Spatial Vision, Vol. 4, Nos 2/3, pp. 183–207 (1989)

Three stages and two systems of visual processing

GEORGE SPERLING
Human Information Processing Laboratory, Department of Psychology and Center for Neural Sciences, New York University, New York, NY 10003, USA

Received for publication 1 July 1989

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 the discriminability of strong signals masked by stimulus noise; postsensory noise 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 stimulus *luminance*, and a second-order nonlinear (nonFourier) rectifying regime that uses the absolute value (or power) of stimulus *contrast*. 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 higher processing levels. Rectification sacrifices the statistical efficiency (noise resistance) of the first-order regime for efficiency of neural connectivity and computation.

1. INTRODUCTION

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

bandpassed 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 been confronted before, but consideration 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.

2. 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, b) 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.

2.1. 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

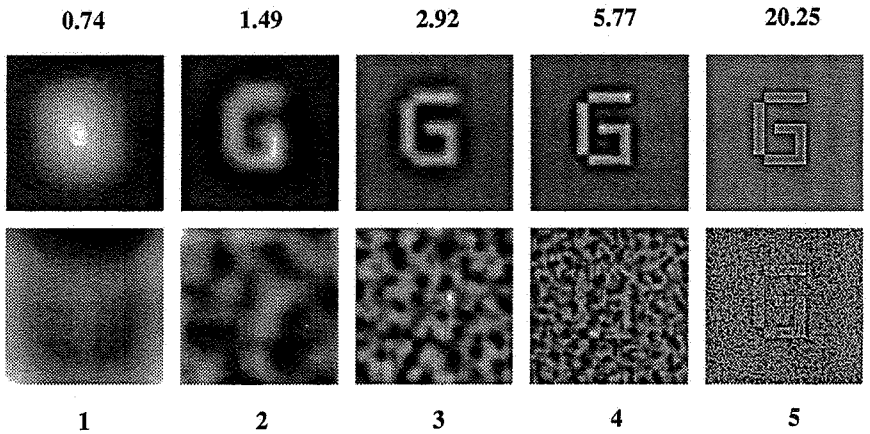

Figure 1. Upper: A sample of the letter G filtered in five spatial frequency bands. The number above the band indicates the 2D mean frequency (cycles per letter height) of the approximately two-octave wide band. Lower: The filtered letter plus noise in the same bands with a signal-to-noise ratio of 0.50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

signal strength until the effective signal-to-noise ratio is sufficient to 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 noise whose expected power is proportional to signal power. Several authors have noted the distinction between additive and multiplicative noise (e.g., Legge and Foley, 1980; Carlson and Klopfenstien, 1985; Legge *et al.*, 1987; Pavel *et al.*, 1987).

Loss of information that results from too-sparse sampling of the stimulus also can be regarded as a form of multiplicative noise (e.g., Legge *et al.*, 1987). To sample a signal means multiplying it by δ in the neighborhoods of the sampled points and by zero elsewhere. (The constant δ is chosen so that the energy of the original signal is preserved in the sampled signal.) For many applications, multiplying parts of a signal by zero (in sampling) is equivalent to multiplying parts of it by random numbers (multiplicative noise). In both cases, as the perturbation increases, *information* eventually is lost (noise), and the lost information cannot be recovered by increasing signal strength (*multiplicative* noise).

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 six examples. In attempting to detect a spatial sine wave grating embedded in noise, with a fixed signal-to-noise ratio, the contrast of the display as a whole 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 *et al.*, 1982; Jamar and Koenderink, 1985). In Parish and Sperling's (1987b) letter-in-noise discrimination task, only the signal-to-noise ratio matters.[1] In discriminating direction of motion, once a contrast of about 0.05 is reached, further increases in contrast do not improve performance (Nakayama and Silverman, 1985). Similarly, in audition, typically the signal-to-noise ratio (and not the absolute signal strength) determines performance. 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 a 4:1 range of contrasts by Pavel *et al.* (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

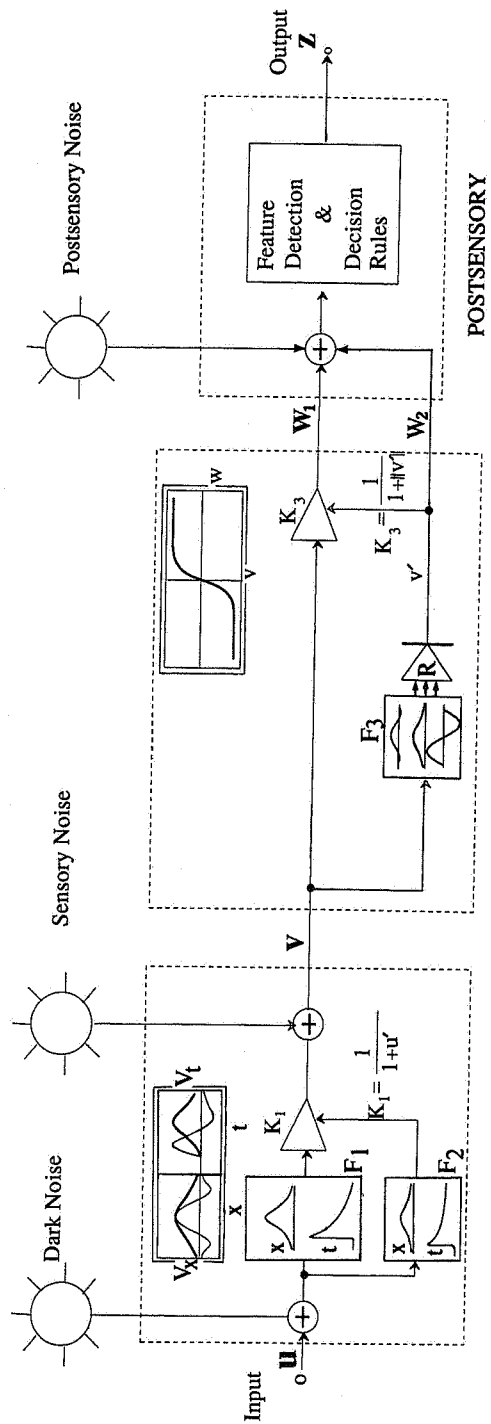

Figure 2. Three stages of visual processing. Sums indicate stationary noise sources, + indicates summation components, boxes indicate linear filters, triangles indicate amplifiers (gain control) and R indicates an array of rectifiers. Double boxes show input/output relations. *Adaptation.* The visual 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 in the double boxes x and t. The surround F_2 controls the gain of amplifier K_1 to produce Weber law light adaptation (Sperling and Sondhi, 1968). The dark-adapted impulse responses are shown by the dark lines v_x vs. x and v_t vs. 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 triangle R, control the gain of the corresponding amplifiers K_3. Only one rectifier and amplifier are shown. A typical input–output function **w** vs. **v** is shown in the double-box insert. *Postsensory stages.* The first-order gain-controlled signal $\mathbf{w}_1$ and the second-order gain-controlling signals $\mathbf{w}_2$ are combined with each other and with noise. The postsensory attentive, decision, mnemonic, and response processes are not detailed. The overall system output is **z**.

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. 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.

2.2. *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 postsensory processing (which includes perceptual, attentional, mnemonic, decision, and response processing). 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 and Enroth Cugell, 1986). For example, the flow chart 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.

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.

2.3. *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.

2.4. *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, then 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 (Weber's Law). The essential fact of sensory noise is that, when viewed from outside the system at moderate to high intensities, apparent noise power increases with absolute modulation rather than

remaining constant. To model noise that apparently increases with the mean luminance (background intensity), 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 intensity gain-control mechanism, mimics an external additive noise that increases as a function of background intensity.

2.4.1. 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: $s/n = \{\text{signal amplitude}\}/\{\text{root-mean-square (RMS) noise amplitude}\}$. Indeed, the effective signal-to-noise ratio of the stimulus at the detector is equivalent to the d' statistic of signal detection theory (Green and 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 assumed 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 *et al.* (1987) and Geisler (1989) applied such an analysis to contrast detection thresholds for sine wave stimuli of spatial frequencies from 5 to 40 c/deg, 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 preneural factors cited above had been taken into account, at the observed thresholds, the stimulus s/n at the detector was constant (Banks *et al.*, 1988). The most parsimonious interpretation is that sensory noise is negligible compared to quantal stimulus noise for these stimuli. For sine wave gratings at lower spatial frequencies than 5 c/deg 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.

2.5. *Postsensory noise*

In the suprathreshold experiments with added external noise discussed above, detection depended only on the signal-to-noise ratio s/n and not 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 and 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 identify (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).

2.6. *The efficiency of detection*

The efficiency *eff* of discrimination is the ratio of s_i^2/n_i^2 required by an ideal observer to the s_h^2/n_h^2 required by a human observer at the same criterion level c of performance: $eff = (s_i/n_i)_c^2/(s_h/n_h)_c^2$. Alternatively, efficiency can be expressed as the squared ratio of human over ideal d' when confronted with the same stimuli: $eff = (d'_h/d'_i)^2$. Efficiency represents an estimate of how much less information than the human the ideal observer needs in order to match the human's performance. 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 and 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 in a range where sensory noise varies enormously suggests that information loss in the visual pathway before the point of postsensory noise was negligible. In terms of noise sources, this means (1) that dark noise and sensory noise were negligible compared to stimulus noise and (2) the postsensory noise in the optimal band was of approximately the same power as the real stimulus noise (because the efficiency was near 0.5). Over the enormous range of spatial frequencies subserved by these channels, efficiency was determined primarily by postsensory noise.

3. 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). The question we address here is: How do the results of Parish and Sperling's letter discrimination experiments relate to what is already known about sine-wave detection? First, Parish and Sperling's (1987a) letter discrimination experiments involved an enormously greater range of low spatial frequencies than are typically used. In the range of retinal spatial frequencies for which data from both kinds of experiments are available, contrast threshold ranges from a minimum of 0.002 at 5–8 c/deg to a maximum of 0.07 at 37 c/deg (e.g., van Nes and Bouman, 1967, cf. Fig. 3). [The frequency of 37 c/deg is the mean retinal frequency of Parish and Sperling's highest frequency band 5d at its longest viewing distance; the most detectable retinal sine frequencies (5c/deg) are produced by Parish and Sperling's frequency bands 3d, 4c, and 5b (Fig. 3) at closer viewing distances.] In the letter-in-noise experiment, observed discrimination efficiency was independent of the mean retinal frequency (varying only with object spatial frequency) whereas, in the sine-wave grating detection experiment, threshold sensitivity for sinusoidal gratings varies from 0.07 to 0.002, a factor of 35, within the same frequency range (Fig. 3). Indeed, the combination of filter frequency with viewing distance in the letter-in-noise discrimination experiment produced retinal frequencies that varied over a range of more than 200:1 (Fig. 3), and discrimination was independent of retinal spatial frequency throughout this entire range.

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

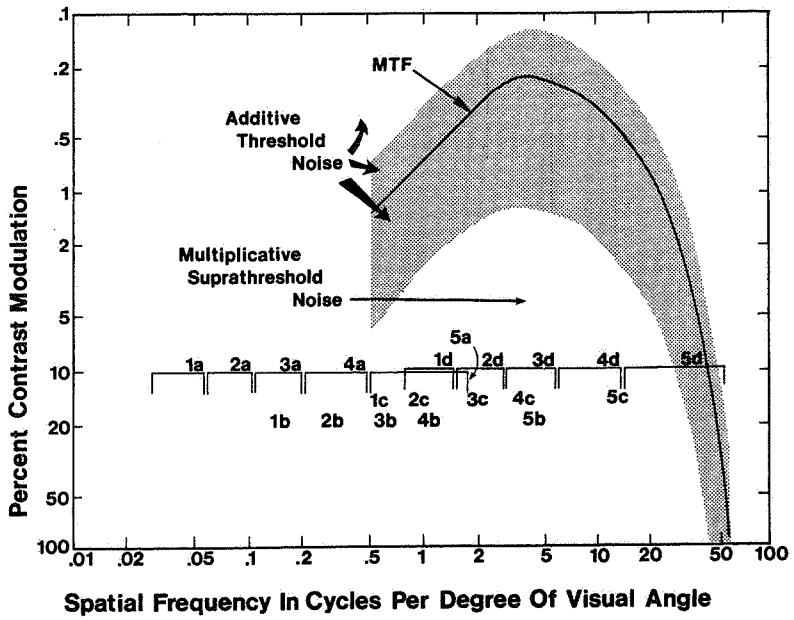

Figure 3. The contrast modulation transfer function (MTF) and the frequency ranges of the letter-in-noise stimuli of the Parish and Sperling letter discriminaton experiments. The MTF gives the contrast detection threshold for sine gratings as a function of their retinal spatial frequency; it is based 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 retinal half-bandwidth of an object frequency band (1–5, Fig. 1) at one of the viewing distances (a = closest, d = furthest) 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 $\geqslant 10\%$ to indicate that for all larger contrasts (downward in the figure), performance is independent of contrast, i.e., it is controlled by multiplicative noise.

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, Jamar and 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.

Detection thresholds for sine gratings vary enormously with retinal spatial frequency in precisely the same range of frequencies where the 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 detection of low-contrast spatial gratings is limited by quantum noise in the stimulus and by sensory noise; the discrimination of letters-in-noise 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.

3.1. 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 5c/deg has the lowest sensory noise, but it has the same decision noise as the channel that processes 37 c/deg, which has 35 times more sensory noise.

3.2. 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 (Reisz, 1928; Robinson and Dadson, 1956; Jesteadt *et al.*, 1977; see Scharf and 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.

3.3. 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 to 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 good approximation, moving closer to or further from the source corresponds to an overall intensity change. If changing intensity were to change the internal frequency relations, the internal object descriptor would have to be intensity (distance) dependent, an enormous complication. Near threshold, the internal representation of sounds (or visual stimuli) inevitably reflects the ear's (or the eye's) frequency sensitivity. Above threshold, it would be desirable for object descriptions to be intensity invariant. Indeed, the further above threshold, the less auditory and visual discriminations vary as a function of frequency.

4. 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, apparently, the opposite occurs. For example, in Parish and Sperling's (1987a) letter discrimination task with two-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?

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. 4(b) and 4(c) 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 arcmin. 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 ink droplets and paper fiber or about how a CRT screen is populated with spots of light-sensitive phosphor—local texture that obscures the larger landscape. This is the problem of being unable to see the forest for the trees (Sperling and Parish, 1985). The scale of observation is inappropriate for the object being observed.

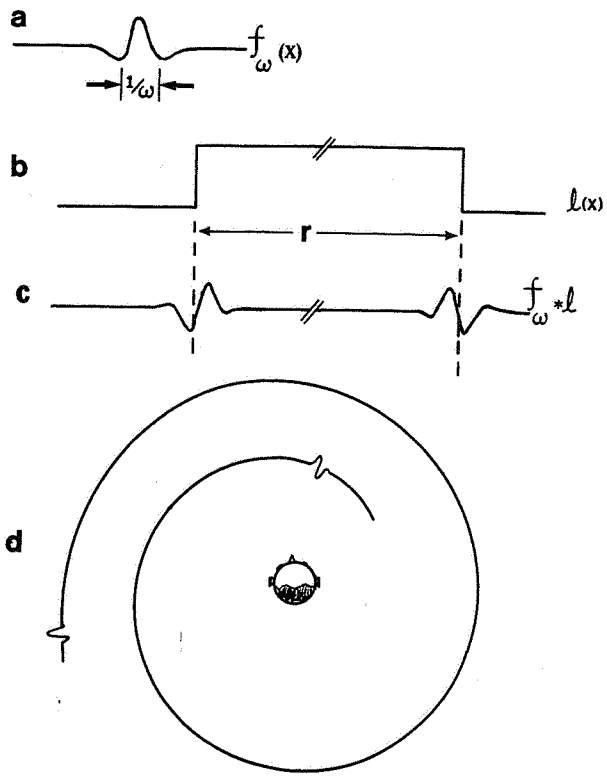

Figure 4. When is it impossible to see both the forest and the trees? The retinal image of a bandpass filtered boundary can remain independent of filter frequency f_ω by varying the viewing distance d. For two boundaries that are physically separated by a distance ϕ, the retinal distance r is ϕ/d. Example: (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 varies wth ω, not the boundary images. (d) Ultimately, for very large ω, r grossly exceeds 360 deg. 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).

4.1. 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 center frequency of the most efficient band for letter recognition

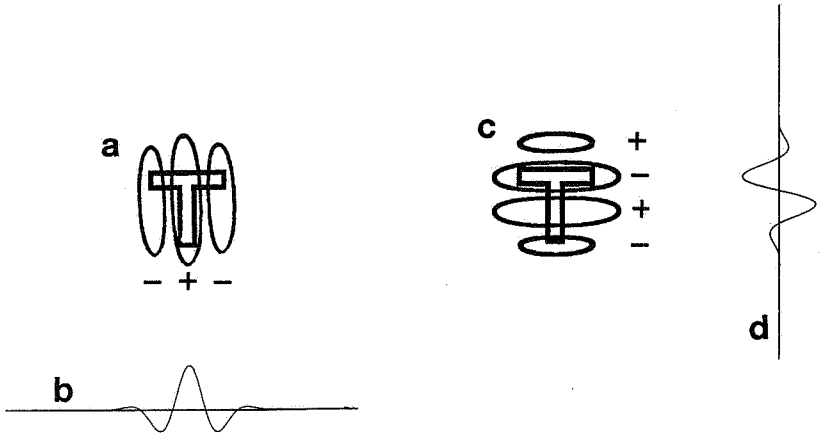

Figure 5. The letter T and receptive fields that have a center frequency of 1 cycle per letter height (in their higher-frequency dimension). (a) The letter T centered in an even symmetric receptive field. The + and − signs indicate the sign of the field's response to spots of light in the indicated areas. (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.

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.

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.

5. 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.

5.1. Demodulation

5.1.1. 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.*

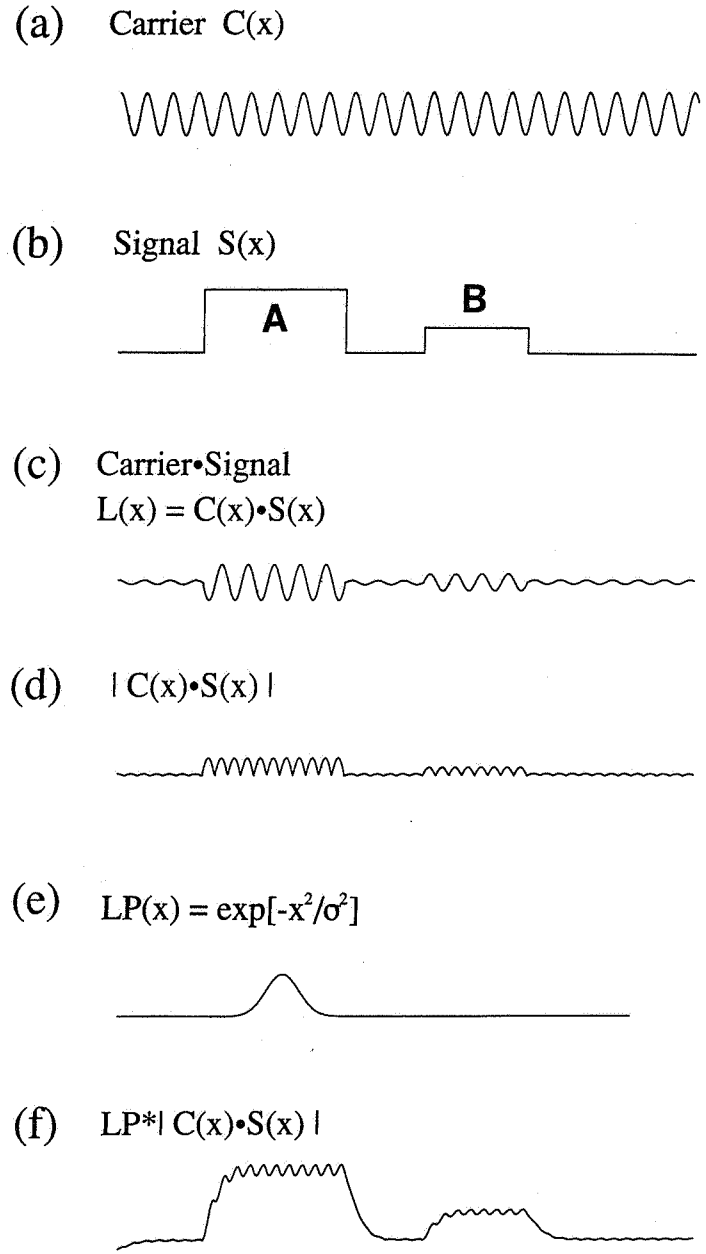

Figure 6. Carrier frequencies, amplitude modulation, and demodulation. (a) A carrier frequency $C(x) = \sin(2\pi f_c x)$. (b) A signal $S(x)$ that consists of an object A which has a large amount of the carrier f_c (one of its characteristic spatial texture frequencies), a second object B which has an intermediate amount, and the background which has a small amount. (c) A representation of the actual frequencies in the image, the image contrast distribution: $L(x) = S(x) \cdot C(x)$, an amplitude modulated carrier. (In visual scenes, the phase of the carrier is not 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), $LP * |S(x) \cdot C(x)|$. (The $*$ indicates convolution). The original signal $S(x)$ has been mostly recovered.

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 (Fig. 6b) but a larger amount of other spatial frequencies. A texture functions, like a color, to characterize an object.

There is a close 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. The difference is that the phase of an AM carrier is consistent throughout the signal; the phase of texture-defined objects is not.

5.1.2. Neural mechanisms of fullwave and halfwave rectification. A simple form of demodulation involves fullwave rectification (taking the absolute value) of the signal (Fig. 6d). The modulated carrier is rectified and then lowpass filtered (Fig. 6e 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 (e.g., 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.

When there are separate on- and off-channels, to preserve the sign of the signal at subsequent synapses, the target synapses for on- and off-neurons must operate in opposite directions (excitation or inhibition, see Fig. 7a). Fullwave rectification is accomplished when the target synapses of the on- and off-channels operate in the same direction (see Fig. 7b). Fullwave rectification means that the high-frequency sensors of the carrier frequency 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 (Fig. 7c) 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 and 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.

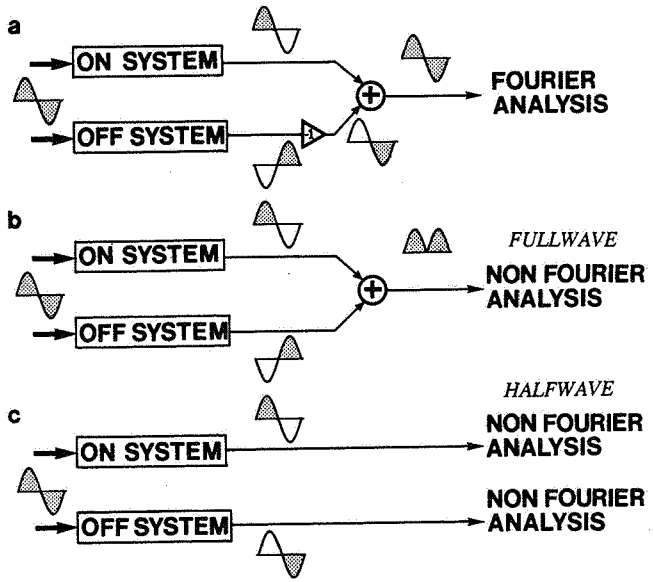

Figure 7. How linear transformations, fullwave rectification, and half-wave rectification can be accomplished in the visual system. *On-System* refers to neurons that have an on-center/off-surround receptive field organization (Kuffler, 1953) and that carry signals representing positive local contrasts relative to the surround *Off-System* refers to off-center/on-surround neurons that 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) 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 could be the essential nonlinearity in a second-order processing scheme.

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 cycles/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.

5.2. Direct evidence for two computational regimes in motion and texture-slant perception

5.2.1 The x, t *cross-section of a motion stimulus.* Perhaps the most convincing way to demonstate 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 a 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

Figs. 8(a) and (b). Figure 8(a) shows a frame-by-frame representation of a rightward moving black bar; Fig. 8(b), shows the corresponding x, t cross-section. Superimposed on the bar's x, t cross-section in Fig. 8(b) is a sinewave. This sinewave is the x, t cross-section 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.

5.2.2. Motion stimuli that can be perceived in either of two directions. Figure 8(c) shows a space–time representation of a motion stimulus that has conflicting cues—a contrast-reversing bar, based on Anstis' (1970) reversed phi phenomenon. The bar steps sideways across a gray field, alternating its contrast between black (-1) and white ($+1$) on each step. The stimulus is Gaussian windowed in space–time so that only a few steps in the middle of the screen are maximally visible.

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. 8(b); and its Fourier sinewave components are slanted downward to the right, indicating rightward motion.

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

5.2.3. First-order motion perception. From an algorithmic point of view, a motion extraction system can be regarded as consisting of three consecutive types of computation: linear filtering, motion extraction, and decision. The first two, filtering and motion extraction, are carried out in parallel everywhere in visual space. A visual stimulus is processed first by linear filters in x, y, t that determine the range and amount of spatio-temporal frequencies that enter into the rest of the system, i.e., the linear filters determine the range of frequencies to which the motion system is sensitive. A second, inherently nonlinear stage, performs elementary local motion extraction by cross-correlation (Reichardt, 1957; van Santen and Sperling, 1984), spatio-temporal energy analysis (Adelson and Bergen, 1985), or some other relatively simple motion-extraction algorithm. At subsequent stages, the extracted motion from the various different motion detectors in each neighborhood and from various locations is compiled, and a decision is reached that is appropriate to the response demands.

For motion stimuli, Chubb and Sperling arrive at a functional discrimination between first-order (Fourier, direct) and second-order (nonFourier, rectified) processes.[2] The first-order regime is referred to as a Fourier process because it is well modeled by the type of computation described above. That is, when the stimulus contains Fourier motion components in the range of spatio-temporal frequencies to which humans are sensitive, the amount and direction of these Fourier components directly predicts the amount and direction of perceived motion.

5.2.4. Second-order motion perception. The contrast-reversing bar of Fig. 8(c) is an example of a class of stimuli for which the direction of perceived motion is opposite

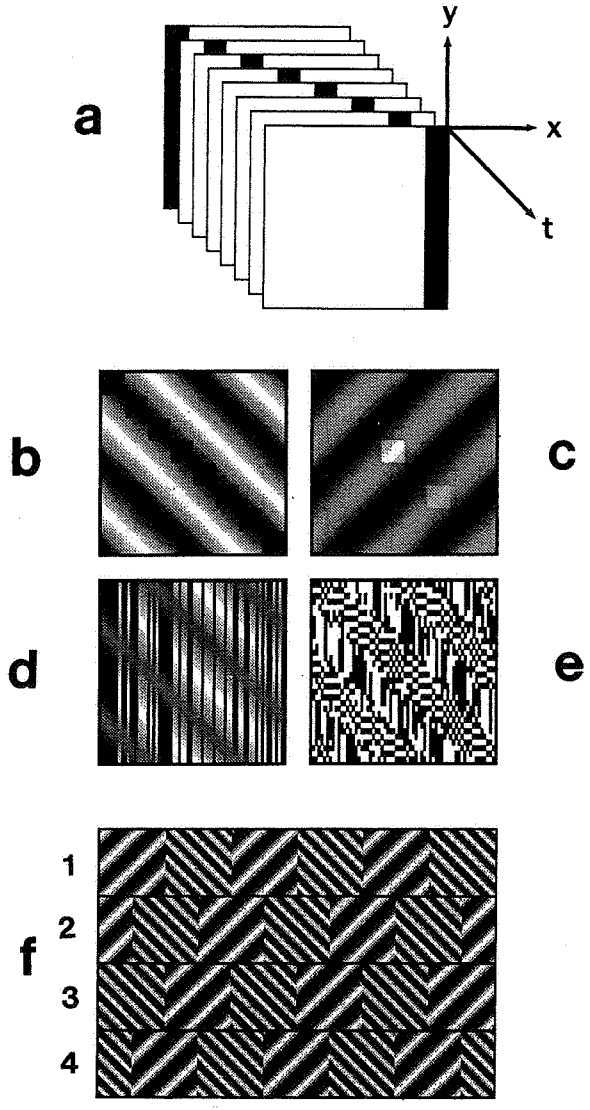

Figure 8. Stimuli for analyzing second-order processing. (a) An x, y, t representation of successive frames of a motion stimulus—a black bar moving rightward. (b) An x, t cross-section of (a). A sinewave grating, representing a dominant Fourier component, has been superimposed on the x, t cross-section. 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 near (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-Wiesel 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 orienation in x, y) to standard analysis.

to the direction of the Fourier motion components. Still other stimuli are perceived to move in the absence of any appropriate Fourier motion components. These 'non-Fourier' motion stimuli are detected by a second-order motion system. Second-order motion perception is now known to involve a stage of fullwave rectification (Fig. 7b) between the initial linear filtering and subsequent motion extraction. Second-order motion operates over larger retinal distances than does the first-order (Fourier) system. Additionally, consistent with the lower statistical 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 a fullwave and a halfwave second-order computation. Chubb and 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) did not find significant second-order motion. Thus, the predominant mechanism of second-order motion perception involves fullwave rectification. Fullwave rectification also is the dominant mechanism in second-order texture-slant processing of the x, y patterns that represented the x, t cross-sections of the motion stimuli in their motion experiments.[3]

In motion perception, there is a well-established distinction between short-range and long-range motion processes (e.g., Braddick, 1974; Pantle and Picciano, 1976; Westheimer and McKee, 1977; Victor and Conte, 1989b). The inadequacy of first-order motion processing has been amply documented by Ramachandran *et al.* (1973), Sperling (1976), Lelkens and Koenderink (1984), Pantle and Turano (1986), and Victor and Conte (1989a). The properties adduced for the long-range motion perception are generally those described for second-order motion above, plus a relative insensitivity to the eye of origin of successive stroboscopic stimuli. To these can be added the observations of Dosher *et al.* (1989) and Landy *et al.* (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). To supplement these theoretical proposals, van Santen & Sperling (1984, 1985) generated complex stimuli in which the direction and amount of perceived motion was opposite to intuition. Experimentally measurements of the perceived motion were quantatively predicted by their first-order, elaborated Reichardt model.

Chubb and Sperling (1988b, 1989a, b) provided a computational specification of a second-order motion system. They also provided methods for producing stimuli that can be proved mathematically to be directly aimed at one or the other system. One consequence is that it was easily shown that (retinal) short-range and long-range are inadequate system descriptions because there is a broad intermediate range in which both computations operate.

5.2.5. The equivalence of x, y spatial slant to x, t velocity. The velocity of a moving vertical line is the slope of its x, t cross-section (Fig. 8a, b). The slant of the line is the angle corresponding to the slope: slant $= \tan^{-1}$ (slope). For any stimulus, velocity in x, t and slant in x, y are equivalent up to a simple monotonic transformation. The problem of left-vs.-right direction of motion discrimination for x, t motion stimuli involves formally identical computations to the problem of left-vs.-right slant discrimination for x, y spatial texture stimuli (van Santen and Sperling, 1985; Chubb and Sperling, 1987, 1988b).

To extract local texture slant in scenes, Knutsson and Granlund (1983) proposed a computation in which similarly slanted odd and even linear filters (line and edge detectors) added their squared (rectified) outputs. The core of their algorithm anticipates Adelson and Bergen's (1985) remarkably similar motion algorithm, which in turn, was shown (by van Santen and Sperling, 1985) to be equivalent to other, previously proposed motion mechanisms, including the Reichardt correlation detector (e.g., van Santen and Sperling, 1984) and an elaborated Watson and Ahumada (1983) detector. Thus, in confronting first-order motion stimuli, the theories of Fourier-based motion perception and Fourier-based texture-slant perception have followed a parallel evolutionary path to the same pinnacle of success; when confronted with second-order motion and texture stimuli, the theories arrive at the same abyss.

5.2.6. Drift-balanced and microbalanced second-order motion/texture stimuli. Strong evidence for two computational regimes is obtained in studies of slant detection in textured x, y patterns as well as in studies of direction discrimination in one-dimensional x, t motion perception. Figures 8(d) and (e) show demonstrations of stimuli that show obvious apparent motion (when presented as x, t motion stimuli) and obvious slant (orientation) when presented in x, y, as in the illustration.

The stimuli of Fig. 8(d) and (e) are driftbalanced: that is, they are exemplars of random stimuli in which the *expected* motion (or slant) is exactly equal for every pair of oppositely-directed component Fourier frequencies (Chubb and Sperling, 1988b). The overlaid sine gratings of Fig. 8(b) and (c) are an example of two oppositely-directed Fourier components—their slants in the x, t cross-sections are equal and opposite. The stimuli of Fig. 8(d) and (e) are not only driftbalanced, they are also microbalanced. This means, roughly, that every little area, whatever its shape, in these stimuli is driftbalanced. This means that the obvious slant in these x, y stimuli would be 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. Any oriented x, y filter or oriented x, t motion detector would have the same expected response for any microbalanced stimulus (in particular, zero, for the stimuli of Fig. 8d and e) as for their mirror-images. Rectification is required to make the x, t motion or x, y slant in these stimuli accessible to standard slant or motion analysis (Chubb and Sperling, 1988).

5.2.7. Texture quilts prove linear filtering precedes rectification. Figure 8(f) shows an example of a texture quilt (Chubb and Sperling, 1989a). The essential ingredient of a texture quilt is a moving patch of texture. The trick is that the spatial phase of the texture in the moving patch is uncorrelated from frame-to-frame.

The temporal version of texture quilt illustrated in Fig. 8(f) exhibits consistent

apparent motion; and the spatial version exhibits consistent slant. Although the adjacent patches in the quilt of Fig. 8(f) differ both in spatial frequency and in slant, preliminary experiments indicate that apparent motion is perceived in texture quilts with patches differing only in spatial frequency or only in slant. These results mean that the early texture grabbers are spatial-frequency selective and that at least some are orientation specific.

Chubb and Sperling (1990) prove that to make the overall motion in such a texture quilt 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. In terms of a perceptual computation, this means that the texture with which each patch of the texture quilt is filled must first be extracted and rectified before the information can be used in a larger-scale slant-extraction or motion-extraction computation to reveal the overall pattern.

Since the work of Schade (1952) and DeLange (1954), first-order Fourier-based computations have been the cornerstone of psychophysical analysis.[4] The examples of Fig. 8(c, 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 the second-order processes for perceiving motion and texture-slant.

5.3. *Direct evidence for two computational regimes in distance judgments*

As with motion perception and texture-slant perception, distance estimation experiments also yield evidence for two perceptual processing systems, one Fourier and one rectifying system. In a three-bar distance estimation experiment, an observer must judge whether a central line is equally spaced between two flanking bars (bisection). In a two bar task, the observer directly judges the distance between two widely separated bars.

5.3.1. Direct estimates of distance. 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 both are black (Burbeck, 1987). Indeed, when 'bars' are defined by patches of high frequency gratings so that 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 by means of which the stimulus grating patch is represented internally as 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. Linear receptive fields are inherently phase sensitive with responses that vary from negative to zero to positive depending on just where in the receptive field the stimulus patches fall. Receptive fields would have to be duplicated for all orientations, distances, pairs of frequencies, and for at least four pairs of spatial phases (sin, cosine in each patch). 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 in judging distance between same-contrast and opposite-contrast patterns. Again, there 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.

5.3.2. Bisection. Klein and Levi (1985) used a central line to bisect a spatial interval defined by two horizontal flanking lines. The flanking lines were either directly above and below the central line or displaced sideways. Observers judged which interval was smaller. 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.

5.4. *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). Critical issues remain unresolved: How do first- and second-order computations combine to determine perception? To what extent are the noise sources associated with first- and second-order computations 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

This work was supported by USAF, Life Sciences Directorate, Visual Information Processing Program, Grants 85-0364 and 88-0140.

REFERENCES

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

Anstis, S. M. (1970). Phi movement as a subtraction process. *Vision Res.* **10**, 1411–1430.

Banks, M. S., Geisler, W. S. and Bennett, P. J. (1987). The physical limits of grating visibility. *Vision Res.* **27**, 1915–1924.

Barlow, H. B. (1956). Retinal noise and absolute threshold. *J. Opt. Soc. Am.* **46**, 634–639.

Barlow, H. B. (1957). Increment thresholds at low intensities considered as signal/noise discrimination. *J. Physiol.* **136**, 469–488.

Berliner, J. E. and Durlach, N. I. (1973). Intensity perception. IV. Resolution in roving-level discrimination. *J. Acoust. Soc. Am.* **61**, 1677–1685.

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 Res.* **27**, 417–427.

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

Carlson, C. R. and Klopfenstein, R. W. (1985). Spatial-frequency model for hyperacuity. *J. Opt. Soc. Am. A* **2**, 1747–1751.

Cavanagh, P. (1989). Motion: The long and the short of it. *Spatial Vision*, **4**, 103–129.

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

Chubb, C. and Sperling, G. (1988a). Processing stages in Non-Fourier Motion Perception, *Invest. Ophtalmol. Visual Sci.* **29**, 266.

Chubb, C. and Sperling, G. (1988b). Drift-balanced random stimuli: A general basis for studying non-Fourier motion perception. *J. Opt. Soc. Am. A* **5**, 1986–2006.

Chubb, C. and Sperling, G. (1989a). Second-order motion perception: Space–time separable mechanisms. *Proceedings: 1989 IEEE Workshop on Motion.* IEEE Computer Society Press, Washington, DC, pp. 126–138.

Chubb, C. and Sperling, G. (1989b). Two motion perception mechanisms revealed by distance driven reversal of apparent motion. *Proc. Natl. Acad. Sci.* USA **86**, 2985–2989.

Chubb, C. and Sperling, G. (1990). Texture quilts: Basic tools for studying motion from texture. *J. Math. Psychol.* **34**, in press.

DeLange, H. Dzn. (1954). Relationship between critical flicker-frequency and a set of low frequency characteristics of the eye. *J. Opt. Soc. Am.* **44**, 380–389.

Dosher, B. A., Landy, M. S. and Sperling, G. (1989). Kinetic depth effect and optic flow: 1. 3D shape from Fourier motion. *Vision Res.* **29**, in press.

Green, D. M. and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics.* Wiley, New York.

Geisler, W. S. (1989). Sequential ideal-observer analysis of visual discriminations. *Psychol. Rev.* **96**, in press.

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

Ives, H. E. (1922). A theory of intermittent vision. *J. Opt. Soc. Am.* **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 Res.* **22**, 407–416.

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

Jesteadt, W., Wier, C. C. and Green, D. M. (1977) Intensity discrimination as a function of frequency and sensation level. *J. Acoust. Soc. Am.* **61**, 169–177.

Klein, S. and Levi, D. M. (1985). Hyperacuity thresholds of 1 sec: Theoretical predictions and empirical validation. *J. Opt. Soc. Am. A* **2**, 1170–1190.

Knutsson, 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 AID.* IEEE Computer Society, Silver Spring, MD, pp. 206–213.

Kuffler, S. W. (1953). Discharge patterns and functional organization of mammalian retina. *J. Neurophysiol.* **16**, 37–68.

Landy, M. S., Sperling, G., Dosher, B. A. and Perkins, M. (1987). From what kind of motions can structure be inferred? *Invest. Ophthalmol. Visual Sci. (Suppl.)* **28**, 233.

Legge, G. E. and Foley, J. M. (1980). Contrast masking in human vision. *J. Opt. Soc. Am.* **70**, 1458–1471.

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

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

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

Marr, D. (1982). *Vision. A Computational Investigation into Human Representation and Processing of Visual Information.* W. H. Freeman and Company, San Franciso.

Nakayama, K. and Silverman, G. (1985). Detection and discrimination of sinusoidal grating displacements. *J. Opt. Soc. Am. A* **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. *Invest. Ophthalmol. Visual Sci.* **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*, Department of Psychology, New York University, Pp. 30.

Parish, D. H. and Sperling, G. (1987b). Object spatial frequency, not retinal spatial frequency, determines identification efficiency. *Invest. Ophthalmol. Visual Sci. (Suppl.)* **28**, 359.

Pavel, M., Sperling, G., Riedl, T. and Vanderbeek, A. (1987). The limits of visual communication: The effect of signal-to-noise ratio on the intelligibility of American Sign Language. *J. Opt. Soc. Am. 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 Res.* **13**, 1399–1401.

Reichardt, W. (1957). Autokorrelationsauswertung als Funktionsprinzip des Zentralnervensystems. *Z. Naturforsch.* **12b**, 447–457.

Riesz, R. R. (1928). Differential intensity sensitivity of the ear. *Phys. Rev.* **31**, 867–875.

Robinson, D. W. and Dadson, R. S. (1956) A redetermination of the equal-loudness relations for pure tones. *Br. J. Appl. Phys.* **7**, 166–181.

Schade, O. H. (1952). Optical and photelectric analog of the eye. *J. Opt. Soc. Am.* **46**, 721–739.

Scharf, B. and Buus, S. (1986). Audition 1. Stimulus, physiology, thresholds. In: *Handbook of Perception and Performance. Vol. 1.* K. Boff, L. Kaufman, and J. Thomas (Eds.) Wiley, New York, pp. 14–1 to 14–71.

Shapley, R. and Enroth-Cugell, C. (1986). Visual adaptation and retinal gain control. *Prog. Retinal Res.* **3**, 263–266.

Sperling, G. (1976). Movement perception in computer-driven visual displays. *Behavior, Res. Methods Instrum.* **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 Parish, D. H. (1985). Forest-in-the-trees illusions. *Invest. Ophthalmol. Visual Sci. (Suppl.)* **26**, 285.

Sperling, G. and Sondhi, M. M. (1968). Model for visual luminance discrimination and flicker detection. *J. Opt. Soc. Am.* **58**, 1133–1145.

van Nes, F. L. and Bouman, M. A. (1967). Spatial modulation transfer in the human eye. *J. Opt. Soc. Am.* **57**, 401–406.

van Santen, J. P. H. and Sperling, G. (1984). Temporal covariance model of human motion perception. *J. Opt. Soc. Am. A* **1**, 451–473.

van Santen, J. P. H. and Sperling, G. (1985). Elaborated Reichardt detectors. *J. Opt. Soc. Am. A* **2**, 300–321.

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

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

Wallach, H. and O'Connell, D. N. (1953). The kinetic depth effect. *J. Exp. Psychol.* **45**, 205–217.

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

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

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

NOTES

1. Informal observations. Variations in contrast and luminance were not reported in Parish and Sperling (1987a).
2. The nomenclature was suggested by P. Cavanagh (1990).
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 boundary 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, Adelson and Bergen's (1985) detector of directional motion energy is equivalent to the Reichardt motion model (van Santen and Sperling, 1984, 1985) and to Knutsson and 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 second-order motion or orientation.
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.