 SOURCE OF FUNDING NUMBERS			
	PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO
Building 410 Bolling AFB, DC 20332-6448	61102F	2313	A5	

11 TITLE (Include Security Classification)

Three Stages and Two Systems of Visual Processi-ng

12 PERSONAL AUTHOR(S)
George Sperling

13a. TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Reprint	FROM ____ TO ____		32

16 SUPPLEMENTARY NOTATION

17	COSATI CODES		18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

DTIC
ELECTE
AUG 18 1989
S
B
D

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21 ABSTRACT SECURITY CLASSIFICATION
☒ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS	UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL	22b TELEPHONE (Include Area Code)	22c OFFICE SYMBOL
Dr. John F. Tangney	(202) 767-5021	NL

89 8 1 074

Three Stages and Two Systems of Visual Processing

George Sperling

Human Information Processing Laboratory
Psychology Department, New York University
6 Washington Place
New York, NY   10003

Address correspondence to:

Prof. George Sperling
Psychology Dept.   NYU
6 Washington Place
New York, NY   10003

Three Stages and Two Systems of Visual Processing

George Sperling

Human Information Processing Laboratory, Department of Psychology,

New York University, New York, NY   10003

ABSTRACT

Three stages of visual processing determine how internal noise appears to an external observer: light adaptation, contrast gain control, and a postsensory/decision stage. Dark noise occurs prior to adaptation, determines dark-adapted absolute thresholds, and mimics stationary external noise. Sensory noise occurs after dark adaptation, determines contrast thresholds for sine gratings and similar stimuli, and mimics external noise that increases with mean luminance. Postsensory noise incorporates perceptual, decision, and mnemonic processes. It occurs after contrast-gain control and mimics external noise that increases with stimulus contrast (i.e., *multiplicative* noise). Dark noise and sensory noise are frequency specific and primarily affect weak signals. Only postsensory noise significantly affects to strong signals, and it has constant power over a wide spatial frequency range in which sensory noise varies enormously.

Two parallel perceptual regimes jointly serve human object recognition and motion perception: a first-order linear (Fourier) regime that computes relations directly from luminances, and a second-order nonlinear (nonFourier) rectifying regime that uses the absolute value of detector outputs. When objects or movements are defined by high spatial frequencies (i.e., texture *carrier* frequencies whose wavelengths are small compared to the object size), the responses of high-frequency receptors are *demodulated* by *rectification* to facilitate discrimination at the next hierarchical processing levels. Rectification sacrifices the statistical efficiency (noise resistance) of the first-order regime for efficiency of connectivity and computation.

Three Stages and Two Systems of Visual Processing

George Sperling

New York University

Address correspondence to:

George Sperling
Psychology Department, NYU
6 Washington Pl., Rm. 980
New York, NY 10003

Telephone: (212) 998-7868

Running head: Stages and Systems of Visual Processing

Draft, Revision 2

Three Stages and Two Systems of Visual Processing

George Sperling

Human Information Processing Laboratory, Department of Psychology,

New York University, New York, NY 10003

Bandpass filtering refers to the processing of images or sounds so that they contain only a narrow range--typically one or two octaves--of component frequencies. In audition, bandpass filtering is used to create stimuli that stimulate only a small portion of the basilar membrane. By studying psychophysical responses to stimuli filtered in different bands, information processing mediated by each portion of the basilar membrane can be studied.

In vision, the aim of bandpass filtering is to create stimuli that stimulate only one or a small number of the visual channels that operate in parallel to process visual stimuli. Ideally, stimuli filtered in high frequency bands would stimulate only receptors (channels) with small receptive fields. Stimuli filtered in low frequency bands would stimulate only channels that have large receptive fields. (The term channel is used here to designate an information processing system characterized by receptors of a particular size.) As in audition, there is substantial interest not only in how stimuli that are confined to a single band are processed, but also in how information from stimuli in different bands is perceptually combined.

With the advent of affordable graphics processors, bandpass filtering has become an increasingly widespread stimulus manipulation in vision. Working with bandpasssed stimuli raises to the fore some important issues that are the subject of this article. With hindsight, we see, as usual, that some of these issues have have been confronted before, but advent of bandpassed stimuli offers important new insights. In other cases, new stimuli and procedures raise new questions and offer new opportunities. This paper coordinates data that have emerged from paradigms that utilize bandpass filtered stimuli together with a variety of other data in order to arrive at some general principles of sensory information processing.

1. Visual Noise at Three Stages of Processing

Consider first a study that was originally designed to determine whether image spatial frequencies or object spatial frequencies were critical for object discrimination. Parish and Sperling (1987a, 1987b)

filtered individual capital letters in five different spatial frequency bands (Fig. 1). They studied the role of three factors in the ability of subjects to identify these letters when they were embedded in noise: (1) the signal-to-noise ratio, (2) the object-relative spatial frequency band in which the letters-plus-noise were filtered, and (3) the viewing distance (which determined retinal spatial frequency). They found that identification accuracy was independent of viewing distance over a range of more than 30:1. In this wide range, retinal spatial frequency did not matter in determining recognition accuracy; only object spatial frequency mattered. On the other hand, visual sensitivity to sine gratings at threshold varies enormously within the same range of retinal frequencies. In this section, we examine sine-grating detection and letter discrimination in order to define the various sources of noise that limit visual performance.

Insert 1 here.

APOSR-TR. 89-1131

Additive and Multiplicative Noise

We consider two kinds of noise: additive noise and multiplicative noise. The term additive noise is used here to denote a stationary noise source that is independent of the signal and is added to the signal. Additive noise can be overcome by increasing signal strength until the effective signal-to-noise ratio is sufficient support the desired level of performance.

Multiplicative noise is proportional to the signal, that is, it multiplies the signal. For example, in a binary (dark-grey/light-grey) image, reversing the contrast of (multiplying by -1) a randomly chosen 10 percent of the pixels would be a form of multiplicative noise. Increasing the intensity or the contrast of the image would not alter its signal-to-noise ratio. Multiplicative noise is equivalent to adding a noise whose expected power is proportional to signal power. Several authors have noted the distinction between additive and multiplicative noise (e.g., Carlson & Klopfenstien, 1985; Legge and Foley, 1980; Legge, Kersten, & Burgess, 1987; Pavel, Sperling, Riedl, & Vanderbeek, 1987). Loss of information that results from too-sparse sampling of the stimulus also can be regarded as a multiplicative noise because in cannot be overcome by increasing signal strength (e.g., Legge, et al, 1987).

Multiplicative noise cannot be responsible for detection or discrimination thresholds that are reached by reducing the strength of a signal. Because multiplicative noise declines proportionately with diminishing signal strength, weak signals are not worse off than strong signals. Because sufficiently low-luminance or sufficiently low-contrast visual signals are not visible, we infer that the internal noise that limits vision at low contrasts is better represented as additive rather than as multiplicative noise.

There are many visual discrimination tasks in which increasing stimulus luminance or contrast does not improve performance. Consider five examples. In attempting to detect a spatial sine wave grating embedded in noise, the contrast of the display has no effect on performance once a critical contrast is reached (Pelli, 1981). In detecting spatial amplitude modulation of a one-dimensional spatial noise, once about eight times the contrast threshold for the noise is reached, further increases in overall contrast do not make the modulation more detectable (Jamar, Campagne, & Koenderink, 1982; Jamar & Koenderink, 1985). In Parish & Sperling's (1987b) letter-in-noise discrimination task, only the signal-to-noise ratio matters (Note 1). In discriminating direction of motion, once a contrast of about 0.05 is reached, further increases in contrast do not improve performance (Nakayama & Silverman, 1985). In audition, similar kinds of results in which only the signal-to-noise ratio (and not absolute loudness matters) are the norm. For example, when a noisy radio broadcast is loud enough to be distinctly heard, making it louder does not make it more intelligible. The visual analog, the independence of the discriminability of noisy, dynamic visual signals upon the stimulus contrast at which they were viewed was verified over an 4:1 range of contrasts by Pavel, Sperling, Riedl, and Vanderbeek (1987). In such cases, human performance appears to be characterized by multiplicative noise.

From a theoretical point of view, it is important to note that systems, which appear upon external examination to have identical multiplicative noise, may have vastly different internal mechanisms for generating their behavior. Viewed externally, the internal operation of *multiplying* the noise by a factor k before adding it to the signal is equivalent to the internal operation of *dividing* the signal by k before adding it to the noise. Both result in the same internal signal-to-noise ratio s/n. The equivalence of dividing signals by k and multiplying noise by k suggests gain control as a physiologically plausible internal mechanism to mimic multiplicative noise: The gain-control multiplies input signals by $1/k$ before a

constant-power internal noise is added.

Three Sources of Visual Noise.

To understand how internal noise sources appear when viewed from the outside, it is useful to consider three stages of visual processing: light adaptation, contrast gain control, and decision. Figure 2 illustrates a flow chart for the computations carried out by these early stages. The particular mechanisms indicated in Fig. 2 for light adaptation and for contrast-gain control are based on physiologically plausible principles. They are vastly oversimplified and serve to illustrate the functional principles of the processes of light adaptation and gain control rather than the precise details (cf. Shapley & Enroth Cugell, 1986). For example, the flow charts omits the division of signals into two distinct pathways that carry only positive and only negative signals (the on-center and the off-center neurons), parallel spatial frequency channels are not explicitly treated, there is no gain control for w_2, and so on.

Insert 2 here.

Three stationary noise sources are illustrated in Fig. 2; each has constant expected power and an unchanging frequency spectrum. The three stages at which noise is added are (1) directly at the input, (2) after light adaptation, and (3) after contrast-gain control.

Dark Noise.

In absolute darkness, the spontaneous activity of the visual receptors, rods and cones, is represented as dark noise (Barlow, 1956, 1957). Dark noise is prior to any processes responsible for light adaptation. To be reliably detected against a totally dark background, a signal must exceed not only the level of dark noise but also the combined level of all noise in the visual pathways. However, it would be expected that, through evolution, absolute threshold would be determined primarily by dark noise. That is, for receptors to serve most efficiently, their amplification gain would have increased (through evolution) up to the point where the receptor noise itself was the limiting factor.

Sensory noise.

Sensory noise is the limiting noise in the detection of weak signals against uniform backgrounds. For example, by definition, a spatial sine wave grating with a contrast of 0.0001 has an absolute modulation that is proportional to its mean luminance. The brighter the illumination, the greater its absolute modulation. If there were no sensory noise, than increasing the absolute modulation of a spatial sinewave grating by increasing its mean luminance at constant contrast ultimately would increase its absolute modulation to the point of visibility (even with quantal noise in the stimulus). However, at high luminances, grating stimuli are visible very nearly in proportion to their contrast, not to their absolute modulation. The essential fact of sensory noise is that, when viewed from outside the system at moderate to high intensities, its apparent power increases with absolute modulation rather than remaining constant. To model noise that apparently increases with the mean luminance (background luminance), the sensory noise source is placed after (central to) the gain control that modulates visual responsiveness as a function of intensity. Constant sensory noise, placed after the gain-control mechanism, mimics an external additive noise that increases as a function of background intensity.

Sensory noise, Weber's law, quantal fluctuations. Weber's law asserts that the minimum detectable increment in intensity ΔS increases in direct proportion to background intensity S on which it is superimposed; at threshold: $\Delta S / S = k$, a constant. Assume that, at threshold, a constant signal-to-noise ratio is required at the detector itself: $s/n =$ signal amplitude / root-mean-square (RMS) noise amplitude. Indeed, the effective signal-to-noise ratio of the stimulus at the detector is equivalent to the d' statistic of signal detection theory (Green & Swets, 1966). Internal noise after adaptation to the background is equivalent to external noise whose RMS power increased in proportion to the background intensity: either results in Weber's law behavior because, to maintain a constant signal-to-noise ratio, the threshold increment would have to increase in direct proportion to the mean background. Thus sensory noise is hypothesized to be the source of Weber's law.

Most visual stimuli are produced by sources that can, for practical purposes, be approximated as quantal emitters. This means that, even with a nominally constant stimulus, the number of quanta collected by the retina in any given area varies from occasion to occasion and is characterized by a Poisson distribution. The *variance* of the Poisson distribution is equal to its mean, therefore, the RMS power of

quantum noise increases in direct proportion to the *square root* of the luminance of visual stimuli. Because quantal noise increases with the stimulus amplitude, it usually is considered in conjunction with sensory noise.

The full analysis of quantal noise in the stimulus itself together with such factors as the blur of the visual optics and the spacing of retinal cone receptors is quite complex. For example, Banks, Geisler & Bennett (1988) and Geisler (1989) applied such an analysis to contrast detection thresholds for sine wave stimuli of spatial frequencies from 5 to 40 cycles per degree, at mean luminances from 3.4 to 340 cd/m^2 (10.7 to 1068 trolands). Stimuli at each frequency consisted of seven sine cycles; i.e., the stimuli of different spatial frequencies were scaled replicas of each other. Once all the all the factors preneural cited above had been taken into account, Banks et al found that, at the observed thresholds, the stimulus s/n at the detector was constant. The most parsimonious interpretation is that sensory noise is negligible compared to quantal stimulus noise for their stimuli. For sine wave gratings at lower spatial frequencies than 5 cpd and for more intense stimuli at all spatial frequencies, sensory noise becomes quite significant relative to quantal noise. At very low levels of background luminance, dark noise becomes important (Geisler, 1989). Indeed, a model such as that of Fig. 2, together with threshold data obtained at different adaptation levels, offers a clear distinction between and independent estimates of residual sensory noise and dark noise.

Postsensory Noise.

In the superthreshold experiments with added external noise discussed above, detection depended only on the signal-to-noise ratio s/n and not on on the contrast at which these signals-plus-noise were viewed. In terms of a model, the dependence of objective performance measures (such as direction-of-motion judgments, intelligibility scores, letter discriminations) on s/n and their independence of stimulus contrast is represented by a contrast-gain control that equates all signals that exceed a minimal contrast level. For example, the input/output function illustrated in the contrast-gain control box of Fig. 2 is shaped like a logistic function with an asymptotic output of -1 for large negative contrasts, and an asymptotic output of +1 for large positive contrasts. A constant noise source that was located centrally to (added after) such a gain control would appear to an external observer to be equivalent to an external noise source that was directly proportional to contrast in those ranges of input where the gain-control was near its

asymptotes.

From a functional point of view, all noise sources that are added after contrast-gain control will appear externally to be multiplicative noises, proportional to stimulus contrast. There are many such sources. Consider a two-alternative forced-choice intensity discrimination task. In successive intervals, an observer is presented with, for example, sounds of intensity 40 and 41 db and required to say which interval contained the louder sound. Generally, observers do better in a pure block of trials (only two sounds 40 and 41 db occur) than in a mixed block (e.g., trials with 40 and 41 db mixed in with trials containing 60 and 61 db, a "roving" discrimination--Berliner & Durlach, 1973). In the pure block, the inability of the human observer to equal the performance of the ideal observer is attributed to a combination of (human) sensory and decision noise. In the mixed block, there is additional "context" noise due to an attentional/mnemonic component. In an identification task, where observers must name each stimulus (e.g., 40, 41, 60, 61 db) their performance can be characterized as being further degraded by mnemonic noise.

The relative levels of performance in any two complex detection, discrimination, or identification tasks will be determined by a combination of shared noise sources and task-specific noise sources (e.g., MacMillan, 1987). All these postsensory noise sources are grouped together under the heading of postsensory noise, representing perceptual, contextual, decision, attentional, mnemonic, and response processes that, according to the task, add noise after contrast-gain control.

To recapitulate: In vision, at threshold, sensitivity is governed by the intrinsic additive noise of the visual system (Pelli, 1981). Above threshold, matters apparently are quite different. "The notion of the observer's equivalent noise, which has been so useful in understanding detection, is found not to be relevant at suprathreshold contrasts." (Pelli, 1981, p. 121). However, to formulate coherent theories of performance, we need merely to enlarge the concept of equivalent noise to include noise sources that, to an external observer, appear to vary with adaptation (because they are located after adaptation gain control) and noise sources that appear to vary with stimulus contrast (because they are located after contrast-gain control).

The Efficiency of Detection.

The efficiency eff of detection or of discrimination is the ratio of s_i/n_i required by an ideal observer to the ratio s_h/n_h required by a human observer at the same criterion level c of performance:

$$eff = \left[\frac{s_i}{n_i}\right]_c^2 / \left[\frac{s_h}{n_h}\right]_c^2$$. For example, in a visual display of n independent, equivalently informative

pixels, eff is the fraction of the n pixels that the ideal observer needs to observe in order to match human performance.

Experimental determinations of efficiency establish an upper bound on the power of the human internal noise sources. Parish & Sperling (1987a) determined the efficiency of human discrimination in identifying visual letters masked by noise. When both the letters and noise were passed through a filter centered at 1.05 cycles per letter height, efficiency exceeded 0.40. Furthermore this high efficiency was observed over a 30:1 range of viewing distances. At the different viewing distances, these stimuli are transduced by visual channels characterized by vastly different retinal spatial frequencies. The constant high efficiency suggests that information loss in the visual pathway before the point of postsensory noise was negligible. In terms of noise sources, this means that dark noise and sensory noise were negligible compared to stimulus noise, and the postsensory noise was of same order of power as the real stimulus noise. Over the enormous range of spatial frequencies subserved by these channels, efficiency was determined primarily by postsensory noise.

2. Letter Discrimination, Noise, and the Spatial Modulation Transfer Function (MTF)

The MTF, also called the *contrast transfer function*, is the function that gives the contrast modulation of a sinewave grating at its threshold of detection as a function of its spatial frequency (Fig. 3). For the sine waves that fall in the range of retinal spatial frequencies investigated in Parish & Sperling's (1987a) letter detection experiments, contrast threshold ranges from a minimum of 0.002 at 5 - 8 cpd to a maximum of 0.7 at 37 cpd (e. g., van Nes and Bouman, 1967, cf Fig. 3). The frequency of 37 cpd is the mean retinal frequency of Parish & Sperling's highest frequency band b_5 at their longest viewing distance. The most detectable retinal frequencies (5 - 8 cpd) are produced by frequency bands b_3, b_4, and b_5 (Fig. 1) at closer viewing distances. In all these letter-discrimination conditions, observed discrimination efficiency was independent of the mean retinal frequency whereas threshold sensitivity for sinusoidal gratings varies from 0.7 to 0.002, a factor of 35, within this frequency range (Fig. 3). Indeed, the combination with of filter

frequency with viewing distance produces retinal frequencies that vary over a range of more than 200:1. Figure 3 illustrates the division of spatial frequencies into three regions:

(1) The top region which represents invisible sinusoidal gratings--their contrast is below detection threshold.

(2) A middle region, indicated in grey, in which detection is governed by quantal and sensory noise. In this region, increasing stimulus contrast improves performance.

(3) The lower region in which postsensory noise predominates. Here, noise is proportional to contrast so performance is independent of contrast. The numbers indicate the center frequency (projection on the x-axis) of various bandpass stimulus conditions of the Parish-Sperling study, and the approximate contrast level (0.1, projection on the y-axis) at which performance becomes independent of stimulus contrast.

Previously, Jamer & Koenderink (1985) had noted an apparent independence of spatial-frequency in the detection of amplitude modulated noise gratings. They investigated a relatively small range of frequencies and did not determine the efficiency of detection. In letter detection, the enormous range of frequency invariance, and the extremely low level of decision noise (as demonstrated by comparison with ideal detectors) is truly astounding.

Insert 3 here.

Detection thresholds for sine gratings vary enormously with retinal spatial frequency in precisely the same range of frequencies where discrimination threshold for letters-in-noise is constant. The difference between the two experiments is readily interpreted in terms of the levels-of-noise model. The grating detection expe⋯ is limited by quantum noise in the stimulus and by sensory noise; the letter-in-noise discrimination experiment is limited by postsensory noise. Whereas letter-in-noise discrimination is unaffected by stimulus contrast over a wide range, stimulus contrast is the dependent variable in the grating detection experiment. Indeed, the grating detection experiment can be viewed as indicating the effective power of quantal plus sensory noise as a function of spatial frequency. We say "effective power" because

there is no provision in the simple stage model for input amplification that may vary as a function of spatial frequency; input gain is incorporated into sensory noise.

Conclusions. Representing grating detection and letter-in-noise discrimination as noise limited processes, yields the following conclusions: (1) Sine grating detection at low stimulus contrasts is limited by quantal noise (Banks, et al, 1988) and by sensory noise (Pelli, 1981) each of which varies little with stimulus contrast but varies greatly with retinal spatial frequency and with mean luminance. (2) Letter detection at stimulus contrasts greater than about 0.10 is limited by apparently multiplicative noise that is proportional to stimulus contrast but is independent of spatial frequency (for a 100-fold range of retinal spatial frequencies). (3) When letters are discriminated in external noise which deliberately is not negligible, the effective internal noise apparently varies multiplicatively with stimulus contrast. These empirical relationships follow from the stage model of Fig. 2; and they are illustrated in Fig. 3. (4) Sensory and postsensory noise are independent and vary differently with spatial frequency. For example, the channel that transduces 5 - 8 cpd has the lowest sensory noise, but it has the same decision noise as the channel that processes 37 cpd, which has 35 times more sensory noise.

Analogous phenomena in psychoacoustics. A similar pattern of strong frequency dependence of threshold detection and frequency independence of high-intensity discrimination occurs in psychoacoustics. For example, absolute intensity-detection thresholds $\Delta I(f)$ for sinusoidal pressure waveforms vary enormously as a function frequency f. At high signal levels, detection thresholds for sinusoidal increments $\Delta I(f)/I$ hardly vary with frequency (Robinson & Dadson, 1956; Reisz, 1928; Jesteadt, Wier, & Green, 1977; see Scharf & Buus, 1986, for a review). Detection limits at low input levels are quite different from discrimination limits at high input levels. The nature of these differences is dictated by requirements of having maximally sensitive receptors and of operating over an enormous dynamic range. Since these problems are shared by many modalities, we should not be surprised at functionally similar solutions.

Advantage of above-threshold gain that is independent of frequency. A visual object is characterized by relations between its component spatial frequencies. When the object is viewed from nearer or further, these relations do not change, they are merely transposed up or down the frequency axis. If the visual system had important gain differences between different spatial frequencies, then these differences would

have be incorporated into object descriptors in order to preserve object invariance with scale changes. Clearly, object description at a high level can be more economical when the low level description accurately represents the object's spatial frequency content. Constant gain across frequencies is the simplest way to begin a scale-invariant description.

An auditory object (e.g., a voice or a tune) is characterized by the relations between the component auditory frequencies. To a first order, moving closer or further corresponds to an overall intensity change. If changing intensity changed the internal frequency relations, the internal object descriptor would have to be intensity (distance) dependent. At sounds near threshold, their internal representation will, to some extent, inevitably reflect the ear's sensitivity. Above threshold, it would be desirable for object descriptions to be intensity invariant. Indeed, the further above threshold, the less loudness varies as a function of frequency.

3. Why You Can't See the Forest for the Trees: The Economics of Connectivity

In audition, frequencies above 20,000 Hz are too high to be audible and amplification will not make them audible. In vision, the opposite occurs. For example, in Parish & Sperling's (1987a) letter discrimination task with 2-octave-wide bands, the higher the center frequency of the band, the more discriminable the letters. Is there an upper visual object frequency at which the trend to improved discriminability reverses? What happens as visual stimuli are filtered in higher and higher spatial frequency bands?

Insert 4 here.

Consider first an ideal letter stimulus in which the letter has perfect zero/one step edges (Fig. 4b). When such an edge is bandpass filtered, for example, by a difference of Gaussians filter (Fig. 4a), the results are alternate dark/light stripes centered at the edge (Fig. 4c). When the frequency spectra of different filter bands are related by simple translation on a log frequency axis (i.e., the frequency filters differ only in scale), the basic shape of the bandpass filtered edge is independent of the center frequency of the filter. By suitably scaling the abscissa, we arrive at a canonical representation like that of Fig. 4b and

4c in which, as frequency band changes, only the distance between edges changes, not the shape of the edge representation.

Obviously, the higher the filter's center frequency, the narrower the edge representation. But spatial bandpass filters operate on object--not retinal--frequencies. As the frequency of a band is made higher and higher, the viewer can preserve a constant retinal frequency by approaching the letter closer and closer, ultimately, with a microscope. Locally, the edges filtered at different center frequencies f will produce exactly the same retinal images when the viewing distance is proportional to $1/f$. What changes with viewing distance is the retinal distance between opposite edges of a letter stroke. In normal viewing, the thickness of a letter stroke may be a few minutes of arc. With sufficient magnification, the width of the letter stroke ultimately comes to occupy degrees or even hundreds of degrees of visual angle. For extremely high frequency bands, when a letter has been enlarged sufficiently to make its edges visible, it is physically impossible to view the whole letter at one time. This is the classical problem of trying to read a newspaper under a high power microscope. Even individual letters become unrecognizable.

For the cases of letters printed with real ink on real paper, or illuminated letters on real CRT screens, the scaling problem is similar to the case of ideal letters. High spatial frequencies represent local texture information about the ink and paper or about how the CRT screen is populated with phosphor--local texture that obscures the larger landscape. This is the problem of being unable to see the forest for the trees (Sperling & Parish, 1985). The scale of observation is inappropriate for the object being observed.

Economy of connection. What is the appropriate scale of observation? This is dictated by the principle of *economy of connection.* To compute relations, sensors must be connected to each other. It is uneconomical for every sensor in a large field to be connected with and to compute its relations to every other sensor; typically sensors are connected only to immediate neighbors and to nearby neighbors. A sensor and its similar neighbors form a kind of module. The size of the visual receptive field viewed by a module is inversely related to the module's characteristic spatial frequency. In this arrangement, the optimal scale of observation is when the object is of the same order of size as the receptive field of the sensors so that the object can be entirely described within a module. Indeed, the spatial frequency band that is most efficient for letter recognition is one cycle per object, i.e., the same order of size as the object. The size relation between letters and the spatial frequencies that were empirically found to be most

efficient in identifying them is illustrated in Fig. 5. Note that several such spatial frequency filters, in different orientations and phases, would be required to discriminate between the 26 upper-case letters.

Insert 5 here.

When, on the other hand, much smaller-sized sensors are used to describe a large object then, in a hierarchically organized system, it requires communication between modules, communication that occurs only at higher levels. Empirically, using high spatial frequencies to describe large objects results in a loss of perceptual efficiency. Below, we consider some reasons why communication between modules might entail a loss of information.

4. Two Processing Systems

The basic thesis of this section is that there are two processing systems: a Fourier system that uses phase information and makes local computations within a small local area (a module); and a nonFourier system that discards phase information and coordinates computations made in different modules. We approach these general issues by considering an analogy from radio communication.

Demodulation.

High frequency carriers. In AM (amplitude modulated) radio communication, the amplitude of a high frequency *carrier* wave is modulated by the voice frequencies that are to be transmitted. Voice frequencies of up to about 10,000 Hz are transmitted as amplitude modulations of a 100,000 Hz carrier frequency. The process of extracting the low-frequency modulating signal from the high-frequency carrier frequency is *demodulation.*

In visual object recognition, an analogous process of modulation occurs when an object A whose overall shape is--by definition--characterized by frequencies around one cycle per object, is differentiated from its surround by higher frequencies. This would occur if the object had a surface texture that differed from the background texture. In that case, a spatial filter tuned to one of the dominant spatial frequencies in A, say f_a, would record a large response wherever A was present, and smaller responses elsewhere.

Another object, B, might contain an intermediate amount of f_a but a larger amounts of other spatial frequencies (Fig. 5b). Textures function, like colors, to characterize objects.

Insert 6 here.

There is a perfect analogy between a characteristic texture frequency and an AM carrier frequency. The goal of demodulation is the same in both instances. In AM modulation, demodulation means estimating how much carrier signal (its amplitude) is present at each instant in time. In a texture-defined object, the problem for the visual system is estimating how much carrier signal is present at each point in space.

A simple form of demodulation involves fullwave rectification (taking the absolute value) of the signal (Fig. 5d). The modulated carrier is rectified and then lowpass filtered (Fig. 5e and f) to remove the carrier and higher frequencies; only the original modulating signal remains. In the visual system, after the initial receptors, positive and negative signals are carried in separate channels (for example, on-center, off-center neurons). An alternative method of transmitting positive and negative quantities is to modulate the resting firing rate of a neuron up and down. The advantage of using separate positive and negative channels is that zero signal means zero impulses per second, and so the average firing rate is minimized.

Insert 7 here.

When there are separate on- and off-channels, to preserve the sign of the signal at at subsequent synapses, the target synapses for on- and off- neurons must operate in opposite directions (excitation or inhibition). Fullwave rectification is accomplished when the target synapses of the on- and off-channels operate in the same direction (see Fig. 7). In terms of the high-frequency sensors of the carrier frequency, fullwave rectification means that the sensors communicate information about their location and the magnitude (but not the sign) of their responses to the next higher level of the system. On the other hand, halfwave rectification corresponds to independent analyses of the on- and off-channel signals, a process

that has been proposed as a mechanism for locating luminance boundaries (Watt & Morgan, 1985).

Converting the output of high frequency detectors to lower frequencies (demodulation) is a critical component of object recognition because objects are defined most efficiently and most economically in the lowest feasible frequency range. The computational advantage of a hierarchical demodulatory scheme is that pattern recognition at the higher level can use a single computation that is independent of the scale or the contrast of the sensors that are transmitting information from lower levels. Because, in this context, demodulation involves going from higher to lower spatial frequencies, the pattern recognition algorithm can operate at the lowest frequency. Using the lowest possible frequencies is computationally efficient because of the economy of connection: A neuron and its immediate neighbors span the field of interest.

In letter discrimination, the experimentally measured efficiency of discrimination was highest ($eff = 0.4$) at 1 cycle/object, the lowest usable band of spatial frequencies. Efficiency decreased to 0.1 at 10 cycle/object. Informational inefficiency is an unavoidable consequence of rectification because a computation that discards the sign of the input cannot be as efficient as one that takes sign into account. However, statistical inefficiency is a consequence of, not direct evidence for, demodulation or rectification. For direct evidence, we turn to other paradigms.

Direct evidence for two computational regimes in motion perception.

Perhaps the most convincing way to demonstrate two computational systems is to embed two conflicting cues, one aimed at each system, in the same stimulus. The best examples occur in the domain of motion stimuli. The image of moving stimulus is a three-dimensional (3D) function that gives luminance $l(x,y,t)$ as a function of x,y,t. To represent this 3D function on a printed page, we use x,t) cross-sections that omit the y dimension as illustrated in Fig. 8a and b. Figure 8a shows a frame-by-frame representation of a rightward moving black bar; Fig. b, shows the corresponding x,t cross-section. Superimposed on the bar's x,t crossection in Fig. 8b is a sinewave. This sinewave is the x,t crossection of a sinusoidal grating that is moving at the same velocity as the bar. This particular moving grating

represents one of the largest Fourier components of the moving bar.

Insert 8 here.

Figure 8c shows a space-time representation of a motion stimulus that has conflicting cues--a contrast-reversing bar, based on Anstis's (1970) reversed phi phenomenon. The bar steps sideways across a gray field, alternating its contrast between black (-1) and white (+1) on each step.

In the x, t cross-section, the bar moving to the right appears as a contrast-reversing diagonal slanting to the right. However, the Fourier sinewave components of the contrast-reversing bar are slanted down and to the left, indicating Fourier motion to the left. By rectifying the contrast-reversing bar, i.e., taking the absolute value of its contrast, the result is the stimulus of Fig. 8b; and its Fourier sinewave components are slanted downward to the right, indicating rightward motion.

When the contrast-reversing bar is viewed from near, it seems obviously to move to the right. However, when it is viewed peripherally, or from a distance, or at very low contrast, it apparently moves to the left (Chubb & Sperling, 1988a, 1989b). This clearly indicates that observers make two different kinds of motion computations.

For motion stimuli, Chubb and Sperling arrive at a functional discrimination between first-order (Fourier, direct) and second-order (nonFourier, rectified) processes. (Note 2.) They refer to the first-order regime as a Fourier process because it is well modeled by linear filters that utilize the Fourier decomposition of the stimulus. The second-order system, which involves rectification, operates better over larger retinal distances than does the Fourier system. Consistent with the lower efficiency of rectification, the second-order system has higher contrast thresholds than the Fourier system. Certain values of the parameters of viewing, (such as small retinal size, peripheral retinal location, and low stimulus contrast) increase the relative strength of the first-order versus the second-order computation.

While the contrast-reversing bar is a simple demonstration stimulus, it does not enable one to discriminate between different second-order computations. Chubb & Sperling (1989b) demonstrate a sideways stepping, contrast-reversing grating, a stimulus which displays obvious second-order motion and

in which halfwave rectification, alone or in combination with any reasonable temporal transformation, can be excluded. In displays that were designed to exclude fullwave rectification and admit only halfwave rectification or Fourier motion analysis, Sperling and Chubb (1987) found only such weak second-order motion, that they did not preclude alternative explanations. Thus, the predominant mechanism of second-order motion perception involves fullwave rectification. Fullwave rectification also is the dominant mechanism in second-order texture-orientation processing of the x, y patterns that represented the x, t cross-sections of the motion stimuli in their motion experiments (Note 3).

In motion perception, there is a well-established distinction between short-range and long-range motion processes (e.g., Braddick, 1974; Pantle & Picciano, 1976; Westheimer & McKee, 1977; Victor & Conte, 1989b). The inadequacy of first-order motion processing has been amply documented by Ramachandran, Madhusudhan Rao, and Vidyasagar (1973), Sperling (1976), Lelkens & Koenderink (1984) Pantle & Turano (1986), and Victor & Conte (1989a). The properties adduced for the non-first-order system are generally those described above, plus a relative insensitivity to the eye of origin of successive stroboscopic stimuli. To these can be added the observations of Dosher, Landy, & Sperling (in press) and Landy, Sperling, Dosher, & Perkins (1987) that first-order motion supports the kinetic depth effect (KDE, Wallach & O'Connell, 1953) whereby 3D structure is perceived in 2D moving stimuli, whereas KDE induced by second-order motion stimuli is weak and of enormously lower resolution (e.g., Prazdney, 1987).

The computations of first-order motion are well embodied in the quite similar models of Watson and Ahumada (1983), van Santen and Sperling (1984), and Adelson and Bergen (1985), which van Santen & Sperling (1984, 1985) supplement with surprising predictions of first-order relationships that are verified experimentally. What Chubb & Sperling (1988b, 1989a, 1989b) have added is a computational specification of a second-order motion system together with methods for producing stimuli that can be proved to be directly aimed at one or the other system. As a consequence, it is easily shown that (retinal) short-range and (retinal) long-range are inadequate system descriptions because there is a broad intermediate range in which both computations operate.

Orientation and motion perception. Strong evidence for two computational regimes is obtained in studies of orientation detection in textured patterns as well as in studies of direction discrimination in one-

dimensional motion perception. Indeed, these two problems involve formally identical computations (van Santen & Sperling, 1985; Chubb & Sperling, 1987, 1988b). Figures 8d and 8e show demonstrations of stimuli that show obvious apparent motion (when presented as motion stimuli) and obvious slant (orientation) when presented in x,y, as in the illustration.

The stimuli of Fig. 8d and 8e are driftbalanced: that is, they are exemplars of random stimuli in which the *expected* motion (or orientation) is exactly equal for every pair of oppositely-directed component Fourier frequencies (Chubb & Sperling, 1988b). (The overlaid sine gratings of Figs. 8b and 8c are an example of two oppositely-directed Fourier components--their slants in the x,t cross-sections are equal and opposite. The stimuli of 8d and 8e are microbalanced. This means, roughly, that every little area, whatever its shape, in these stimuli is driftbalanced. Thus, the obvious orientation in these x,y stimuli is invisible to every linear Hubel-Wiesel cell (i.e., neurons with receptive fields such as illustrated in Fig. 5). The motion in the x,t versions of the stimuli is invisible to any standard (Fourier or Reichardt-equivalent) motion detector. Rectification is required to make the x,t motion or x,y orientation in these stimuli accessible to standard motion or orientation analysis (Chubb and Sperling, 1988).

Figure 8f shows an example of a texture quilt (Chubb & Sperling, 1989a). To make the overall motion in such a stimulus accessible to motion analysis requires an initial stage of selective spatial filtering (texture grabbing) followed by rectification and standard motion (or texture) analysis. No purely temporal transformation, no matter how complex and nonlinear, can make this motion (or texture) accessible to first-order analysis (Chubb & Sperling, 1989a). The squares of the texture quilt are each filled with their unique carrier frequencies, and these frequencies must first be extracted and demodulated to reveal the larger pattern.

Since the work of Schade (1952) and DeLange (1954), first-order Fourier-based computations have been the cornerstone of psychophysical analysis (Note 4). The examples of Figure 8c,d,e show the limitations of first-order linear analysis and the necessity of postulating second-order computations. Texture quilts (Fig. 8f) provide a fine tool for studying the spatial properties of second-order motion.

Distance estimation. Distance estimation experiments also yield evidence for two processing systems, one Fourier and one rectifying system, in spatial vision. In a three-bar distance estimation experiment, an observer must judge whether a central line is equally spaced between two flanking bars. In

a two bar task, the observer judges the distance between two widely separated bars. On a grey background, for widely separated bars, it matters not whether any of the bars is black and the other white, or whether both are white or black (Burbeck, 1987). Indeed, when "bars" are defined by patches of high frequency gratings so that the the bars themselves do not differ in average luminance from their surround, distance judgments are as accurate as with solid bars.

That observers accurately judge the distance between widely separated grating patches virtually guarantees a demodulatory process in which the grating patch is converted to a solid patch. To solve the distance task with a first-order computation, i.e., with linear receptive fields and without demodulation, would involve horrendous complications. The linear receptive fields needed for distance judgments are dumbbell-shaped receptive fields with one end of the dumbbell in each patch. Receptive fields would be phase sensitive with responses that varied from negative to zero to positive depending on just where in the receptive field the stimulus patches fell. Receptive fields would have to be duplicated for all orientations, distances, and pairs of frequencies. Otherwise, for example, distance judgments would be impossible if the two bars being judged were of different spatial frequencies. In fact, the distance between two grating patches of different spatial frequencies is judged as accurately as the other distances (Burbeck, 1988). Demodulation resolves all these problems of first-order computations at once by transposing distance judgments to the lowest common domain.

For closely spaced bars, there is a significant difference between the same-contrast and opposite-contrast patterns. Thus there again is the telltale *rectification-times-size interaction* that indicates two processing regimes: rectification dominating at large retinal sizes, and direct computation at small retinal sizes.

Klein and Levi (1985) were led to propose two processing regimes based on a size-times-stimulus-type interaction in a bisection type of distance judgment. The observer's task was to estimate which of two horizontal flanking lines was closer to a central line. The flanking lines were either directly above and below the central line or displaced sideways. With large retinal images, the sideways displacement was immaterial; with small retinal images, it was critical. This difference between results at close spacings and far spacings of lines in psychophysical judgments led Klein and Levi (1985) to postulate two regimes of detection mechanisms. They proposed a regime for small-size computations that relied on efficient linear

filters (direct computation), but they did not propose a specific regime for large-size computations. However, the failure of the first-order small-size computation to account for the large-size results is consistent with the rectification proposal, although Klein and Levi's results do not specifically require rectification.

Two Processing Regimes: Conclusions.

For bandpass filtered objects, different computations will be carried out depending on whether the object can be coded by neighboring sensors or whether it requires the coordination of information from distantly separated sensors. Nearest neighbor computations can use linear filters and can be highly efficient (first-order computations). Distant computations require demodulation (which is carried out by fullwave rectification) and information that is coordinated at higher levels of the computational hierarchy (second-order computations). The second order computations, because they use rectification for demodulation, sacrifice statistical efficiency (impaired compared to ideal detectors) for computational simplicity (improved relative to attempting the computation at the same hierarchical level). An interesting unresolved issue that would relate the stages and systems of this paper concerns the extent to which the noise sources associated with each type of computation (first-order, second-order) are shared or independent.

It seems obvious that counting and labeling (rectification) operations will predominate over linear processes at higher perceptual and cognitive levels of processing. The surprise has been that simple rectification occurs so early in processing, being involved in retinal gain control and in the earliest stages of motion and pattern analysis. Presumably the appearance of rectification early in visual processing is determined by two factors: its economy of neural connectivity in a hierarchically organized nervous system and its ecological adequacy in our natural environment.

Acknowledgments

Reference Notes

Note 1. Informal observations. Variations in contrast and luminance were not reported in Parish & Sperling (1987a).

Note 2. The nomenclature was suggested by P. Cavanaugh (1988) in "Motion: The long and the short of it." Paper at the Workshop on Visual Form and Motion Perception: Psychophysics, Computation, and Neural Networks. Boston University, March 5, 1988.

Note 3. Although it is embedded in a much more complex framework, Grossberg and Mingolla (1985) incorporate a fullwave rectifying stage in their general model of texture and boundery perception that appears to deal with second-order stimuli. However rectification and similar nonlinear operations such as squaring do not, in and of themselves, imply second-order processing. For example, the Adelson and Bergen's (1985) detector of directional motion enery is equivalent to the Reichardt motion model (van Santen & Sperling, 1984, 1985) and to Knutsen & Grandlund's (1983) texture-orientation model. All of these models embody a nonlinear squaring stage (or the equivalent) and they merely perform first-order computations; none can detect the second-order motion or orientation.

Note 4. Ives (1922) anticipated subsequent linear theories of visual threshold phenomena but he was ignored by the psychophysicists of his time because they did not understand linear systems theory.

References

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

Anstis, S.M. (1970). Phi movement as a subtraction process.

Vision Res., **10,** 1411-1430, 1970. Banks, M. S. Geisler, W. S., & Bennett, P. J. (1989). *Vision Research,* **29.** (In press.)

Barlow, H. B. (1956). Retinal noise and absolute threshold. *Journal of the Optical Society of America,* **46,** 634-639.

Barlow, H. B. (1957). Increment thresholds at low intensities considered as signal/noise discrimination. *Journal of Physiology,* **136,** 469-488.

Berliner, J. E. & Durlach, N. I. (1973). Intensity perception. IV. Resolution in roving-level discrimination. *Journal of the Acoustical Society of America,* **61,** 1677-1585.

Braddick, O. (1974). A short-range process in apparent motion, *Vision Res.,* **14,** 519-527.

Burbeck, C. A. (1987). Position and spatial frequency in large-scale localization judgments. *Vision Research,* **27,** 417-427.

Burbeck, C. A. (1988). Large-scale relative localization across spatial frequency channels. *Vision Research,* **28,** 857-859.

Carlson, C. R. and Klopfenstein, R. W. (1985). Spatial-frequency model for hyperacuity. *Journal of the Optical Society of America - A,* **2,** 1747-1751.

Chubb, C. and Sperling, G. (1987). Drift-balanced random stimuli: A general basis for studying non-Fourier motion perception, *Investigative Ophthalmology and Visual Science* **28,** 233.

Chubb, C. and Sperling, G. (1988a). Processing stages in Non-Fourier Motion Perception, *Investigative Ophthalmology and Visual Science,* **29,** 266.

Chubb, C. and Sperling, G. (1988b). Drift-balanced random stimuli. A general basis for studying non-Fourier motion perception. *Journal of the Optical Society of America - A,* **5,** 1986-2006.

Chubb, Charles, and George Sperling. (1989a). Second-order motion perception: Space-time separable mechanisms. *Proceedings: 1989 IEEE Workshop on Motion.* Washington, D.C: IEEE COmputer Society Press. Pp. 126-138.

Chubb, Charles, and George Sperling. (1989b). Two motion perception mechanisms revealed by distance driven reversal of apparent motion. Proceedings of the National Academy of Sciences, USA, 86, 2985-2989.

DeLange, H. Dzn. (1954). Relationship between critical flicker-frequency and a set of low frequency characteristics of the eye. *Journal of the Optical Society of America*, 44, 380-389.

Dosher, B. A., Landy, M. S. and Sperling, G. (In press). Kinetic depth effect and optic flow: 1. 3D shape from Fourier motion. *Vision Research*, 29.

Green, D. M. and Swets, J. A. (1966). Signal detection theory and psychophysics. New York: Wiley, 1966.

Geisler, W. S. (1989). Sequential idea-observer analysis of visual discriminations. *Psychological Review*, 96. (In press.)

Grossberg, S. and Mingolla, E. (1985). Neural dynamics of perceptual grouping: Textures, boundaries, and emergent segmentations. *Perception and Psychophysics*, 38, 141-171.

Ives, H. E. (1922). A theory of intermittant vision. *Journal of the Optical Society of America*, 6, 343-361.

Jamar, H. H. T., Campagne, J. C., and Koenderink, J. J. (1982). Detectability of amplitude- and frequency-modulation of suprathreshold sine-wave gratings. *Vision Research*, 22, 407-416.

Jesteadt, W., Wier, C. C., and Green, D. M. (1977) Intensity discrimination as a function of frequency and sensation level. *Journal of the Acoustical Society of America*, 61, 169-177

Jamar, H. H. T., and Koenderink, J. J. (1985). Contrast detection and detection of contrast modulation for noise gratings. *Vision Research*, 25, 511-521.

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

Knudsson, H. and Granlund, G. H. (1983). Texture analysis using two-dimensional quadrature filters. *IEEE Computer Society Workshop on COmputer Achitecture for Pattern Analysis and Image Database Management -- CAPAIDM. Silver Spring, Maryland: IEEE COmputer Society. Pp. 206-213.*

Kuffler, S. W. (1953). Discharge patterns and functional organization of mammalian retina. Journal of Neurophysiology, 16, 37-68.

Landy, M. S., Sperling, G., Dosher, B. A., and Perkins, M. (1987). From what kind of motions can structure be inferred? *Investigative Ophthalmology and Visual Science*, 28, No. 3, *ARVO Supplement*,

233.

Legge, G. E. and Foley, J. M. (1980). Contrast masking in human vision. *Journal of the Optical Society of America*, **70**, 1458-1471.

Legge, G. E. Kersten, D., and Burgess, A. E. (1987). Contrast discrimination in noise. *Journal of the Optical Society of America - A*, **4**, 391-404.

Lelkens, A. M. M. and Koenderink, J. J. (1984). Illusory motion in visual displays. *Vision Research*, **24**, 1083-1090.

MacMillan, N. A. (1987). Beyond the categorical/continuous distinction: A psychophysical approach to processing modes. In S. Harnad (Ed.), *Categorical Perception*. New York: Cambridge University Press. Pp. 53-85.

Marr, D. (1982). *Vision. A computational investigation into human representation and processing of visual information*. San Franciso: W. H. Freeman and Company.

Nakayama, K., and Silverman, G. (1985). Detection and discrimination of sinusoidal grating displacements. *Journal of the Optical Society of America, A: Optics and Image Science*, **2**, 267-274.

Pantle, A. and Picciano, L. (1976). A multistable movement display: Evidence for two separate motion systems in human vision, *Science*, **193**, 500-502.

Pantle, A. and Turano, K. (1986). Direct comparisons of apparent motions produced with luminance, contrast-modulated (CM), and texture gratings, *Investigative Ophthalmology and Visual Science*, **27**, 141.

Parish, D. H. and Sperling, G. (1987a). Object spatial frequencies, retinal spatial frequencies, and the efficiency of letter discrimination. *Mathematical Studies in Perception and Cognition*, *87-8*, New York University, Department of Psychology, 1987. Pp. 30, Figs. 8.

Parish, D. H. and Sperling, G. (1987b). Object spatial frequency, not retinal spatial frequency, determines identification efficiency. *Investigative Ophthalmology and Visual Science, ARVO Supplement*, **28**, No. 3, 359.

Pavel, M., Sperling, G., Riedl, T., & Vanderbeek, A. (1987). The limits of visual communication: The effect of signal-to-noise ratio on the intelligibility of American Sign Language. Journal of the Optical Society of America - A, **4** 2355-2365.

Pelli, D.G. (1981). *Effects of Visual Noise*. (Ph.D. dissertation, University of Cambridge, Cambridge,

England).

Prazdny, K. (1987). Three-dimensional structure from long-range apparent motion. *Perception*, 15, 619-625.

Ramachandran, V. S. Madhusudhan Rao, and Vidyasagar, T. R. (1973). Apparent movement with subjective contours. *Vision Research*, 13, 1399-1401.

Riesz, R. R., (1928). Differential intensity sensitivity of the ear. *Physics Review*, 31, 867-875

Robinson, D. W., and Dadson, R. S., (1956) A redetermination of the equal-loudness relations for pure tones. *British Journal of Applied Physics* 7, 166-181

Schade, O. H. (1952). Optical and photelectric analog of the eye. *Journal of the Optical Society of America*, 46, 721-739.

Scharf, B., and Buus, S., (1986). Audition 1. Stimulus, Physiology, Thresholds. In Boff, K., Kaufman, L. & Thomas, J. (Eds.), *Handbook of Perception and Performance. Vol. 1*. New York: Wiley, 1986. Pp. 14-1 to 14-71.

Shapley, R. and Enroth-Cugell, C. (1986). Visual adaptation and retinal gain controls. *Progress in Retinal Research*, 3, 263-246.

Sperling, G. (1976). Movement perception in computer-driven visual displays. Behavior, Research, Methods and Instrumentation, 8, 144-151.

Sperling, G. and Chubb, C. C. (1987). Non-Fourier motion and texture perception. Paper presented at *AFOSR Program in Visual Information Processing*, Annapolis, MD, December 16, 1987.

Sperling, G. and D. H. Parish. (1985). Forest-in-the-Trees Illusions. *Investigative Ophthalmology and Visual Science, ARVO Supplement*, 26, 285.

van Nes, F. L. and Bouman, M. A. (1967). Spatial modulation transfer in the human eye. *Journal of the Optical Society of America*, 57, 401-406.

van Santen, J. P. H., and Sperling, G. (1984). Temporal covariance model of human motion perception. *Journal of the Optical Society of America A: Optics and Image Science*, 1, 451-473.

van Santen, J.P.H. and Sperling, G. (1985). Elaborated Reichardt Detectors, *J. Opt. Soc. Am. A: Optics and Image Science*, 2, 300-321.

Victor, J. D., and Conte, Mary, M. (1989a). Cortical interactions in texture processing: scale and dynamics. *Visual Neuroscience*, in press.

Victor, J. D., and Conte, Mary, M. (1989b). *Motion mechanisms have only limited access to form information.* Manuscipt.

Wallach, H., and O'Connell, D. N. (1953). The kinetic depth effect. *Journal of Experimental Psychology,* **45,** 205-217.

Watson, A. B. and Ahumada, A. J. Jr. (1983). A linear motion sensor. *Perception,* **12,** A17.

Watt, R. J. & Morgan, M. J. (1985). A theory of the primitive spatial code in human vision. *Vision Research,* **25,** 1661-1674.

Westheimer, G. and McKee, S. P. (1977). Perception of temporal order in adjacent visual stimuli. *Vision Res.,* **17,** 887-892.

Figure Captions

Figure 1. Upper: A sample of the letter G filtered in five spatial frequency bands. The number above the band indicates the 2D mean frequency (cycles per letter height) of the approximately two-octave wide band. Lower: The filtered letter plus noise in the same bands with a signal-to-noise ratio of 0.50 in all panels. The effective s/n in the reproduction is somewhat lower. (From Parish and Sperling, 1987a.)

Figure 2. Three stages of visual processing. Suns indicate stationary noise sources, + indicates summation components, boxes indicate linear filters, triangles indicate amplifiers, and the blocked triangle indicates a rectifier. Double boxes show input/output relations. *Adaptation.* The visual input is u, the light-adapted output is v. The center/surround organization is produced by two pairs of separable spatial and temporal filters, F_1, F_2 respectively, whose impulse responses are indicated by in boxes x and t. The surround controls the gain of amplifier K_1 to produce Weber law light adaptation (Sperling & Sondhi, 1968). The dark-adapted impulse responses are shown by the dark lines v_x versus x and v_t versus t. The light lines show the light-adapted input/output relation. Dark noise adds directly to the input, sensory noise adds after light-adaptation. *Contrast-gain control.* The graphs in box F_3 indicate various oriented and nonoriented spatial filters that operate on the light-adapted signal v. The fullwave rectified outputs, indicated by the blocked triangle R, control the gain of the corresponding amplifiers K_3, only one rectifier and amplifier is shown. A typical input-output function w vs v is shown in the insert. *Postsensory stages.* The first-order gain-controlled signal w_1 and the second-order gain-controlling signals w_2 are combined with each other and with noise. The postsensory, decision, mnemonic, and response processes are not detailed. The overall system output is z.

Figure 3. The contrast modulation transfer function (MTF) and the frequency ranges of the letter-in-noise stimuli of the Parish & Sperling letter discrimination experiments. The MTF gives the contrast detection threshold (in percent contrast modulation) for sine gratings as a function of their retinal spatial frequency; it is based on data of van Nes and Bouman, 1967. Stipling indicates the area, near threshold, where quantal noise and additive sensory noise predominate. Noise also predominates in

the whole upper portion of the graph, where stimuli are invisible. Each open, downward facing rectangle indicates the approximate half-bandwidth of a frequency bands (1,...5, Fig. 1) at one of the extreme viewing distances a,d used by Parish and Sperling (1987a). The horizontal placement of the corresponding number/letter symbol indicates the mean *retinal* frequency of the stimulus; symbols (b,c) for intermediate viewing distances also are shown. The stimulus symbols are placed vertically at a contrast ≥ 10% to indicate that for all larger contrasts (downward in the figure), performance is independent of contrast, i.e., it is controlled by multiplicative noise.

Figure 4. When is it impossible to see both the forest and the trees? The retinal image of a bandpass filtered boundary can remain independent of filter frequency f_ω by varying the viewing distance d. For two boundaries that are physically separated by a distance ϕ, the retinal distance r is ϕ/d. Example: (a) Impulse response of f_ω, a family of linear spatial filters that differ only in their scale ω (e.g., ω is the center frequency of the passband). (b) A retinal illuminance distribution $l(x)$ representing the left and right boundary of an ideal white stripe. (c) The retinal illuminance distribution of the filtered image $l*f_\omega$ with viewing distance d chosen so that $d=1/\omega$. This choice of d normalizes (leaves unchanged) the images of the boundaries (neighborhoods of dashed lines in b and c). Only r, and not the boundary images, varies with ω. (d) Ultimately, for very large ω, r grossly exceeds 360 degrees. When it is necessary to view the boundary from extremely close in order to achieve visible detail, it will then be impossible to simultaneously resolve a boundary (see a tree) and to see both boundaries (see the forest).

Figure 5. The letter T and receptive fields that have a center frequencies of 1 cycle per letter height (in their higher-frequency dimension). (a) The letter T centered in an even symmetric receptive field. The + and - signs indicate the sign of the field's response to spots of light in the indicated areas. (b) Horizontal cross section showing the sensitivity of the receptive field as a function of position. (c, d) The letter T within an odd-symmetric receptive field.

Figure 6. Carrier frequencies, amplitude modulation, and demodulation. (a) A carrier frequency $C(x)=\sin(2\pi f_c x)$. (b) A signal $S(x)$ that consists of an object A which has a large amount of the carrier f_c (one of its characteristic spatial texture frequencies), a second object B which has an intermediate amount, and the background which has a small amount. (c) A representation of the actual

frequencies in the image, the image luminance distribution: $L(x) = S(x) \cdot C(x)$, an amplitude modulated carrier. (In visual scenes, the phase of the carrier is not be preserved across objects.) (d) The rectified image. The absolute value of the image $|S(x) \cdot C(x)|$ is the simplest instantiation of fullwave rectification. (e) A lowpass filter (Normal density function). (f) The result of lowpass filtering (d), *LPcitimes* $|S(x) \cdot C(x)|$. The original signal $S(x)$ has been mostly recovered.

Figure 7. How linear transformations, fullwave rectification, and half-wave rectification can be accomplished in the visual system. *On System* refers to neurons that have an on-center/off-surround receptive field organization (Kuffler, 1953) and which carry signals representing positive local contrasts relative to the surround. *Off System* refers to neurons that have off-center/on-surround receptive fields, and which transmit information about negative local contrasts. (a) When synapses from an On-System neuron onto a target neuron are excitatory and Off-System synapses are inhibitory (indicated by the inverting amplifier −1), the sign of input contrasts is preserved and *first-order* (Fourier) motion analysis of the stimulus can occur. (b) Fullwave rectification occurs when both On- and *Off-System synapses are the same (either excitatory or inhibitory)*; this results in *second-order* signal analysis that is "nonFourier." (c) Positive halfwave rectification occurs when the On-System signals are analyzed independently; negative halfwave rectification refers to independent analysis of Off-System signals. Like fullwave rectification, halfwave rectification is a second-order processing scheme.

Figure 8. Stimuli for analyzing second-order processing. (a) An x, y, t representation of successive frames of a motion stimulus—a black bar moving rightward. (b) An x, t cross-section of (a). A sinewave grating, representing a dominant Fourier component, has been superimposed on the x, t cross-section. Note that the detection of direction of motion in x, t is equivalent to the detection of direction of slant in x, y. (c) An x, t cross-section of a windowed, contrast-reversing bar, a stimulus that appears to move leftward from afar (first order motion) and rightward from afar (second-order motion). A sinewave grating, representing a dominant Fourier component, has been superimposed on the x, t cross-section to indicate the direction of Fourier movement. (d, e) x, t cross-sections of microbalanced stimuli whose motion is invisible to first-order motion detectors and whose slant in their x, y representation is invisible to first-order orientation detectors (e.g., Hubel-Weisel cells). (f) A

texture quilt. The four rows represent four successive frames of a dynamic stimulus. The initial extraction of either the low spatial-frequency texture oriented downward left or of the high frequency texture oriented downward right will enable a first-order motion algorithm to extract the overall leftward motion (overall slant downward to the left). Texture quilts remain microbalanced after any purely temporal transformation and require an initial texture extraction followed by rectification to expose their motion in x, t (or orientation in x, y) to standard analysis.

NOTES

Sampling really is multiplicative noise - see Burgess-Kerten-Legge 1987
(DONE)

Want high intensity stimuli to have a uniform freq spectrum so that
  retinal magnif and minific leave spectral relations unaltered

bar: Fig 2 still is under explained

Mention Prazdney