15

COMPUTATIONAL THEORIES OF VISUAL PERCEPTION

Robert Shapley, Terrence Caelli, Stephen Grossberg, Michael Morgan, and Ingo Rentschler

I. INTRODUCTION

The vast quantity of information about visual perception and its neurophysiological foundations is a problem both for experts and for newcomers to the field. The human mind needs some principles of organization, some theoretical generalizations, that compress the sheer weight of knowledge into more manageable concepts. The first part of this chapter, Section II–V, presents contemporary views on the problems with theories of spatial pattern perception, and some proposed formulations that clarify the theoretical issues. The second half of the chapter (Sections VI and VII) deals with models of form, color, and brightness perception that involve special-purpose neural networks. This chapter thus draws upon material in virtually all of the preceding chapters, except for those concerned with development and clinical applications. Ultimately, however, phenomena of development and abnormal vision may provide critical challenges for comprehensive models of human vision.

One of the interesting and unexpected points of agreement among many diverse theoreticians of vision is that a satisfactory account of visual perception demands consideration of the visual system as a *nonlinear system*. This conclusion follows from general arguments about information processing but also from specific functional characteristics of visual performance. For instance, the phenomena of masking of vernier acuity led Watt and Morgan, (1984) to propose a theory requiring a nonlinear transduction, *rectification*, of signals in the visual cortex before the positions of vernier lines were calculated. In a completely different context, Shapley and Gordon (1985) and Grossberg and colleagues (Grossberg and Mingolla, 1985a,b; Grossberg and Todorovic, 1988) observed that rectification or some similar nonlinear stage was required to explain contour perception. A third example of an appeal to nonlinear image analysis arises in the different accounts of the phenomenon of "pop out" or immediate recognition of the odd element in an array (see Chapter 11, this volume). Explanations of preattentive, rapid texture segmentation

Visual Perception: The Neurophysiological Foundations.

require either feature detection (Julesz, 1981; Treisman, 1982) or local energy computations (Caelli, 1985) that may be related to the rectifiers postulated by Morgan, Shapley, and Grossberg and their colleagues.

Another recurrent theme in vision theory is the idea of *parallel processing* by multiple, independent visual channels (see also Chapter 6, this volume). One aspect of parallel processing involves multiple spatial scales of analysis, or, equivalently, multiple spatial frequency channels. This idea is central not only to models of spatial vision, but also in other contexts where the independent, parallel computation of brightness and of bounding contour is an important theoretical novelty.

II. SPATIAL FREQUENCY CHANNELS AND THE FOURIER TRANSFORM THEORY OF VISION

Much of the theorizing about spatial vision since the 1960s has been influenced by the ideas of Campbell and Robson (1968) and colleagues who proposed that spatial patterns were analyzed into their spatial Fourier components and detected by specialized spatial frequency channels (see also DeValois and DeValois, 1988; Chapter 10, this volume). In order to place more recent theoretical ideas in context, we must briefly review the theories of spatial frequency channels and Fourier representations of visual patterns.

A. The Fourier Representation

Any real visual image is equivalent to a sum of sinusoidal grating patterns. This is the visual, spatial expression of Fourier's theorem which states that (almost) any waveform can be represented as a sum of sinusoidal waves. Thus, for any arbitrary one-dimensional luminance profile $L(x)$, the Fourier representation is

$$L(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} L(k)\, e^{ikx}\, dk \qquad (1)$$

The coefficients for each spatial frequency in the summation, $L(k)$, are the Fourier Transform of the original waveform $L(x)$. It is important to realize that in general $L(k)$ is a complex-valued function of spatial frequency k. That means $L(k)$ has a real and an imaginary part; it is usually represented as having an *amplitude* and a *phase*. Thus, in the literature, authors write about the amplitude spectrum and phase spectrum of spatial waveforms.

If there were neurons that were very finely tuned to spatial frequency, the visual system could extract the Fourier coefficients and represent spatial waveforms within the brain in terms of their amplitude spectrum. If, in addition, there were some way of calculating phase at each of the spatial frequencies represented in the amplitude spectrum, one could reconstruct each spatial waveform completely. The first condition, narrow spatial frequency tuning, has in fact been found in many neurons in the primary visual cortex of cats (Cooper & Robson, 1968; Maffei & Fiorentini, 1973; among many others) and of macaque monkeys (DeValois, Albrecht, & Thorell, 1982). Thus, there might well be a representation of the amplitude spectrum of visual images in the activity of the population of cells in primary visual cortex.

The representation of the phase spectrum is problematical. Spatial phase is encoded in the firing patterns of *simple* cortical cells in primary visual cortex, both in cats and monkeys (Movshon, Thompson & Tolhurst, 1978c; Pollen & Ronner, 1981; DeValois et al., 1982; Spitzer & Hochstein, 1985). However, in *complex* cells, spatial phase is encoded very poorly or not at all (Spitzer & Hochstein, 1985).

Explicit proposals that primary visual cortex (V1) computes the Fourier amplitude spectrum were made by Glezer, Ivanoff, and Tscherbach (1973) and Robson (1975). These ideas were implicit in the earlier research of Blakemore and Campbell (1969) and Graham and Nachmias (1971). They have formed the basis for many subsequent psychophysical investigations of spatial frequency channels, among them

the papers by Graham (1980), Robson and Graham (1981), Caelli and Hübner (1983), Watson, Barlow, and Robson (1983), and Field and Nachmias (1984). An important qualification of the Fourier representation approach was included in Robson's (1975) proposal, namely, that Fourier amplitude spectra were only calculated over relatively small patches of the visual field. This proposal was intended to take into account the known large variation of visual properties with position in the visual field due to, among other factors, *retinal inhomogeneity*.

Localized Fourier-like analysis has been proposed to account for spatial pattern discriminations and hyperacuity (Wilson & Gelb, 1984; Wilson, 1986). The localization is accomplished by having spatial filters stationed at every retinal point. The Fourier-like analysis is accomplished by making the localized spatial filters moderately narrow in spatial frequency bandwidth and by postulating that there are several spatial filters with diverse spatial scales (optimum spatial frequency) at each position. In order to fit real data, it is important that the responses of each filter are passed through a static nonlinearity (simulated psychometric function) that is accelerating at low contrasts and decelerating at high contrasts. Pooling of responses of similar filters across space is required in order to account for spatial frequency discrimination. Pooling is done by taking the pth root of the sum of the p powers of all responses to be pooled. Thus, in Wilson and Gelb's (1984) model for spatial frequency discrimination, given any two patterns P_1 and P_2 the difference in response of the ith spatial frequency-tuned filter is given by

$$F_i(x) = F_i(P_1) - F_i(P_2) \tag{2}$$

and the pooled activity in one of the mechanisms is

$$F_m = [\sum_{i=1}^{n} (F_i(x)P]^{1/p} \tag{3}$$

The maximal response across the population of mechanisms determines which mechanism performs the discrimination.

III. PATTERN ACUITY AND HYPERACUITY

A central issue in contemporary models of spatial vision concerns the theoretical explanation for *hyperacuity*, the capability of human observers to make fine spatial discriminations on a spatial scale much smaller than the distance between photoreceptors in the retina. One possibility is that hyperacuity reveals the spatial scale of representation of visual position in the brain. An alternative view is that the hyperacuity limit depends on sensitivity and size of receptive fields but does not tell us about the representation of space in the visual system.

The view that hyperacuity reveals the spatial scale of visual representations derives from the scheme for spatial representation put forward by Marr (1982). According to this view, the major task of the visual system is to form a symbolic description of the outside world. In order to form this description, the visual system has to rely initially on the retinal image, the two-dimensional projection of light upon the back of the eye. One of the key ideas of Marr's theory of vision is that the first stage in vision is the formation of a symbolic description of the image itself: the so-called Primal Sketch. The Primal Sketch is a two-dimensional spatial description of the important structures in the image. The aim of this early symbolic process is to identify *features* of the image that are likely to be of importance in later scene-based analysis: features such as edges, bars, blobs, texture boundaries, corners. The Primal Sketch can be thought of as the internal representation of a cartoon of the image. Indeed, Marr (1982) adduced the fact that we so readily perceive meaning in cartoons as evidence for the existence of a natural counterpart to the cartoonist's symbols in the visual system.

Some important aspects of the Primal Sketch are the following: (a) The Sketch is symbolic. This statement is a little vague, but one can take it to mean that out of the practically infinite manifold of possible assertions implicit in the image, a symbolic process makes only a small number explicit, using a defined set of primitives (features) to do so. This relates to the

next point: (b) The Sketch is inflexible. It is not adapted to symbolize any possible structures in the image. The sketch has available only a small set of primitives and is thus restricted a priori in what it can represent. This aspect of the Primal Sketch is challenged implicitly by evidence for plasticity of internal representations presented later in this chapter. (c) The Sketch is data-driven. It is not subject to top-down influences. This means that (linear or nonlinear) transducers transform the image to produce the outputs of early vision, the responses of filters or detectors for the primitives in the Sketch. When our cognitive processes fail to understand the logic of one of these early inflexible processes, we call the result an "illusion." An example of an illusion which probably originates at the Primal Sketch level is the Münsterberg illusion, shown in Fig. 1. The key to understanding the Münsterberg illusion is that bandpass, spatial-frequency filtering of the image produces locally tilted elements along the mortar lines (Grossberg & Mingolla, 1985a; Morgan & Moulden, 1986). These locally tilted components cause a global impression of tilt along the mortar lines, as in the Fraser effect, here illustrated in Fig. 2. (d) The Sketch represents spatially localized primitives. Spatial primitives in the Sketch are given a positional tag to locate them in the image. The tag is closely similar to Lotze's (1886) "local sign." It follows that the representational stages following the Primal Sketch must be able to interpret positional tags and use them to describe spatial relations.

To test the idea of a Primal Sketch we need to investigate the way in which human observers discriminate between and recognize simple forms and textures. Watt and Morgan (1983, 1984, 1985) have argued that pattern discriminations will be carried out with high acuity if they use the natural symbols of vision. Were this assertion true, one could discover the primitives of the Sketch (if such things exist; see Chapter 11, this volume) by comparing pattern dis-

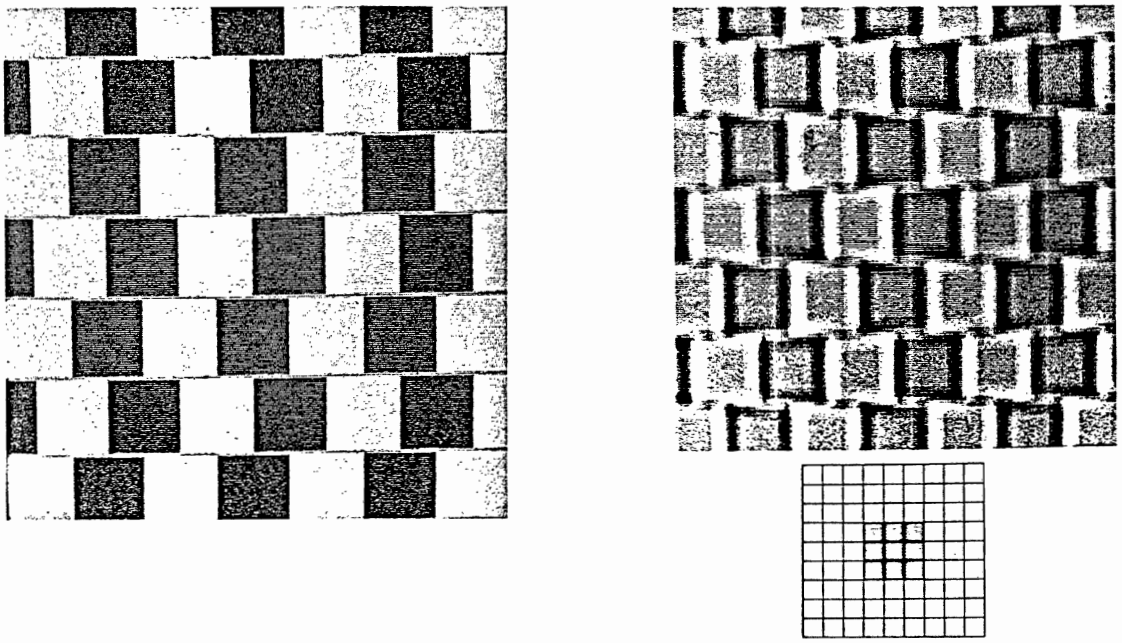


FIG. 1 At left is shown the Münsterberg or cafe wall illusion. The gray mortar lines are actually all horizontal (i.e., parallel), but they appear to be converging and diverging. At right is shown the results of band pass filtering the image with a Laplacian mask (inset). Note that filtering produces locally tilted elements along the previous mortar lines. From Morgan and Moulden (1986).

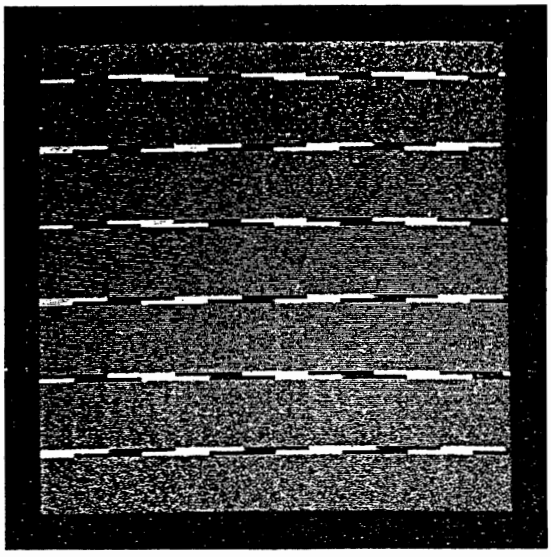


FIG. 2 A version of the Fraser twisted cord illusion, using phase-shifted rectangular elements instead of the usual tilted lines. If this stimulus is subjected to orientation filtering, it provides a maximal input to detectors tuned to orientations just off the horizontal. These local signals of tilt are apparently integrated to give an impression that the line is tilted as a whole, as in the Münsterberg illusion (Fig. 1).

crimination acuity for different spatial patterns. A useful class of pattern tasks for this purpose are "hyperacuities" (Westheimer, 1979a) such as vernier acuity or spatial interval acuity. For such hyperacuities, the minimum positional offset is an order of magnitude smaller than the distance between cone photoreceptor centers in the foveal cone mosaic. Hyperacuity tasks are useful for a number of reasons: thresholds can be measured by objective, forced-choice procedures; stimuli can be carefully controlled in contrast and spatial frequency content; and results are repeatable across observers both qualitatively and quantitatively.

It is possible that if we could understand how pattern acuity tasks are performed then we would be much closer to a viable theory of pattern vision in general. There is at present a sharp division between two theoretical approaches to the explanation of hyperacuity. One is the spatial primitives approach, exemplified in the literature by Watt and Morgan (1985). This theoretical position is the descendant of Marr's ideas about the Primal Sketch (Marr, 1982). The aim of this approach is to use hyperacuity data to constrain models of the early symbolic processes in vision, the local spatial primitives or features that are combined into the Primal Sketch. The emphasis in spatial primitive (feature) models is on the very restricted range of image variables which can be made explicit in the early stages of vision. In contrast to the spatial primitives approach are the models of Geisler (1984), Klein and Levi (1985), Parker and Hawken (1985), Wilson (1986), and Shapley and Victor (1986). Morgan mordantly refers to these theories as "Primal Soup" theories. He asserts that the original Primal Soup was an unstructured broth of amino acids and other useful things which evolved into life by aggregation and accretion. The proposed neural counterpart to this Primal Soup is the entire set of neurons in the primary visual cortex, each of which can be characterized by its spatial frequency and orientational selectivity. In models of pattern discrimination, such as that proposed by Wilson (1986), two patterns can be discriminated if they stimulate any subset of the Soup differently.

If two psychophysically discriminable patterns stimulate different populations of single units, it must be the case that at least one neuron exists which has a reliably different response to the two patterns. It ought to be possible to identify such units physiologically, and an impressive start in this direction has been made by Parker and Hawken (1985), who showed that single cells in monkey visual cortex should be able to respond to step changes in position of a bar within their receptive fields with a displacement hyperacuity threshold similar to that of a human observer. This prediction is based on the sensitivity of the most sensitive visual cortical cells to contrast reversal and on the small size of cortical receptive fields in the monkey (Parker & Hawken, 1985). If one asks how such a cell, or the class of neurons that it represents, should come to control the psychophysical discrimination, then one reaches the heart of the problem with Primal Soup theories.

No model of visual pattern discrimination, and certainly not the spatial primitives model of Watt and Morgan (1983, 1984), can deny that pattern discrimi-

nation thresholds are limited by differential neural activity. It is impossible that two patterns should be discriminated if they have identical effects on the brain, so if two patterns are discriminated, they must produce different neural responses. Theories such as Wilson's (1986) perform the essential function of trying to identify the subpopulation of units which limit performance. Using this approach one can answer such questions as, "Why is vernier acuity not better than 2 arc sec?" Geisler's (1984) ideal detector model addresses the same problem from the standpoint of photoreceptor mechanisms for differentiating stimuli. Shapley and Victor (1986) used a similar approach to demonstrate that positional hyperacuity was encoded in the impulse trains of retinal ganglion cells and was a natural consequence of linear filtering and a high signal-to-noise ratio. In fact, all of the Primal Soup theories point to the facts that pattern hyperacuities are limited by processing constraints in the visual cortex and that explicit signals for very small positional offsets in the hyperacuity range are present in the output of the retina. This is a complete turnaround from previous theoretical positions that postulated the need for central averaging and interpolation mechanisms to rescue miniscule hyperacuity signals lost in noise (Crick, Marr & Poggio, 1980; Barlow, 1981).

It can be argued that answering questions about the limits of hyperacuity falls far short of explaining how vernier or other hyperacuity judgments are made by the observer. The Primal Soup theories are not theories of explicit representation unlike, say, the Primal Sketch. The Primal Soup theories place no limits on what can and cannot be represented. Therefore they cannot be used to draw conclusions about many of the principal facts needed in an analysis of perception. Here we present these principal facts in the form of a series of assertions.

1. Observers Can Be Instructed How to Perform Pattern Discrimination Tasks

The point is simple and obvious but often neglected in theoretical analyses of pattern discrimination. Typical instructions might be: "I am going to show you two lines and I want you to tell me which one is longer by pressing the appropriate button on the response box." Subjects have no difficulty in understanding what these instructions mean, and they perform the task effortlessly. They do not need a long series of practice trials and they do not need feedback in order to achieve their best performance. Observers must therefore be basing their decisions on a visual representation of the pattern in which relations of length have been made explicit enough to be accessed through language.

How could this be done in the Primal Soup? The subject would have to discover that longer patterns stimulated one set of units and shorter another set. The problem of finding the right set of units would be hard enough if just two patterns had to be discriminated but it is even worse because:

2. Thresholds Can Be (and Usually Are) Determined Using a Range of Stimulus Values

The most commonly used methods of determining thresholds are derived from the Method of Constants. A range of stimuli is presented along some dimension and the subject must say whether each stimulus is "x" or "y." In the case of length discrimination, to continue the example given above, the question would be whether each stimulus is longer or shorter than the standard. If the observer were suddenly to be presented with a line longer than he or she had seen before, we can take it for granted that he or she would classify it unhesitatingly in a correct way. Thus, to explain pattern vision it is not enough to have a model that discriminates between patterns at threshold. The model must also explain how suprathreshold patterns are classified correctly and how novel stimuli can be recognized. In short, a complete model for pattern vision would have to explain how things look. This is exactly what Primal Soup theories fail to do. However, whether Primal Sketch theories explain pattern classification is unclear.

3. Pattern Acuity Judgments are Usually Directional

When observers discriminate between two lines of different length, they can say not only that they are

different but also which one looks longer. Similarly, vernier judgments are made on the basis of the direction of the spatial offset. Any satisfactory theory should account for directional ability. The Primal Soup theories are usually not explicit about this ability to identify the directionality of the discrimination and what it would cost in terms of sensitivity. This is a crucial criticism since it indicates the computational incompleteness of most Primal Soup theories of pattern discrimination. However, a similar criticism could be directed at Primal Sketch theories.

4. Discriminations Are Hard if They Are Not Based on Natural Continua

This is a proposition for which there is little evidence, but it might be asserted anyway in order to indicate what the relevant evidence might be, as a guide to future research. Pattern discrimination tasks use natural categories, such as "larger versus smaller," or "shifted left versus shifted right." But if any arbitrary collection of filters can be grouped together for purposes of discrimination, it should be possible to construct an infinite set of discriminations which do not correspond to natural categories. An example would be a texture composed of a set of random-phase spatial frequency components in a passband defined by its center frequency and bandwidth. Suppose one finds that with textures which have a bandwidth of 0.5 octaves, two such textures can be discriminated when the difference in center frequency is 0.1 octaves. Consider now a set of ten such textures spaced 0.1 octaves apart from one another; we require the observer to respond "+" to even numbered stimuli and "−" to odd numbered textures. Since all members of the set are discriminable by definition, they all reliably stimulate different mechanisms, so it should be possible according to Primal Soup theories to give the mechanisms appropriate weightings to solve the discrimination. As another example, vernier offsets of less than 20 sec arc in either direction could be called "sheep" and those absolutely greater "goats." It is interesting to compare this prediction with the experiments by Rentschler, Hübner and Caelli (1988) on classifying mirror-symmetric Gabor functions described below. They found that with sufficient pretraining subjects could reach 85% accuracy at discriminating such "unnatural" patterns. They conclude from these experiments that there are adaptive spatial filters in the visual system that can be modified to allow "unnatural" spatial discriminations.

5. Patterns Can Be Described as Well as Recognized

Perhaps one of Marr's most valuable contributions to the theory of image interpretation was his criticism of the traditional emphasis on pattern "recognition." What matters about seeing the telephone is not that one can say "phone" but that one can say things like "it consists of two separable parts connected by a cord; one part is roughly oblong with depressions in it into which the other parts fit." Similarly, observers can look at a vernier target for the first time and *describe* it. It is hard to believe that the processes which are involved in making this explicit description are not also involved in vernier discriminations. However, this conclusion must be labeled conjecture rather than inference since there is little evidence at present that ties the visual processes involved in description with those responsible for pattern discrimination.

6. Localized Edges are Important in Pattern Acuity

This sounds like a glimpse of the obvious, but consider whether it ought to be true. If the image of a face has the phases of all its Fourier components jumbled, it will appear to be a mess, but it remains uniquely specified over a set of filters. Thus, it should remain discriminable from other faces treated in the same way, according to Primal Soup type theories. Using the same line of argument, one can imagine a set of one hundred faces the observer can discriminate and even name reliably.

The appearance of images is certainly influenced by their amplitude spectrum, and, consequently, altering the amplitude spectrum by filtering changes their appearance (Ginsburg, 1984). Harvey (1986) used a series of band pass amplitude filtered images of a human face to find the filtered face that most

closely approximated the mental image of the face in memory. Using multidimensional scaling techniques, he concluded that visual memory behaves like a band pass filter, centered on 3.8 cycles per degree, with a bandwidth of 2.8 octaves.

The appearance of images is specified to a much greater degree by their phase than by their amplitude spectra. If an image is produced with the amplitude spectrum of an image A and the phase spectrum of a different image P, it will resemble image P rather than image A (Oppenheim & Lim, 1981; Piotrowski & Campbell, 1982; see also Fig. 3). The explanation is that the phase spectrum determines the spatial structure of a signal. There is evidence that the visual system's response to spatial structure is dependent on *local computations*: either local reponses to edges, bars, and blobs or local spectral energy in even and odd symmetric filters (e.g., Caelli & Moraglia, 1986). These may be two different verbal descriptions of the same computation. Whatever the nature of the local computation, it must depend on the positioning of Fourier components in the image and thus on their relative phase.

Primal Soup theories were designed to explain dis-

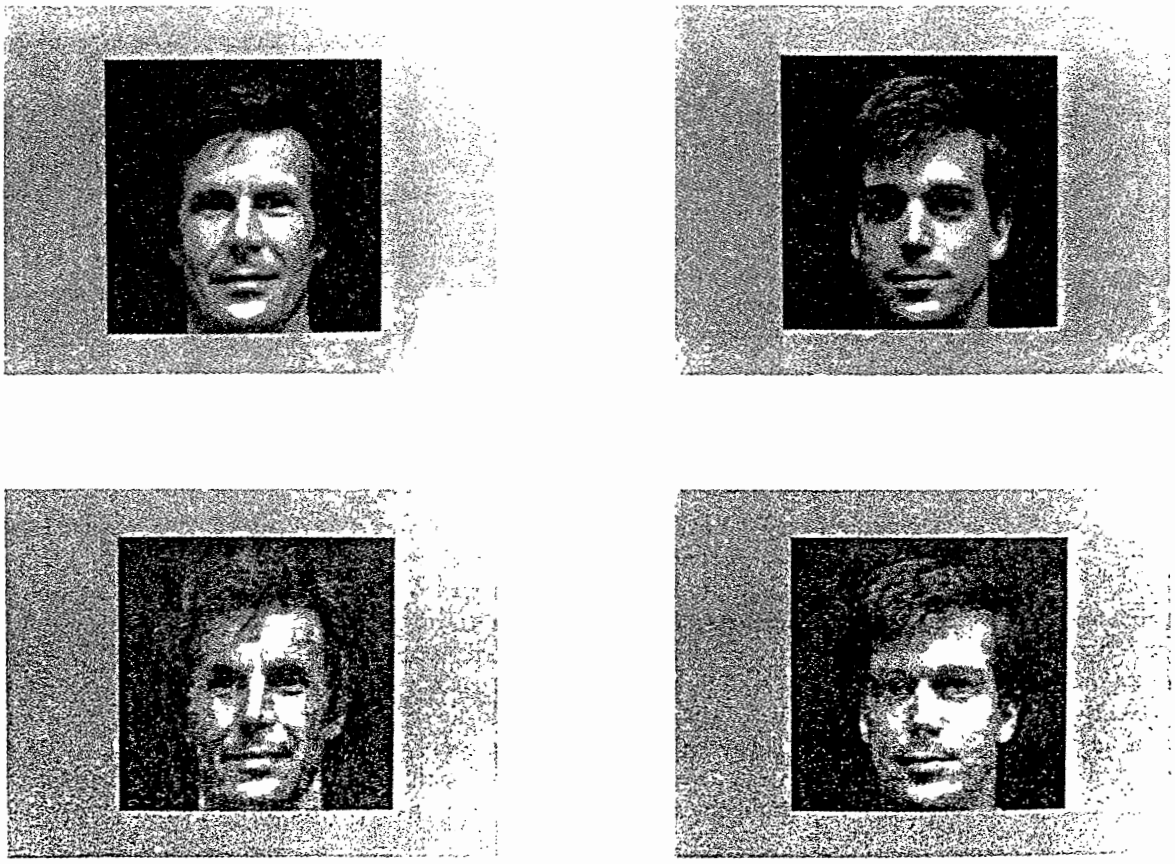


FIG. 3 Effect of scrambling amplitude spectra while keeping phase spectra unchanged between images. The top row of images comprises two different faces. In the second row, the images have had their amplitude spectra varied while keeping the phase spectra unchanged. A comparable distortion of phase spectra with no change in amplitude spectra produces unrecognizable faces.

crimination and hyperacuity, not appearance. These theories have demonstrated that discriminations and hyperacuity may be understood as threshold detection problems: threshold detection in the presence of a masking pattern, the discriminandum. They were not designed to handle appearance. The success of Primal Soup theories in accounting for hyperacuity performance reveals that the assumption that studying hyperacuity should lead to the mechanisms for producing appearance is not necessarily true.

IV. COMPUTATIONAL RULES FOR SPATIAL VISION

Modeling biological spatial vision in terms of selected signal types and system parameters is not entirely adequate. For example, though masking and adaptation studies show "tuning" to spatial frequency, orientation, and phase parameters of simple grating patches (e.g., Wilson, McFarlane & Phillips, 1983; Breitmeyer, 1984), these curves do not predict masking effects with more natural images. Repeatedly it has been found that the properties of *local image luminance profiles* are directly relevant in such tasks. Transform domain parameters (e.g., Fourier amplitudes and phases) are only predictive when they determine significant characteristics of the luminance profiles (Caelli & Moraglia, 1986). This is the reason that, in the development of models of spatial vision based on Fourier methods, there has been a progression from models in which spatial phase was not preserved to a model such as Wilson and Gelb's (1984) in which position information is retained.

What is required for an adequate paradigm for spatial vision is not details of psychophysical procedure, nor the enumeration of parameter tuning for postulated neural mechanisms, but, rather, *general computational constraints* that are necessary and sufficient to solve given perceptual tasks. This idea is reminiscent of Marr's (1982) parsing of computational vision into levels of description, the highest level of which is the computational theory. Work by Caelli (1988) and Rentschler et al. (1988) suggests that the type of detectors, their tuning characteristics, and use of detector responses by decision-making processes may be adaptive to the signal and task demands. Then the true *invariants* for visual information processing consist in the types of computations available.

One of the major tasks of the visual system, according to this computational paradigm for vision, is object and texture segregation. This means that the system must parse retinotopically registered, viewer-dependent projections of object-reflected light into parts which mean something to the observer. There may be a natural division of such segmentation mechanisms into two classes. These are often referred to as *boundary-extracting* and *texture-encoding* mechanisms in the literature on image processing. They register luminance information in terms of two components over a range of spatial scales. The boundary extraction processes are usually thought to be determined by threshold outputs of detectors with retention of contrast sign. On the other hand, the texture processing system is viewed as registering the energy of the outputs of filters (or detectors) at each position, producing a distribution of activity that does not encode the sign of the detector's reponse. Several authors have shown that energy response measures are more representative of human texture discrimination performance than are texton or feature detector theories (Laws, 1980; Ade, 1983; Caelli, 1985; Bergen & Adelson, 1988). We now compare and contrast these different approaches to texture segmentation.

The perception of textural segments involves the ability of observers to carry out a labeling process whereby each position is allocated to one of a number of textured regions. These regions are usually perceived as contiguous, so forming a complete covering of the image surface. The task of a complete model for texture segmentation is to produce an algorithm that reflects human performance on this task.

Theories of texture segmentation have varied, including statistical or global integrative computations (Julesz, 1981) to detector-based models based on notions of fixed, highly tuned filters or "textons" (Laws, 1980; Julesz, 1981; Treisman, 1982; Julesz & Bergen, 1983; Caelli, 1985; Bergen & Adelson, 1988) or adaptive filters derived from the second-

order statistics of texture stimuli (Ade, 1983). Most current theories involve the notion of image decomposition by filters having point-spread functions of different shape, typically having center–surround (isotropic band pass) or orientation-specific even and odd symmetric profiles. However, the human visual system apparently does not use the outputs of such detectors in some texture segmentation tasks. This can be seen in Fig. 4 where two points are illustrated: (1) local orientation detection is clearly dependent on local contrast sign while (2) if local orientation is detected, the visual system is amazingly insensitive, in segmenting texture, to the sign of contrast. Such observations and others have led Caelli and colleagues to conclude that, in the texture mode, the appropriate filter output is energy or some other even nonlinear function of contrast and merely measures the magnitude of detector activity rather than its sign. The actual point-spread functions of the filters or detectors that are used for texture segmentation are still unknown.

One of the more serious logical errors in this area is the simple enumeration of filters or detectors in terms of apparently important components of texture micropatterns. For example, it is known that the crossing-line and end-of-line detectors are simply not necessary to produce texture discrimination in textures generated by Julesz (1981) and Treisman (1982) to confirm this claim. Detailed analyses of these textures demonstrate segmentability by simple

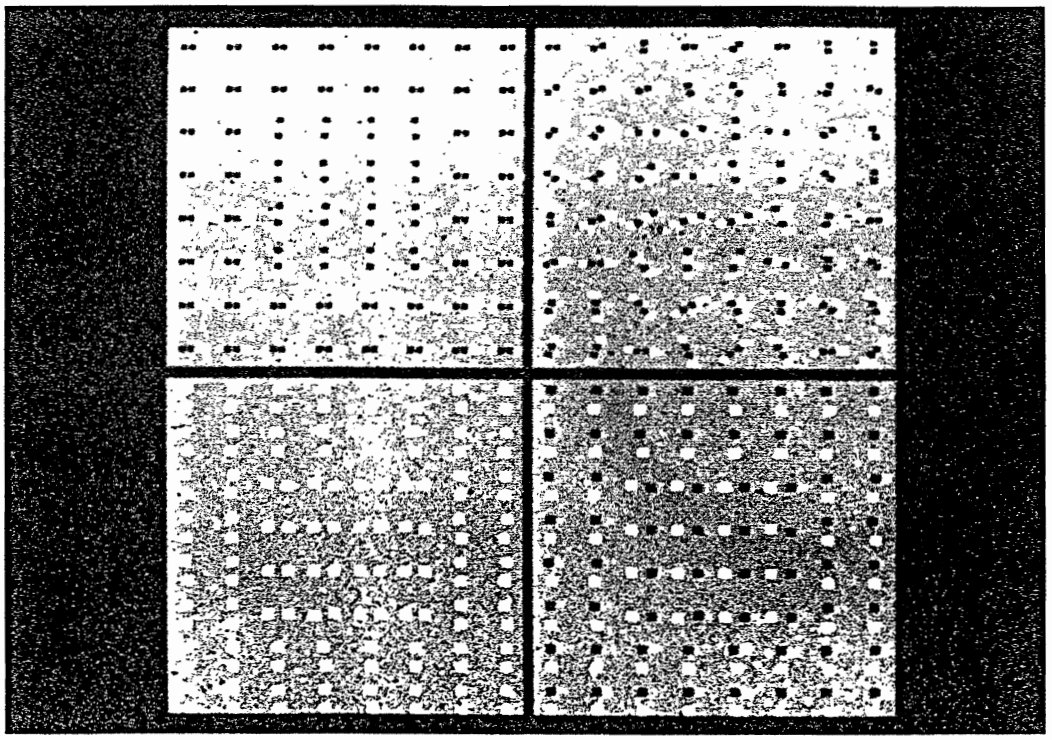


FIG. 4 The top row shows how orientation-specific texture segmentation is dependent on the local contrast of oriented dipoles. The bottom row shows that when sufficient orientation-specific micropattern differences are not present, contrast distribution differences are not sufficient to produce segmentation. From Caelli (1988), © 1988 IEEE.

orientation-specific, or even isotropic, band pass filters equivalent to retinal center–surround receptive fields (Caelli, 1985; Bergen & Adelson, 1988).

In order to understand the degree to which perceptual data tell us anything about the neural filters that allow humans to perform texture segregation, the filter design problem has to be considered in the context of the system's information processing characteristics. To do this we first note that texture segmentation involves the notion of region contiguity. That is, textures, or regions between texture boundaries, are perceived as regions where all positions are labeled as belonging to the same equivalence class. This constraint usually involves some form of *relaxation labeling* where the texture processing system aims at attaining contiguity by updating the labeling according to near neighbor labels—a form of cooperative processing.

Another consideration for filter design is the inherent redundancy in a large number of filters which, by virtue of spectral overlaps, have signal responses that are correlated. It is important to realize that the types of redundancies in detector outputs can be used to reshape current filter characteristics or to develop new detectors to capture the useful information in the texture statistics or luminance gradients.

To capture explicitly these principles of maximizing the spatial contiguity between texture boundaries and minimizing the number of functional detectors, the following associative-network equation has been developed by Caelli (1988):

$$A_i^{t+1}(x,y) = \frac{1}{\alpha\beta}\sum_{r}^{\alpha}\sum_{s}^{\beta} A_i^t\left(x + r - \frac{\alpha}{2}, y + s - \frac{\beta}{2}\right) + \sum_{\substack{j=1 \\ i\neq j}} w_{ij}^t A_j^t(x,y) \tag{4}$$

where $A_i^t(x,y)$ corresponds to the activity (energy) of detector i at position (x,y) and time t. The first determinant of the response is a simple linear version of relaxation where the response at position (x,y) to detector i is reinforced or inhibited as a function of the responses of the same detectors in its neighborhood. One neurophysiological basis for this process could be variable excitatory or activating links between cells in similar cortical areas having similar receptive field profiles where the strength of spreading activation varies directly with the cell's response. The second response determinant [the right hand term in Eq. (4)] contains the redundancy computation. Here the activity of detector i at position (x,y) is updated by the degree to which it is correlated over the textured regions with others at time t. This cooperative process has the effect of forming "attractors" in the detector space whereby detectors whose responses are correlated tend to become more correlated. This rather simple associative network demonstrates how many broadly tuned detectors can be adapted to signal characteristics so as to result in fewer, more appropriately tuned, and orthogonalized profiles.

Figure 5 shows examples of equilibrium ($t = 7$ iterations) results of this system using 24 Gabor functions as initial filter kernels. It is particularly important to note that the 24 detectors (only 12 are shown in Fig. 5) converged on one isotropic, low-pass profile that was sufficient to segment the center from surround regions, so producing contiguity of the texture regions. That is, what started off as a 24-channel model reduced to a single-channel model. In this model the decision process simply involves determining the degree to which detector outputs at each position differ, where the degree of difference indexes the degree of texture segmentability.

Figure 6 shows examples of four different textures segmented according to these principles. The results for column 2 correspond to zero associativity or independence of processing. Here, strength of segmentation closely corresponds to what is perceived both casually and experimentally (Julesz, 1981). Column 3 corresponds to classification strength as a result of filter update by the response correlations. Results here are compatible with those for the 24 channels. Thus, equivalent performance is achieved with a much more economical model. Column 4 illustrates what would happen if the filters had inhibitory connections during updating. It would not work. Inhibition in this context results in region differentiation, a result not consistent with the aim of extracting contiguous textured regions.

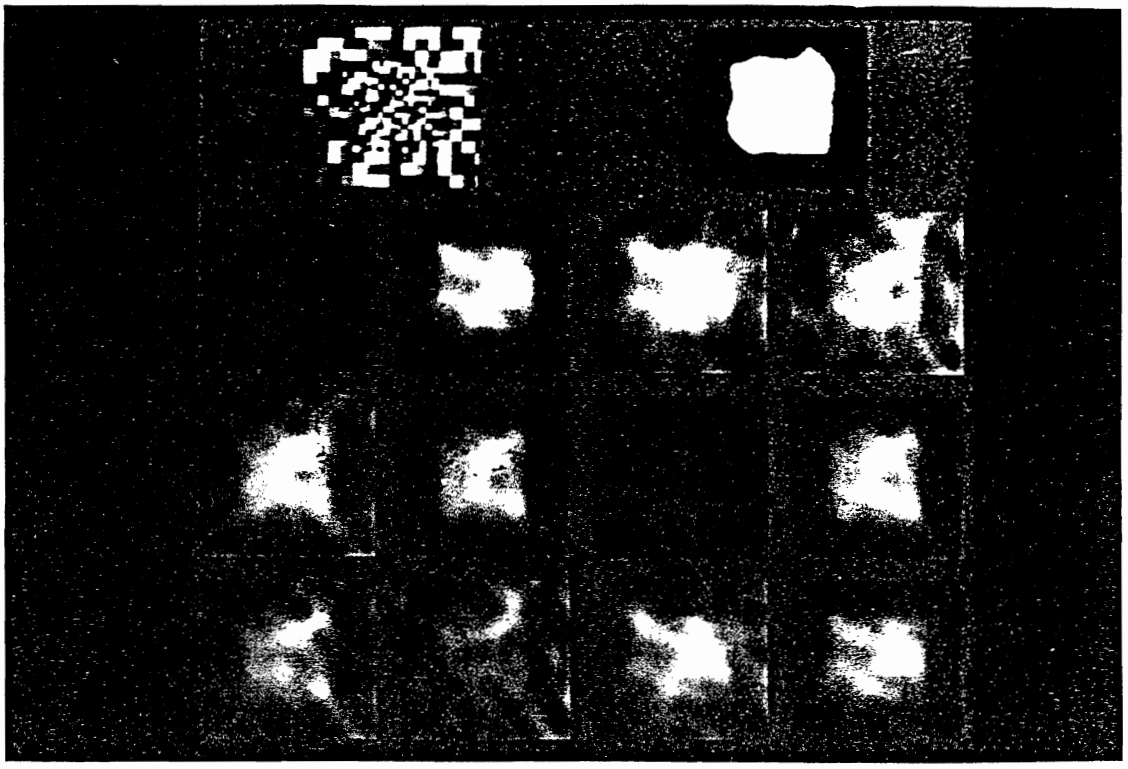


FIG. 5 Final activity profiles of 12 of 24 detectors used in a simulation of the associative processing algorithm in Eq. 4, in response to a texture varying in granularity. All of the detectors initially were two-dimensional Gabor functions but were modified as in Eq. 4 to become low-pass spatial filters. The test pattern is illustrated in the top left panel of the figure, and the texture segregation achieved by the adaptive filters is illustrated in the top right-hand panel.

Present attempts to understand vision in terms of filter theory usually result in hypotheses relating neurophysiology to psychophysics only if hypothetical filters are estimated to have similar values in the two types of tasks. This is not necessary since it is the algebraic form of the model and the proposed computations that are essential ingredients for the depiction of a model.

One final point should be made about the relationship between the computational mechanisms for the texture system and possible neurophysiological substrata. If we assume that the visual system has a "default" set of detectors tuned to, say, orientation and size, then what is claimed by Caelli et al., 1988 is that the type of information extracted from the response profiles either may modify the detectors or may cause new detectors to be produced. It may be that biophysical changes at the cellular level are involved in this modification process.

V. THE PATTERN RECOGNITION CONCEPT OF DIGITAL SIGNAL PROCESSING

The pattern recognition approach advocated by Rentschler et al. (1988) implies that objects are provided with a number of attributes, some of which constitute

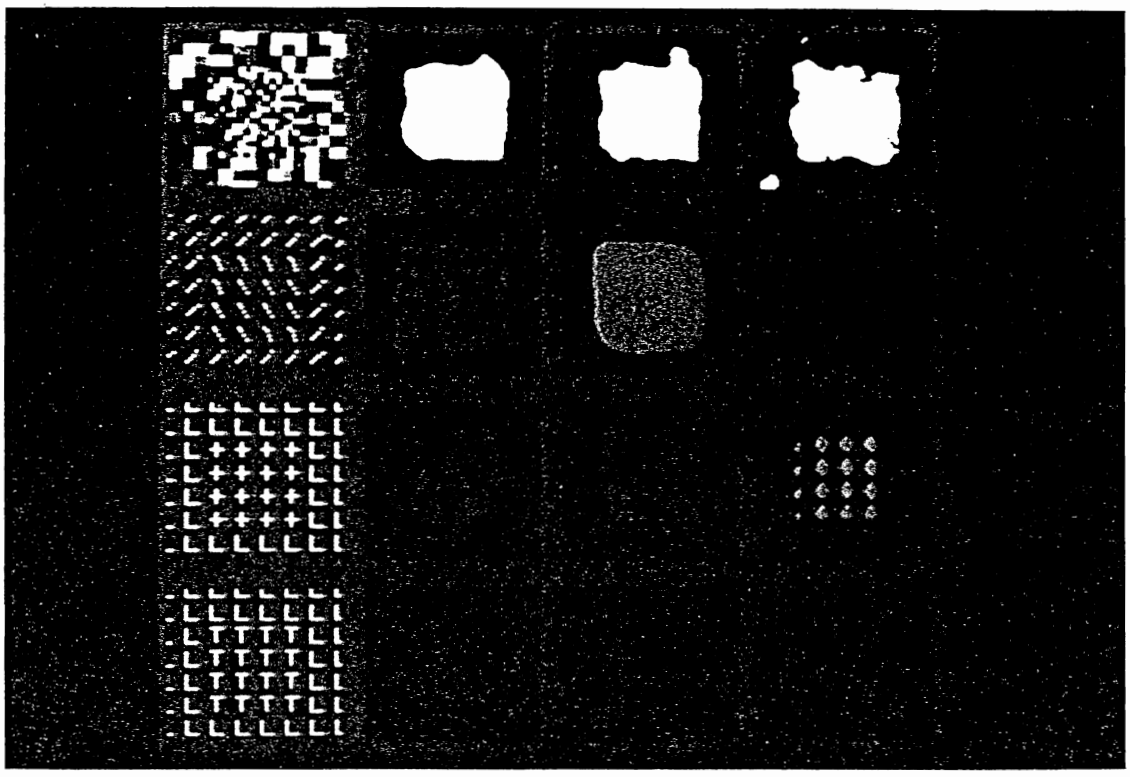


FIG. 6 Segmentation results for 4 textures (column 1) resulting from the outputs of 24 detectors with no associativities (column 2), with associativities (column 3), or with inhibitory mass action (column 4). For details see text.

the peculiarity of a specific object while others which may be called "features," determine the class to which the object belongs (Watanabe, 1985, Chapter 11, this volume). Pattern recognition begins with the observation of a number of variables that is usually quite large and ends with the determination of a single binary variable on which the membership of the given object to a certain class depends. It follows that *the reduction of variables* is essential for solving a problem of pattern recognition. The reduction of dimensionality is often done in several steps. One step may involve a transformation of variables with unimportant ones being rejected. This corresponds to the mapping of the original signal space onto a subspace of lower dimensionality (Watanabe, 1985). Classification, the last step of a pattern recognition procedure, consists of establishing a metric (a distance function) for that subspace and a decision rule for assigning the object to a certain class.

Let the input signal be a digital image composed of $N \times N$ picture elements (pixels). It can be considered as a vector in a vector space of $N \times N$ dimensionality, each pixel being a component of the signal vector. Usually this signal is subjected to an orthogonal transformation (e.g., the discrete Fourier Transform), but it is important to note that such a transformation does not alter the dimensionality of the signal. This implies that the application of a Fourier Transform per se cannot constitute a theory of visual pattern recognition. Feature selection, with the emphasis on dimensionality reduction, depends on an a priori decision, the usefulness of which can only be judged in terms of

classification performance. The resulting signal space of strongly reduced dimensionality is called the feature space.

Pattern classes consist of a set of n feature vectors, or samples, and it is usually assumed that the latter form a cluster in feature space. During "supervised learning," the classifier constructs "class prototypes" as the mean of the feature vectors within a class. Thus, if we write a cluster as:

$$c_i = \{Z_{i1}, Z_{i2}, \ldots Z_{in}\} \tag{5}$$

then the class prototype will be:

$$Z_i = n^{-1} \sum_{j=1}^{n} Z_{ij} \tag{6}$$

After the learning period, classification is made according to which prototype is closest in feature space to the input pattern. This criterion depends on the metric assigned to the feature space. Below we discuss an experiment by Caelli and colleagues in which a Euclidean metric, namely, the square root of the sum of the squares of component differences between pattern and prototype, was used successfully in conjunction with a maximum-likelihood estimate to account for the decision behavior of human observers. In this case the probability for assigning pattern i to class k is

$$p(k|i) = \frac{(d_{ik})^{-1}}{\sum_{j=1}^{m} (d_{ij})^{-1}} \tag{7}$$

with d_{ik} denoting Euclidean distance between pattern i and prototype k. The number of classes is m. A non-Euclidean metric is also described below.

The critical issue about pattern recognition is the conceptual framework within which feature selection and classification are being performed, and not the type of representation of the $N \times N$ dimensional input signals (as spatial or spectral). Moreover, the success of a given solution in a pattern recognition problem does not imply uniqueness of the solution. Watanabe (1981) proved that a necessary, unique solution cannot exist. It follows that theories and results of experiments related to human pattern recognition can provide insight into possible strategies of visual information processing but cannot provide sufficient information about the actual implementation in terms of neural machinery.

A. Classification of Compound Gabor Signals: Evidence for Adaptive Filtering

In a first experiment, Caelli, Rentschler, and Scheider (1987) studied the classification of compound Gabor signals as used also by Caelli, Hübner, and Rentschler (1986) and Rentschler, Hübner, and Caelli (1988) in studies on gray-scale textures. Such signals have luminance profiles defined by

$$g(x,y) = e^{-r^2/\alpha^2}[a \cos(2\pi fx) + b \cos(2\pi \cdot 3fx + \phi)] \tag{8}$$

where α determines the space constant of the isotropic Gaussian aperture with

$$r = (x^2 + y^2)^{1/2} \tag{9}$$

The amplitudes of the first and third harmonics in the Gabor function were a and b, respectively. The phase angle of the third harmonic was denoted ϕ. Results from two experiments are shown where observers learned the classification of learning sets of 15 samples into three categories. The signals were generated on a display with 128 × 128 pixels, 8 bits per pixel. Variations of the signals were restricted to amplitude and phase of the third harmonic. Examples of such images are illustrated in Fig. 7.

A procedure of supervised learning was applied consisting of a variable number of learning units. One unit consisted of three presentations in random order with a stimulus duration of 200 msec. Following each stimulus presentation a number was displayed specifying the pattern as a member of a certain class. At the end of such a unit, the subject was tested as to how he/she was capable of classifying the samples of the learning set. Only one exposure per pattern was used here. If the individual did not reach a criterion of 100% correct, he/she had to undergo another learn-

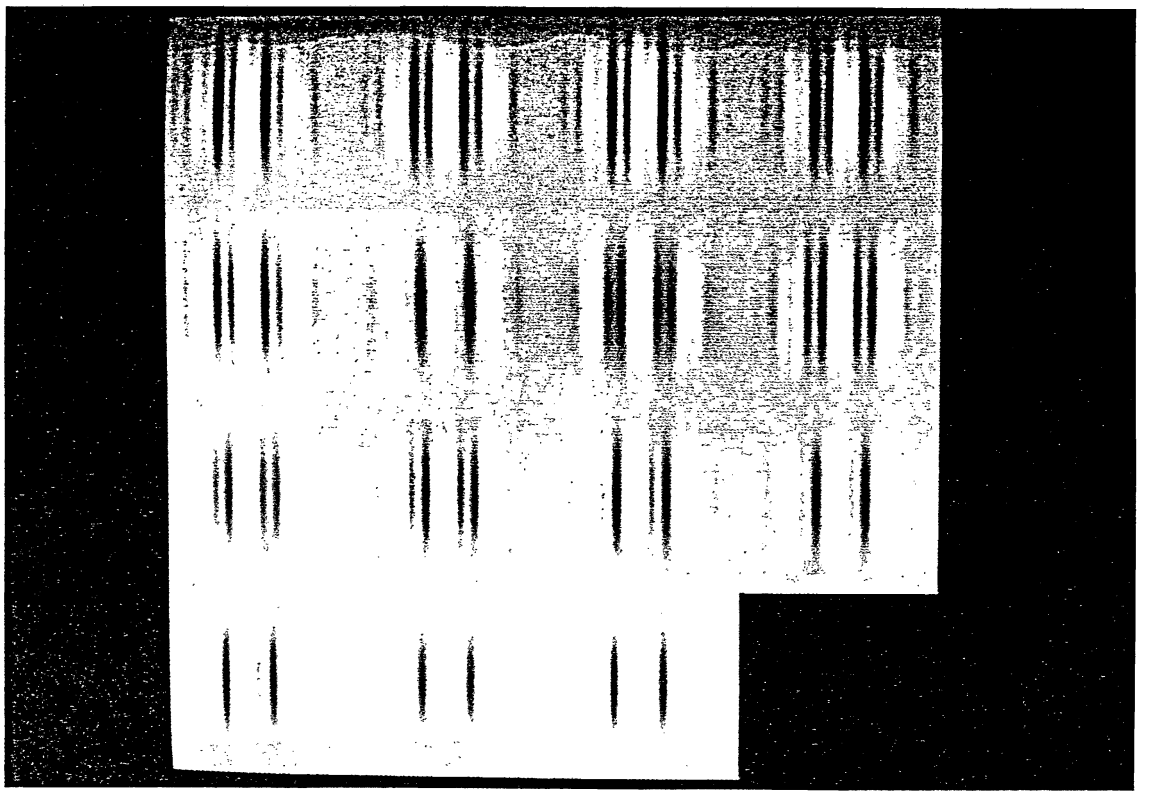


FIG. 7 Compound Gabor signals described by Eq. (8). Variations of the signals were restricted to the amplitude and phase values of the third harmonic.

ing unit, up to a maximum of 15 learning units. Cumulative confusion matrices over all stimuli and test runs were obtained for each subject. Figure 8 shows stimulus configuration 1 and the mean classification performance over four subjects.

The classification model used to account for these results was a least-squares minimum-distance classifier. It uses the Euclidean metric in feature space, but it is different from the regular minimum-distance classifier because it maps the feature space into a decision space, the dimensionality of which equals the number of classes used (for details see Ahmed & Rao, 1975). The feature space is the two-dimensional Cartesian coordinates of evenness and oddness of the third harmonic, that is, $x = b \cos \phi$ and $y = b \sin \phi$. The classification performance predicted by such a model is shown at the right of the Fig. 8.

The consistency between the observed and predicted classification data demonstrates the adequacy of feature selection and the least-squares minimum-distance classification model, involving a Euclidean metric in feature space, in predicting behavior. However, this solution may not be unique. There may be other ways of selecting features and classifying signals that lead to equally good predictions. What can be concluded is that the results cannot be explained on the basis of a fixed number of invariant "channels," or spatial filters. Rather, the view of class pro-

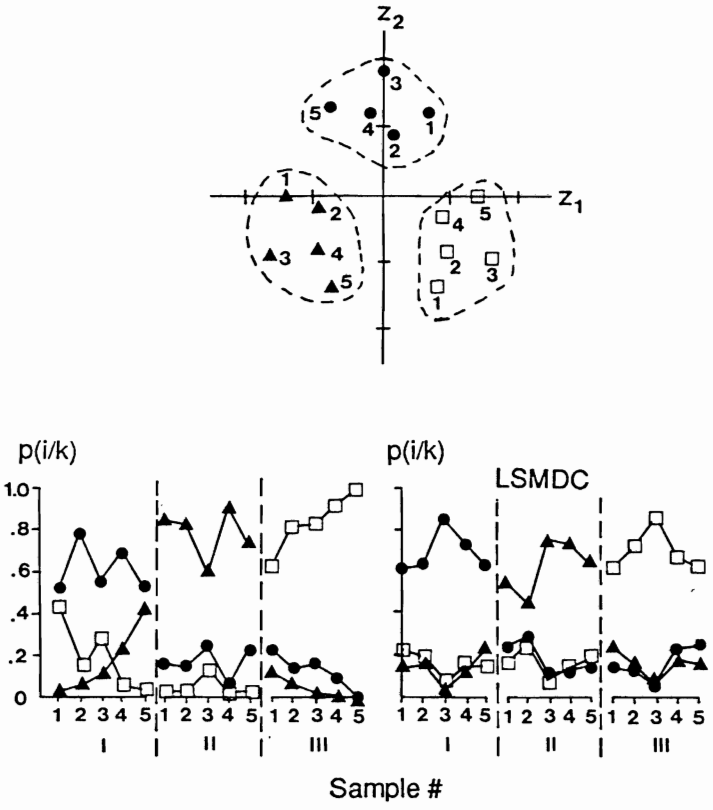


FIG. 8 Mean classification performance for four subjects on stimulus configuration 1 from the set of compound Gabor signals illustrated in Fig. 7 (upper left-hand panel).

totypes as exemplar images (see Watanabe, 1985, for a detailed discussion) which are matched with signals presented for classification is equivalent to defining them as adaptive filters.

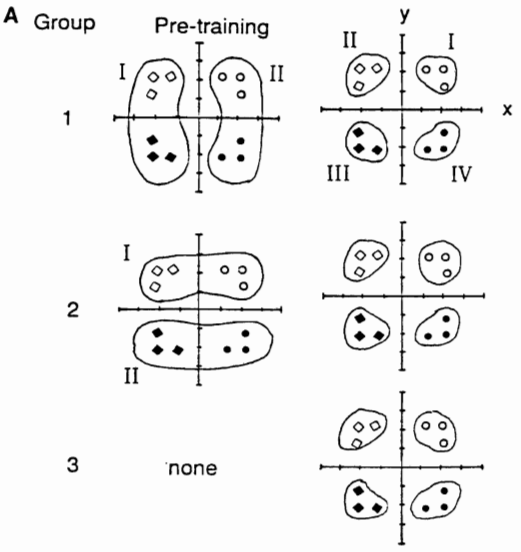
B. Classification of Mirror Image Compound Gabor Signals: Evidence for Effects of Cooperativity

A specific property of the two-dimensional feature space of evenness and oddness is that signal pairs having the fundamental waveform in cosine phase with (x_i, y_i) and $(x_i, -y_i)$ coordinates only differ with respect to the sign of their oddness component and therefore have mirror image luminance profiles. Previously Rentschler and Treutwein (1985) have shown that, independently of scale, mirror image compound grating pairs are indistinguishable in extrafoveal vision. The same is true for texture pairs composed of such signals as micropatterns (Rentschler et al., 1988). Bennet and Banks (1987) have confirmed these findings and interpreted their results in terms of the four channel model of Field and Nachmias (1984). Their main conclusion was that odd symmetric channels do not exist in peripheral vision. What is particularly relevant for the present chapter is that mirror image gratings (Field & Nachmias, 1984; Rentschler & Treutwein, 1985) and compound Gabor signals (I. Rentschler, unpublished observations) are more difficult to distinguish in foveal vision as well. To pursue this issue, Rentschler et al. used the pro-

cedures described above to study pattern classification of compound Gabor signals with mirror image relationships.

Twelve signals, clustered into four groups and symmetrically located in feature space, were used throughout the experiment. Four of the signals (one in each cluster) had the same high amplitude of third harmonic, while eight of them (two in each group) had the same low amplitude. What is important to note is that each signal in the upper half-plane of feature space had its mirror image correspondent in the lower half plane (Fig. 9A). The influence of pretraining on the classification of mirror image signals was studied by means of the following protocol. (1) The four subjects of group 1 were initially trained to 100% criterion in classifying the twelve patterns into two classes that contained three pairs of mirror image signals each. These subjects therefore were not required to distinguish mirror image patterns during pretraining. (2) The four subjects of group 2 were initially trained to criterion in classifying the twelve patterns into two classes with the mirror images being in different classes. Thus, these subjects learned to distinguish mirror image signals. (3) A third group of five subjects received no pretraining. (4) In the main experiment, each of the 13 subjects was then trained to criterion in classifying the same twelve patterns into four classes thus distinguishing between mirror image signals.

The results of the experiment are quantified as confusion matrices (Fig. 9B). A diagonal confusion matrix means that no pattern is confused with any other; classification is perfect. In the actual results, subjects from group 3 (no pretraining) had nondiagonal confusion matrices. Furthermore, the confusion matrix was often asymmetric. For example, patterns from class I were classified as belonging to class IV more often than patterns in class IV were classified as belonging to class I. Unlike those from group 3, subjects in group 1 showed smaller off-diagonal elements in their confusion matrices, and subjects in group 2, who had learned to discriminate mirror image patterns during pretraining, had almost perfectly diagonal confusion matrices. In summary, with increased



B

Group 1 (4Ss, N = 364)

.74	.06	.05	.15
.01	.78	.18	.03
.01	.31	.63	.05
.29	.02	.04	.65

Group 2 (4Ss, N = 102)

.85	.07	.01	.17
.06	.81	.12	.01
.01	.15	.82	.02
.13	.0	.05	.82

Group 3 (1S: N = 138)

.75	.07	.0	.18
.02	.81	.09	.08
.01	.42	.44	.13
.51	.06	.04	.39

FIG. 9 Testing procedures and classification performance on the mirror image Gabor signals. (A) The task of the main experiment was to assign the twelve Gabor signals to four classes. The three groups of observers received different pretraining: Group 1 had to assign the twelve signals to two classes with mirror images belonging to the same class. Group 2 had the same task with the mirror images belonging to different classes. Group 3 had no pretraining. (B) Cumulative confusion matrices for the three groups.

specificity of pretraining the confusion matrix tends to diagonal structure.

These results suggest that the idea of independent, fixed, and invariant channels, so popular in vision research, only applies to very specific situations where the brain is already highly adapted to that specific

task. It would be these specific states where the response to a certain class of signals, or objects, is adequately captured by determining the class prototype or (adaptive) filter. Visual psychophysicists may still wish to call such prototypes "channels." It is conceivable that these mechanisms are built up by synaptic modification as a consequence of learning. Yet, it may be that such channels have functional significance only in the brains of hard-working psychophysicists, or in the brains of their hard-working subjects. The inexperienced are more likely to use a "scatter brain" where the response to a signal consists of a distributed pattern of activity involving a whole ensemble of channels, prototypes, or neurons.

Summarizing these considerations, Rentschler et al. suggest that current models of spatial vision via fixed, invariant, and independent channels are inherently insufficient to explain visual pattern recognition. By definition, these concepts fail to capture active (i.e., signal dependent) processes of feature selection, and they are also inappropriate to examine the issue of cooperative interaction between filter mechanisms. On the other hand, standard procedures from computer pattern recognition can be adopted to investigate such problems in human vision.

VI. THEORIES OF THE NEURAL BASIS OF BRIGHTNESS AND FORM PERCEPTION

A. Relative Strength of Contrast and Assimilation in Brightness Perception

It seems that brightness perception is effortless, rapid, and simple; however, visual scientists have known for years that complex neuronal computations are required for satisfactory performance of this task (Jameson & Hurvich, 1959; Chapter 7, this volume). The visual system is not a photometer, and the perception of brightness is not simply a matter of counting photons. The primary determinant of brightness perception is local contrast, the local difference between luminances on either side of a boundary normalized by the (local) average luminance (Heinemann, 1955; Shapley & Enroth-Cugell, 1984). However, local contrast is not the only physical stimulus attribute that affects brightness perception. The spatial arrangement of nearby bright or dark objects, and the total amount of lightness and darkness in the local vicinity, may influence the apparent brightness of a test object in a way contrary to the effects of contrast. The psychobiological process that mediates this additive effect of local brightnesses has been called *assimilation* because it is similar to color assimilation (Hurvich, 1981). Assimilation is responsible for the filling-in of brightnesses from the borders of objects where contrast is computed. From comparison of psychophysical and neurophysiological experiments on brightness induction, Shapley and Enroth-Cugell (1984) and Reid and Shapley (1988) have concluded that local contrast is computed in the retina. Assimilation, however, is not seen in subcortical visual neurons and therefore must be a process that is introduced in the visual cortex, perhaps even at a higher level than Area V1.

Why is brightness perception so dependent on contrast, and why is there the additional process of assimilation that acts as a corrective to contrast? The key to understanding brightness perception is in comprehending the meaning of brightness constancy and how constancy results from response dependence on contrast.

Animals and men have evolved in a world of reflecting surfaces: water, earth, flowers, fur and skin and feathers. What characterizes a reflecting surface visually is its reflectance. The reflectance is determined by the physical properties of the surface of the object. The reflectance is more or less invariant with respect to illumination. The luminance of an object L_O is proportional to the product of the object's reflectance and illumination, thus

$$L_O = K \cdot R_O \cdot I \tag{10}$$

We know from Wallach (1976) and others that over a wide range of illumination the brightness of a reflecting object does not change, even though its luminance may vary widely, and that the brightnesses of an array of reflecting surfaces are perceived approximately in the order of their reflectances. Land and McCann

(1971) went so far as to say that the visual system was designed to calculate reflectance. We now know that this is not correct, but it is on the right track. The visual system calculates contrast and from this derives relative reflectance, as shown below. Stockham (1972) has pointed out that, under natural conditions, illumination [I in Eq. (10)] is a slowly varying function of time and space, while reflectance is a rapidly varying function because of the reflectance borders (edges) between objects and their backgrounds. Thus, light adaptation and spatial filtering would tend to remove the slowly varying factor in the luminance signal in Eq. (10) and would make the visual system more dependent on the biologically relevant reflectance function.

The early stages of vision compute contrast not reflectance. What this means is that the response of retinal, lateral geniculate and some primary cortical neurons is proportional to contrast over a low-to-medium contrast range, and then may saturate at high contrast. This is indicated by the results in Fig. 10 which show response proportionality to contrast over a hundred-fold range of luminances in cat retinal ganglion cells (R. Shapley & E. Kaplan, unpublished). This result was already adumbrated by the contrast sensitivity measurements of Enroth-Cugell and Robson (1966) who demonstrated invariance of the contrast sensitivity of cat retinal ganglion cells with background luminance over several log units of mean luminance.

An important feature of these results is that constancy of response with contrast is achieved for stimuli that activate only the center mechanism of the receptive field. This means that responsiveness of the visual system to contrast is not a result of center–surround interaction, or, in other words, of lateral inhibition (as in the standard textbook accounts, e.g., Cornsweet, 1970). Rather, the key to understanding dependence of response on contrast is to realize that contrast dependence is primarily a result of the automatic gain control that produces *light adaptation* (Whittle & Challands, 1969; Shapley & Enroth-Cugell, 1984; Chapter 5, this volume). The automatic gain control that regulates the contrast sensitivity of a receptive field center is localized to the center, and thus contrast is computed only locally (see Shapley and Enroth-Cugell, 1984, for a comprehensive review). An elementary calculation demonstrates how light adaptation yields dependence of neural reponse on contrast and therefore discounts the illuminant and produces brightness constancy.

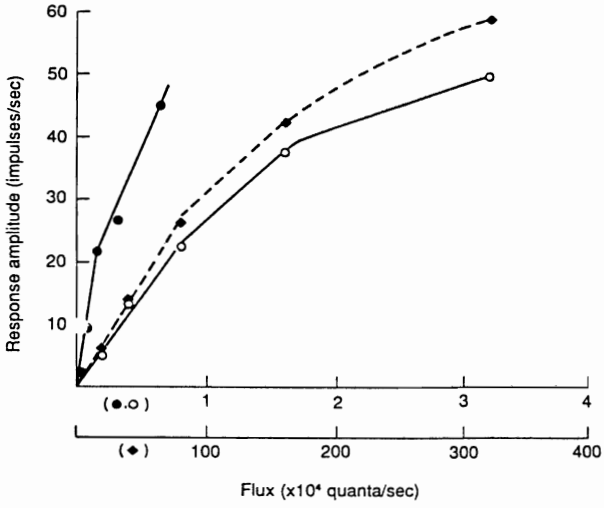


FIG. 10 Responses of an X-type cat retinal ganglion cell to drifting sine-wave gratings matched to the receptive field center, at three mean illuminances. Filled circles are responses to the gratings at the lowest mean illumination. Empty circles are responses to gratings around a mean illumination 10 times higher than the lowest. Diamonds represent responses to gratings at a mean illumination 100 times brighter than the second mean, and 1000 times brighter than the lowest. The diamonds are plotted on the lower axis of the abscissas, while the two sets of circles are plotted against the upper axis. Plotting the data in this way shows that, for empty circles and diamonds, responses are simply proportional to contrast (stimulus flux divided by mean or background flux).

Consider as an example a uniformly illuminated scene with an object on a background. The light from the object side of the border is $I \cdot R_O$ while the light from the background is $I \cdot R_B$ where I is the illumination. Contrast in this case can be defined as $(IR_O - IR_B)/IR_B$ which reduces to $(R_O - R_B)/R_B$. It is evident that contrast is independent of illumination. Now consider the response of a retinal neuron to the same object on the background. Retinal neurons respond to change as when the eye moves the receptive field across the border between background and object.

The effective stimulus in that case will be the difference between the light coming from the object and the background, that is, $IR_O - IR_B$. The response of the neuron will be the stimulus input multiplied by the gain of the input/output transduction. The gain, determined by the light adaptation mechanism, will be approximately proportional to $1/IR_B$. Thus, the response of the neuron will be proportional to $(IR_O - IR_B)/IR_B$, the contrast. And the neural response will also be proportional to $(R_O - R_B)/R_B$ when the illumination is factored out of the fraction. Thus, neuronal response will be proportional to contrast and independent of illumination. Objects of a given reflectance will look equally bright against the same background because they have the same contrasts. Notice that all these statements hinge on the claim that gain is reciprocal with illumination, that is, Weber's law. This law only holds in a limited range of mean illuminations (Shapley & Enroth-Cugell, 1984). When neuronal responses do not follow Weber's law, they also no longer are dependent strictly on contrast and invariant with illumination. This is particularly true at low luminances, as can be seen in Fig. 10. A strong prediction is that brightness constancy will fail when Weber's law does, and this prediction has been confirmed (Heinemann, 1955).

B. Contrast vs. Reflectance: Simultaneous Contrast

Shapley and Reid (1985) showed that it is contrast and not reflectance that the visual system is computing by confirming and extending classical studies on "simultaneous contrast." An example is the elaboration of the classic picture of equally luminant circles on a nonuniform background, as in Fig. 11. Figure 11 shows twelve equally luminant circles placed on rectangular stripes that vary in luminance. This could be interpreted as a scene with twelve equally reflective disks placed on a cloth with twelve different rectangular regions of reflectance. If the visual system is computing local contrast, then the disks should all look different in brightness because their contrasts are not all the same. If the visual system were computing

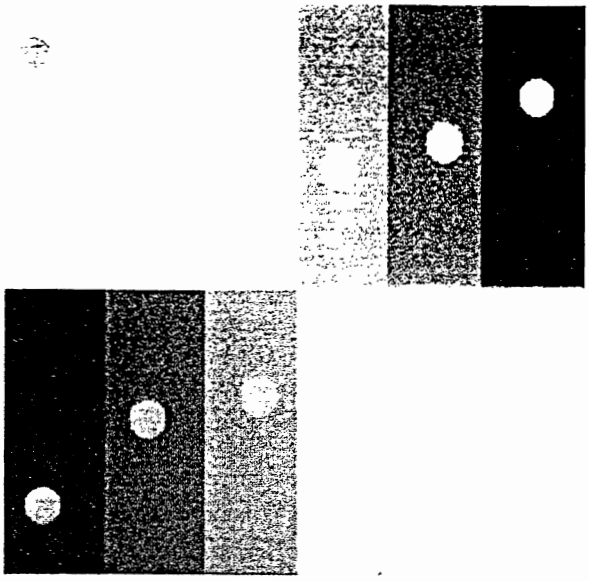


FIG. 11 Twelve equiluminant circles on a nonuniform background look different in brightness. From Shapley (1986).

reflectance as Land and McCann (1971) suggested, then all the twelve disks should look the same shade of gray. It is obvious that they look different, and so at least qualitatively the quantity computed is closer to contrast than to reflectance. Land and McCann (1971) anticipated this refutation of their theory (the 1971 Retinex) in a footnote to their paper in which they asserted that pictures such as Fig. 11 were unnatural because the objects were completely surrounded by their backgrounds. This argument has little merit since in nature isolated objects on backgrounds are the rule rather than the exception. But, accepting the Land and McCann argument as valid for the moment, we can put it to the test by constructing a "Mondrian"-like pattern that Land and McCann consider more natural and by looking again at equally reflective objects on nonuniform backgrounds. This is shown in Fig. 12. It can be seen that the circle and square though equally luminant are not equally bright. An explanation is that neurons are computing local contrast and the visual system is basing its estimate of brightness on contrast-dependent neural re-

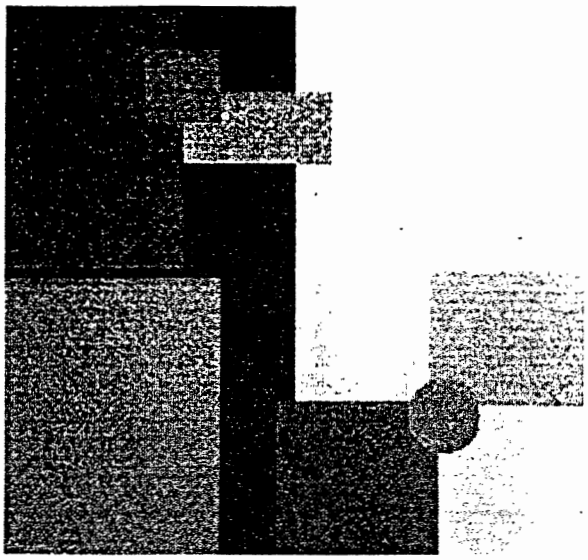


FIG. 12 The square (upper left) and the circle (lower right) are equiluminant but look different in this "Mondrian-like" pattern because the average local contrasts around their borders are quite different. From Shapley (1986).

sponse. Reid and Shapley (1988) have measured the magnitude of this brightness induction in Mondrian-like patterns, and it is virtually identical to that seen in more conventional displays such as Fig. 11.

C. Contrast and Assimilation

Besides local contrast, another component of the brightness computation is assimilation, the additive effect on brightness of an object produced by the brightness of its background. Previously, it has been hard to separate assimilation from contrast because backgrounds have been made brighter and darker by varying their luminances, and this has affected border contrast as well as background brightness. Shapley and Reid (1985) and Reid and Shapley (1988) have used a different approach by varying the brightness of an outer surround so as to vary the brightness of an inner surround by induction while its luminance remained fixed, as in Fig. 13. The inner circles in Fig. 13 are equally luminous, and they lie on annular backgrounds that are equally luminous (though slightly less luminous than the small circles). The outer background is a gradient of luminance such that one annular background appears brighter than the other. As a consequence, the circle on the brighter background appears brighter only by assimilation since its luminance and its *local contrast* are identical to that of the darker circle.

The magnitude and spatial dependence of assimilation were measured by Reid and Shapley (1988) who showed that assimilation was usually weaker than contrast and that it waned with distance. This can be expressed formally as follows. Consider two spots labeled T for test and C for comparison, surrounded by two regions S_T and S_C, respectively, and these in turn surrounded by an outer background B as in Fig. 13. The contrast between T and S_T is denoted C_T and between target C and its background as C_C. Reid and Shapley propose that to compute the brightness of

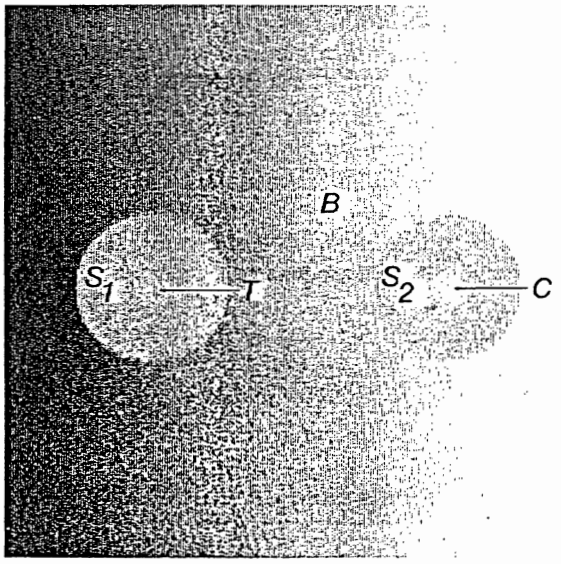


FIG. 13 Two equiluminant circles T and C placed on two equiluminant annuli S_1 and S_2 which look different in brightness because of the luminance gradient in the outer surround B. The inner circles T and C appear different in brightness because of assimilation. From Shapley and Reid (1985).

spots T and C, the visual system does the following: contrasts of the outer surround with the two annuli, C_{ST} and C_{SC} respectively, are multiplied by an assimilation weighting factor, which depends on distance between the borders of the spot and the annulus, and then added to the local contrast around the circle. Thus,

$$B_T = C_T + \alpha C_{ST} \qquad (11)$$
$$B_C = C_C + \alpha C_{SC}$$

The assimilation weighting factor, α, is unity at zero distance and declines to half at a distance of approximately 0.33 degree visual angle (Reid & Shapley, 1988). It is interesting that the Retinex theory (Land & McCann, 1971; Land, 1983) implicitly included an assimilation term that was unity in value, independent of distance (see Reid & Shapley, 1988, Appendix). In such a theory, brightness is a truly global computation since contrast at any border anywhere adds in with equal weight to all others. It is this global nature of the Retinex that endows it with the capability to calculate reflectance, and causes it to compute that objects of equal reflectance are equally bright, regardless of nearby surroundings. Visual neurons and human observers perform a much more *local computation* in order to calculate brightness. Assimilation can be seen as a weak corrective to the visual system's sole reliance on local contrast, but it is not strong enough to overcome the "illusions" in Figs. 11 and 12. The neural implementation of assimilation may be by lateral interactions between cortical neurons as suggested by Reid and Shapley (1988) and included in a formal neural network treatment by Grossberg and Todorovic (1988; see below).

D. Nonlinearity in Boundary Detection

One of the most surprising recent developments in understanding visual form perception is the finding that fundamental neural mechanisms of pattern segregation are essentially nonlinear because these mechanisms ignore the sign of the contrast. The discussion above demonstrates that the signed contrast is crucial for perception of brightness. The sign of the contrast is what separates dark (negative contrast) from light (positive contrast). If, as we now find, form perception depends on a mechanism that ignores sign of contrast, then form and brightness must be computed by separate neural circuits with different rules. This realization has emerged from recent work on texture segmentation, described above, and also from experiments on interpolated or "subjective contours" (Shapley & Gordon, 1985, 1987). The theoretical development of the two-mechanism idea is discussed in detail in Section VII. At this point, we discuss the experimental evidence from interpolated contours.

The best way to understand the evidence from interpolated contours is to see it (Fig. 14). In Fig. 14 there is a circular object on a shaded background. In fact, the luminance in the background is a linear gradient with position, and the mean luminance of the gradient is precisely in the middle of the picture. The luminance within the circular object is also a linear gradient, but increasing in the opposite direction from that in the background. Again, the average luminance of the circle's gradient is reached in the middle of the picture. Along the line down the middle of the picture, luminance is constant. There is no objective border down the center line. In fact, the luminance contrast between object and background are below threshold for several lines on either side of the middle of the picture, as can be verified by the reader by occluding the right- and left-hand sides of the picture leaving a blank zone down the middle. In spite of the absence of the segregating border when viewed in isolation, when the entire picture is viewed the border is completed or interpolated between regions in which it is visible and across regions where it is invisible on its own. This is an example of a subjective contour (Petry & Meyer, 1987). What is most fascinating about Fig. 14 and similar pictures is that the interpolated border links regions of *opposite* sign of contrast.

Various lines of evidence indicate that interpolated borders are caused by neural mechanisms of border detection that are the first stage in form perception (see Chapter 11, this volume). Shapley and Gordon (1987) have offered perceptual evidence on the contrast and spatial dependence of border interpola-

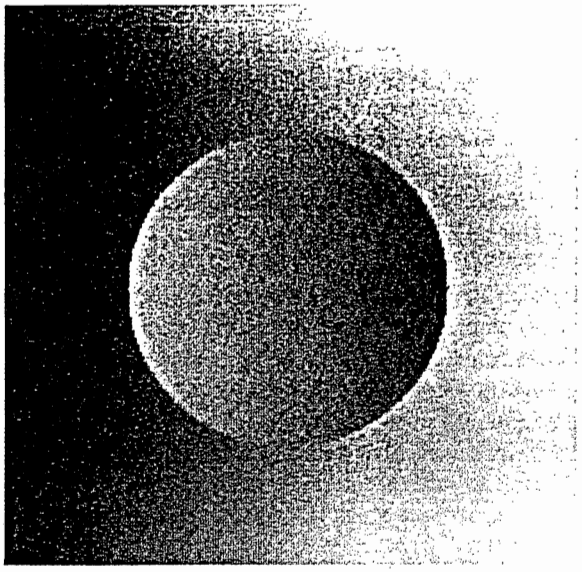


FIG. 14 A circular object that illustrates nonlinear border interpolation. The background region consists of a one-dimensional luminance gradient that has its mean in the middle of the picture. The interior of the circular region also is a luminance gradient with the opposite sign of that of the background, also going through its mean luminance (the same mean as the background) in the middle of the picture. Thus, there is no luminance border down the middle of the picture, yet one sees an interpolated "subjective" contour between regions of opposite sign of contrast.

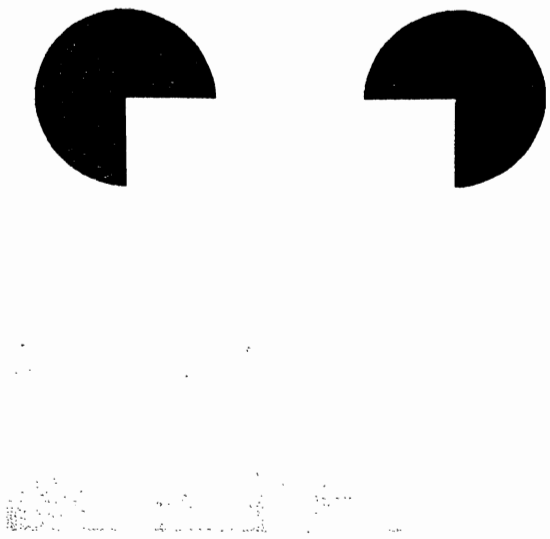


FIG. 15 A Kanizsa square, formed by inducing figures that alternate in sign of contrast. The subjective contours perceived are as strong as those seen with inducing figures of the same sign of contrast, again revealing that the contour sensing process does not respect sign of contrast. From Cohen and Grossberg (1984) and Shapley and Gordon (1985).

tion that implies a hard-wired neural mechanism as opposed to the cognitive mechanisms proposed by Prazdny (1983) and others. The crucial feature of such a border detector is its independence of the sign of contrast, implying an even-order nonlinearity such as rectification or squaring (energy calculation) before signals are combined. Thus, form depends on the magnitude of contrast at a border, while brightness depends on the sign. Another demonstration of the importance only of contrast magnitude and independence of interpolated borders on sign of contrast is offered in Fig. 15, which is a Kanizsa square with inducing figures of alternating sign of contrast (Shapley & Gordon, 1985; see also Cohen & Grossberg, 1984, for a similar picture). The illusory percept is as good as in Kanizsa figures of the same sign of contrast.

VII. NEURAL NETWORK MODELS OF PREATTENTIVE VISUAL PERCEPTION: EMERGENT SEGMENTATION AND FEATURAL FILLING-IN

Grossberg (1984) has developed a neural model for preattentive visual representation of three-dimensional form. He observed that the visual system uses multiple types of locally ambiguous information about edges, textures, shading, multiple spatial scales, and stereopsis to achieve this representation. The representation allows hyperacute spatial resolution and percepts of brightness and color that discount variation in illumination. Understanding how this happens is one of the outstanding problems in vision science.

Two types of visual process play a fundamental role: emergent boundary segmentation and featural filling-in. Grossberg's neural network theory attempts to explain how these processes work (Cohen & Grossberg, 1984; Grossberg, 1984, 1987a, 1987b, 1988; Grossberg & Mingolla, 1985a, 1985b, 1987; Grossberg & Todorovic, 1988). This theory describes neural processing rules for a boundary contour system (BCS), which generates preattentive three-dimensional emergent segmentations, and a feature contour system (FCS), which discounts the illuminant and regulates featural filling-in of brightness and color signals (Fig. 16).

The neural network model for the BCS and FCSs is illustrated in Figure 16. This is the simplified version

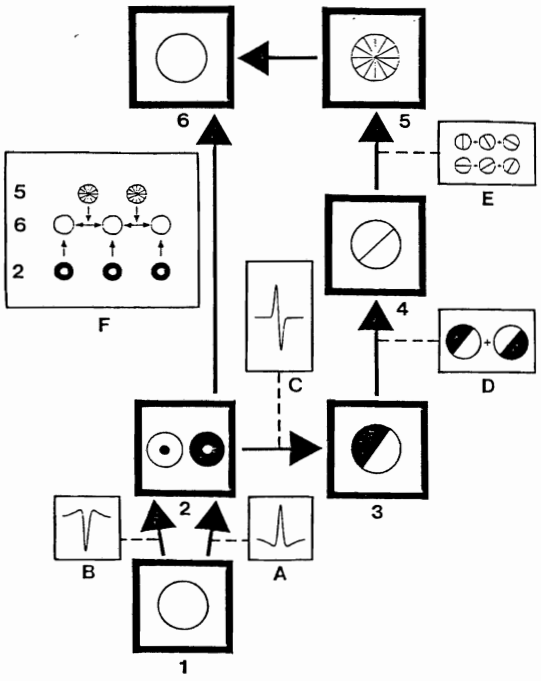


FIG. 16 Overview of the neural network model of Grossberg and Todorovic (1988). The thick-bordered rectangles numbered 1 through 6 correspond to the levels of processing within this simplified model of the FCS processing. The symbols inside the rectangles are graphical mnemonics or icons for types of computational units residing at that level. Arrows depict interconnections between levels. Inset F illustrates how the activity at Level 6 is modulated by outputs from Levels 2 and 5. See text for additional details.

of the full model and is taken from the paper by Grossberg and Todorovic (1988). There are six levels to the neural network model. Level 1 is simply the stimulus distribution sampled on a discrete lattice. Level 2 includes center–surround spatial filters within the FCS constructed so that they do all the work of the retina: spatial filtering and automatic gain control to achieve Weber's law. Levels 2 feeds into both BCS and FCSs. Levels 3, 4, and 5 are within the BCS. Level 3 is a grid of orientation-sensitive neural elements that retain sensitivity to the sign of contrast. These could be thought to be representative of the ensemble of simple cells in Area V1 of the cortex. The elements of Level 4 are orientation-sensitive but contrast sign-insensitive neurons and are thought to correspond to cortical complex cells. Level 5 cells are proposed to be a sheet of neurons that integrate signals from Level 4 "complex cells." The Level 5 cells are the BCS units. Level 6 is composed of a syncytium of cells within the FCS. Any point in this syncytium receives input from a small group of center–surround units from Level 2. Signals from corresponding points in Level 5, the BCS, can reduce diffusion of signals within the Level 6 syncytium by inhibitory gating. Therefore, a BCS signal creates a barrier to filling-in at its target cells. The map of activity across the Level 6 FC syncytium constitutes the internal representation of the visual field in the model although there is some input to the object recognition system (ORS) directly from the BCS and FCS. The network equations for each of the levels in the model are given in the Appendix of the paper by Grossberg and Todorovic (1988). They suggest that the FCS may receive input from the "blob" system in V1 cortex, while the BCS may get its input from the hypercolumn system.

A. Elementary Properties of Boundary Contour and Feature Contour Interactions

The BCS and the FCS were originally introduced to account for paradoxical data concerning brightness, color, and form perception, including the percepts of illusory contours. BCS signals are used to synthesize boundaries, whether real or "illusory," that the

perceptual process generates. FC signals initiate the filling-in processes whereby brightness and colors spread until they hit either their first boundary contour or are attenuated due to spatial spread.

The BCS is insensitive to sign of contrast as evidenced by Fig. 15, discussed earlier (Cohen & Grossberg, 1984; Grossberg & Mingolla, 1985a,b; Shapley & Gordon, 1985, 1987). A color and brightness system, such as the FCS, must remain sign-sensitive. In Fig. 15, two vertically oriented and spatially aligned edges generate an intervening visual boundary across a region that has no color or luminance contrast. The boundary completion propagates *inward* between pairs of inducing elements and is *oriented*. On the other hand, many experiments discussed below indicate that color and brightness signals (produced by hypothesis by the FCS) diffuse outward away from scenic edges in an unoriented manner to "fill in" regions with their own featural quality (Krauskopf, 1963; van Tuijl, 1975; Yarbus, 1967).

The BCS can be used for texture segmentation, as for example in the case of Glass patterns (Glass & Switkes, 1976; Fig. 17). These texture boundaries may be "invisible but recognizable," that is, boundary signals may excite the BCS without having the FCS indicate that they are bounding regions of differing color or brightness.

FIG. 17 A Glass pattern. The emergent circular pattern is recognized although it is not "seen" as a pattern of differing contrasts. From Glass and Switkes (1976).

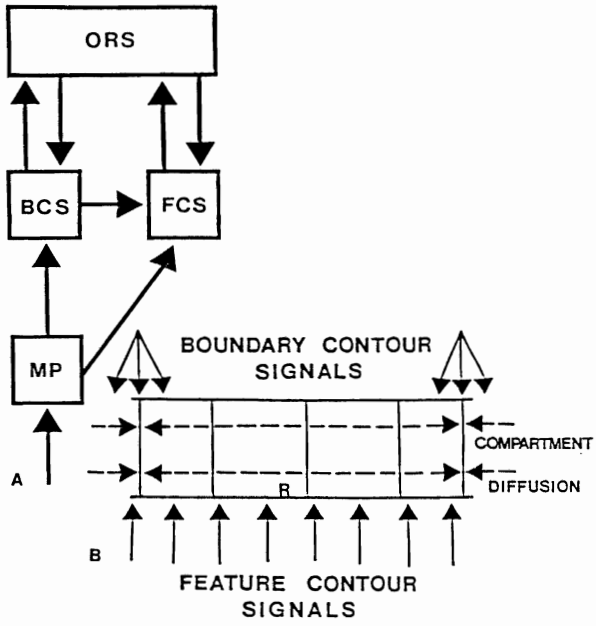


FIG. 18 (A) Macrocircuit of processing stages. Monocular preprocessed signals (MP) are sent independently to both the boundary contour system, BCS, and the feature contour system, (FCS). Interactions of BC system signals within the FC system are needed to support visible brightness and color contours. The BC system also sends bottom-up signals to and receives top-down template signals from the object recognition system (ORS). From Grossberg (1987a). (B) A monocular brightness and color stage within the FC system. Monocular FC signals activate cell compartments which permit rapid electronic diffusion of activity, or potential, across their compartment boundaries, except at those boundaries which receive BC signals from the BC system. Consequently, the FC signals are smoothed and averaged except at boundaries that are completed with the BCS.

The presence within the FCS of different filled-in signals on opposite sides of a boundary is, according to Grossberg's model, necessary for sustaining the perception of figural form. Object recognition can, however, be based on signals from the BCS and FCS to the object recognition system (ORS), which Grossberg postulates is a higher stage that integrates the signals from BCS and FCSs, as illustrated in Fig. 18A,B.

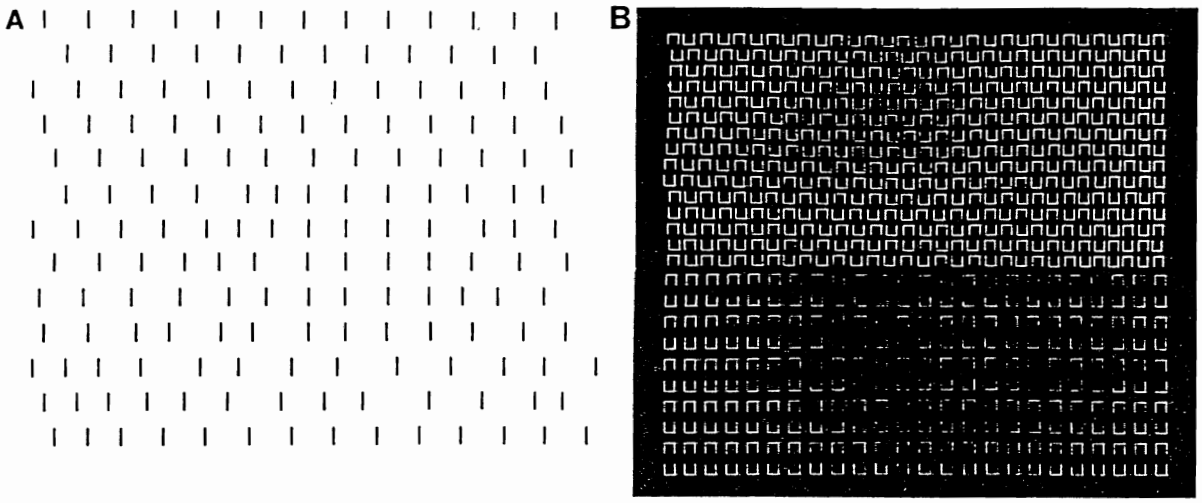


FIG. 19 (A) Emergent features. The collinear linking of short line segments into longer segments is an "emergent feature" which sustains textural grouping. (B) The diagonal grouping at the top is initiated by differential activation of diagonally oriented receptive fields, despite the absence of any diagonal edges in the picture. Horizontal cooperation of signals at the ends of vertical lines generates horizontal subjective contours at bottom. From Beck, Prazdny, and Rosenfeld (1983).

B. Emergent Features in Texture Segmentation

Beck (1983) has identified key variables affecting texture segmentation. Displays like the ones in Fig. 19 indicate that the slopes of texture elements are a critical determinant of grouping, with regions containing many features with similar slopes tending to group. If certain of these features are distributed in a regular manner, collinear groupings of these texture elements may become "emergent features" capable of segmenting one textural region from another (as in Fig. 19A). Such emergent features need not be in line with the slopes of local contrasts (Fig. 19B).

A remarkable aspect of displays such as Fig. 19 is that we see a series of short lines despite the control of perceptual grouping by the long emergent features. This may be explained as follows. Within the BCS, a boundary structure emerges corresponding to the long lines visible in Fig. 19A. This structure also includes short horizontal components near the endpoints of the lines. Within the FCS, these horizontal signals prevent featural filling-in of dark and light contrasts from crossing the boundaries corresponding to the short lines. On the other hand, the output from the BCS to the ORS reads out a long line structure without regard to which subsets of this structure will be seen as dark or light.

C. Hierarchical Resolution of the Boundary Uncertainty Induced by Orientational Sensitivity

The horizontal signals that are generated in response to the ends of vertical lines (Fig. 19B) illustrate a remarkable adaptation of the visual system, which led Grossberg and Mingolla (1985a, 1985b) to assert that *the ends of all thin lines are "illusory"*. The perceived line ends correspond to actual luminance differences and so are not illusory in that sense. However, the boundary signals indicating the position of a thin line end are weak or absent at the stage of BCS processing that corresponds to cortical complex cells (Fig. 20). The line end is detected in response to the spatial pattern of output signals from complex cells by a later stage of BCS processing that corresponds

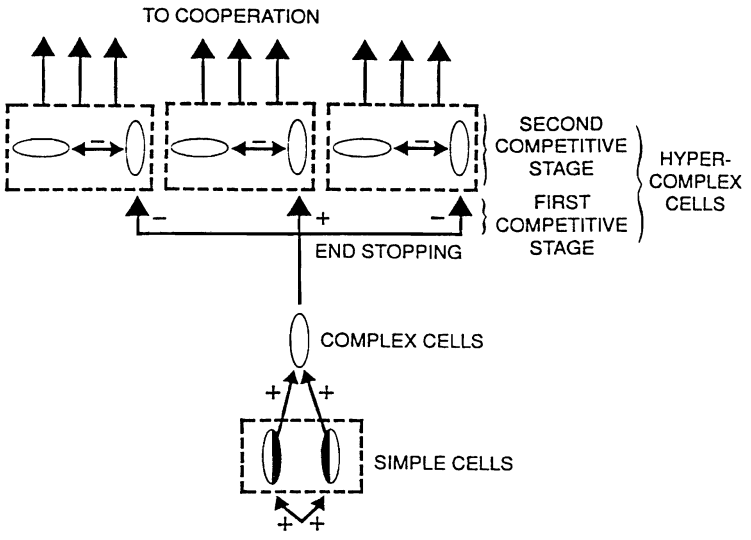


FIG. 20 Early stages of BC system processing. At each position, cells exist with elongated receptive fields of various sizes which are sensitive to orientation, contrast, and sign of contrast. These are the simple cells at Level 3 in the BC system. Pairs of such cells sensitive to like position and orientation but opposite directions of contrast form the input to cells that are sensitive to orientation and contrast magnitude but not contrast sign. Such a complex cell neuron, at Level 4 in the model, is indicated by the empty ellipse. These cells in turn excite like-oriented hypercomplex cells corresponding to the same position and inhibit like-oriented hypercomplex cells corresponding to nearby positions at the first competitive stage (upper dashed boxes). Also, at this level, cells corresponding to the same position but different orientations inhibit each other via a push-pull competitive interaction.

to cortical hypercomplex cells. This stage is not present in the simplified model shown in Fig. 16. Grossberg and Mingolla (1985a) called the hypercomplex responses at the line end end cuts.

Why cannot the complex cells detect the end of a thin line? The model of simple cells and complex cells developed by Cohen and Grossberg (1984) and Grossberg and Mingolla (1985a) is similar to the models proposed by Shapley and Gordon (1985) and Spitzer and Hochstein (1985). At each position, simple cells with like-oriented, but oppositely polarized, receptive fields respond to local image contrasts. Their outputs are rectified before inputting to complex cells of similar orientation. Such complex cells are not sensitive to the direction of image contrast. At the ends of thin lines and corners, these cells are insensitive to all orientations due to the very receptive field elongation that enables them to be sensitive to the orientations of scenic contrasts in edges, texture gradients, and shading that pass through the receptive field. Compensatory processing by later BCS stages is needed to generate end cuts, so that color signals cannot flow from every line end within the FCS (Fig. 18B), or for that matter from every scenic corner. Despite these precautions, colors can sometimes be seen to flow, as noted below, and the processing rules of the BCS provide an explanation of how this happens.

D. Spatial Short-Range Competitive Interactions

Within the BCS, the end cut transformation converts complex receptive fields into hypercomplex and higher-order hypercomplex receptive fields. This requires two successive stages of short-range inhibitory or competitive interactions. These interactions account for results of von der Heydt, Peterhans, and Baumgartner (1984) and have been used to explain many anomalous properties of filling-in and neon color spreading (Grossberg & Mingolla, 1985a; Grossberg, 1987a) that have been reported previously (Van Tuijl, 1975; Van Tuijl & de Weert, 1979; Redies & Spillmann 1981; Redies, Spillmann & Kunz, 1984). Hyperacuity (see Chapter 10) also depends on these inhibitory interactions in the model (Grossberg, 1987a) as does the café wall illusion shown in Fig. 16 (Grossberg & Mingolla, 1985b). Lateral inhibition is postulated between complex cells to like-oriented hypercomplex cells at nearby positions (Fig. 20). This can

account for the effect of a flanking line on the perceived position of a test line, reported by Badcock and Westheimer (1985). Badcock and Westheimer's repulsion effect was independent of sign of contrast and could be generated monoptically or dichoptically, indicating that it would be due to interactions within the BCS at Level 4 or higher.

The phenomenon of end cuts illustrates that the uncertainty of retinal information is not progressively reduced by each successive stage in cortical processing. Each stage may eliminate one uncertainty while it generates yet a new uncertainty, much as in the Heisenberg uncertainty principle. For example, the elongated receptive fields of simple cells and complex cells generate "orientational certainty" in response to scenic edges and bars, but thereby generate "positional uncertainty" at line ends and corners. Subsequent short-range competitive interactions give rise to hypercomplex cell receptive fields which resolve this positional uncertainty by generating end cuts.

E. Spatially Long-Range Cooperative Interactions

Once perpendicular boundary signals are generated at line ends, they are processed by subsequent BCS stages according to the same rules as boundary signals originating in direct response to scenic luminance contrasts. Thus, in the texture displays of Fig. 19, boundary completion and grouping can occur either along the direction parallel to thin line segments or in a direction parallel to line ends. End cuts thus increase the number of potential boundary groupings for each scene. Moreover, because oriented receptive fields at Level 3 in the model are local contrast detectors, rather than edge detectors per se, certain combinations of local contrasts from disconnected scenic elements can at times trigger responses along orientations where there are no local edges. Thus, many potential groupings of parts into wholes are activated during preperceptual processes. In a final percept, a single global grouping is chosen and sharpened while all other possible groupings are actively suppressed.

The competitive interactions shown in Fig. 20 are insufficient for choosing a global grouping if only because their interactions are short-range. These competitive stages of the BCS occur between Levels 4 and 5 in the network of Fig. 16. A spatially long-range cooperative process occurs at Level 5 itself. Figure 21A schematizes the BCS model of this cooperative process. Figure 21B embeds it within the total architecture of the BCS.

A cell at Level 5 of the BCS, modeled as in Fig. 21B, can be active only if it receives sufficiently large inputs from paired populations of similarly oriented and spatially aligned cells at the competitive stage. These cooperative cells thus behave like the logical gates reported by von der Heydt et al., (1984). The cooperative cells were labeled *bipole* cells by Grossberg and Mingolla. These bipole cells (Level 5) feed back excitatory signals to competitive cells (presumed hypercomplex) at the pre-Level 5 stage. The hypercomplex cells, in turn, feed excitation backward to bipole cells. In this way, boundary completion can propagate inward rapidly in a spatially discontinuous manner.

The competitive cells whose orientations and positions receive the largest cooperative signals are favored to win out over less favored cells within the final perceptual grouping. Thus the competitive stages do double duty. They generate end cuts in response to bottom-up signals, and they help to choose the final segmentation in response to top-down signals. The cooperative–competitive feedback network which acts to choose the final segmentation is called the CC loop. Within the CC loop, boundary segments emerge at those locations and orientations that possess enough statistical inertia to survive the cooperative–competitive feedback exchange.

F. All Boundaries Are Invisible

Unless a connected boundary can be synthesized by the BCS, it cannot separate the FCS's signals into domains capable of supporting different colors or brightnesses. Thus if a later event inhibits through the CC

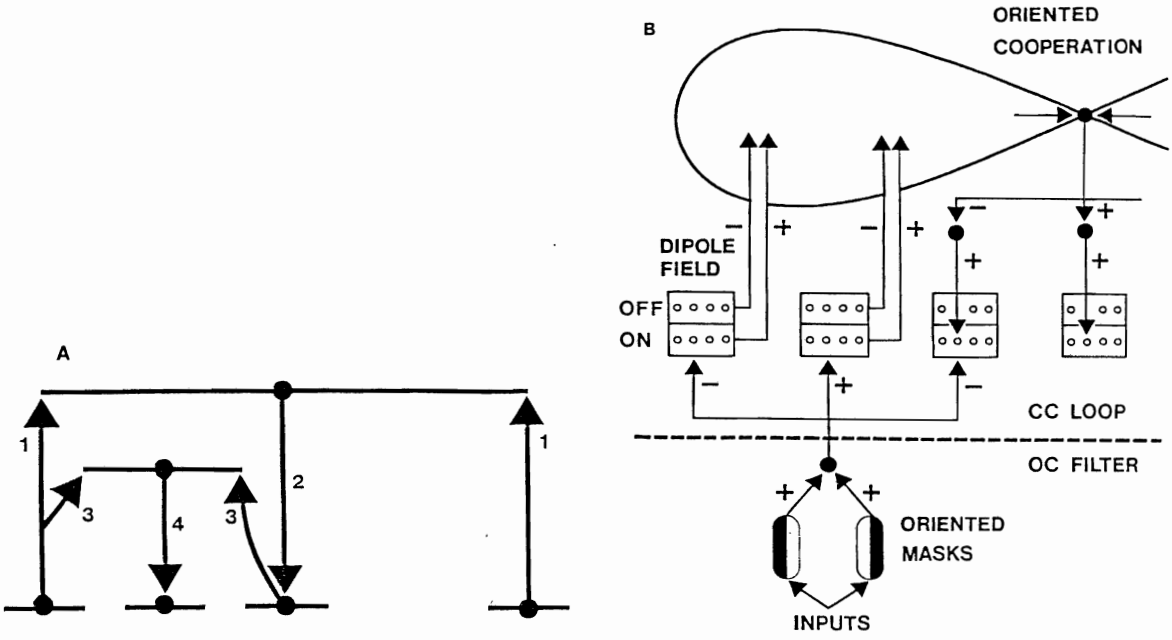


FIG. 21 (A) The pair of pathways 1 from hypercomplex cells activate a bipole cell that generates a positive boundary completion feedback along pathway 2. Then pathways such as 3 activate positive feedback along pathways such as 4. Rapid completion of a sharp boundary between pathways 1 can hereby be generated by a spatially discontinuous bisection process. (B) Circuit diagram of a monocular BCS: Inputs activate oriented simple cell receptive fields which cooperate at each position and orientation to activate complex cells before feeding into an on-center off-surround interaction. This interaction excites like-orientations at the same position and inhibits like-orientations at nearby positions, thereby giving rise to hypercomplex cells. These cells are on-cells within a dipole field: On-cells at a fixed position compete among orientations. On-cells also inhibit off-cells which represent the same position and orientation. Off-cells at each position, in turn, compete among orientations. Both on-cells and off-cells are tonically active. Net excitation of an on-cell excites a similarly oriented cooperative receptive field at a location corresponding to that of the on-cell. Net excitation of an off-cell inhibits a similarly oriented cooperative receptive field at a location corresponding to that of the off-cell. Thus, bottom-up excitation of a vertical on-cell, by inhibiting the horizontal on-cell at that position, disinhibits the horizontal off-cell at that position, which in turn inhibits (almost) horizontally oriented cooperative receptive fields that include its position. Sufficiently strong net positive activation of both receptive fields of a cooperative cell enables it to fire, as in Figure 20. Such firing generates feedback via an on-center off-surround interaction among like-oriented cells. On-cells which receive the most favorable combination of bottom-up signals and top-down signals generate the emergent boundary segmentation.

loop before a connected boundary structure is formed at Level 5, no percept may be visible. This clarifies how metacontrast might work (Breitmeyer, 1984).

Boundary signals are predicted to be invisible within the BCS. Thus, contrast sensitivity of cells within the BCS does not imply visibility of the final percept, according to Grossberg's theory. The generation within the FCS of different filled-in featural contrasts on the opposite side of a completed boundary leads to a visible percept.

G. Binocular Fusion, Rivalry, and McCollough Effect: From Ocular Dominance Columns to Multiplexed Multiple Scale Binocular Segmentation

BCS can be generalized to the binocular case (Grossberg, 1987a,b). Two steps are needed for this generalization. First, V1 cortex is organized into ocular dominance columns (Hubel & Wiesel, 1977). Binocular input is combined at the complex cell stage

producing neural elements that are sensitive to position, orientation, spatial frequency, positional and orientational disparity, yet insensitive to sign of contrast (modeled by Grossberg and Marshall, 1987, 1989). Second, the model exploits the known property of multiple receptive field sizes, each feeding into its own CC loop. Within each size scale, the multiplexed data pattern represented by that scale's complex cells activates a CC loop which selects, amplifies, and coherently groups the binocularly consistent properties within that scale while suppressing the binocularly inconsistent properties. Data about binocular rivalry and the McCollough (1965) effect, respectively, are clarified by the theory's explication of how mechanisms of segmentation and stereopsis interact and of how binocular segmentation and monocular featural filling-in mechanisms interact. Of particular interest is the discovery that unoriented, monocular networks of double-opponent color cells within the FCS can, through interactions with binocular segmentations from the BCS, extract binocularly consistent brightness and color signals whose spatial organization encodes properties of oriented form, as well as of color and brightness. Such a multiplexed representation of form-and-color-and-depth, called by Grossberg a FACADE representation, has been predicted to occur in the prestriate cortical area V4.

H. Explaining Monocular Brightness Data under Constant and Variable Illumination Conditions

The interactions between the BCS and FCSs can be used to account for some brightness phenomena that have not previously been explained by a single theory. Computer simulations (Fig. 21) illustrate how the model handles brightness constancy and brightness contrast. The properties of the neurons at Level 2 in the model (Fig. 16), with their center–surround organization controlled by shunting inhibition are what generates the brightness constancy of responses in the Grossberg model as illustrated in Fig. 22. This shunting network is designed to discount the illuminant. In order to accomplish this, the off–surround of the network cannot be too broad, and there must exist significant automatic gain control within the on-center to generate responses that are contrast dependent. These properties are consistent with the data summarized by Shapley and Enroth-Cugell (1984). These properties

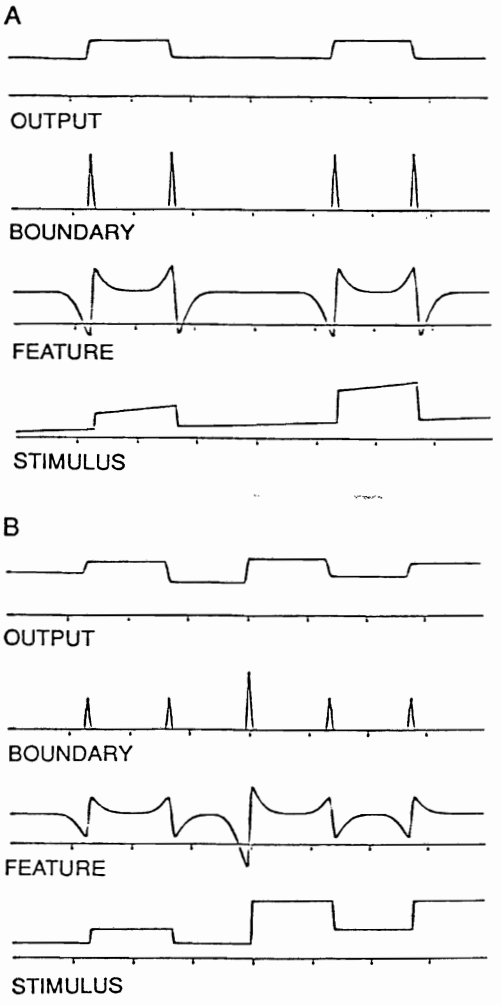


FIG. 22 Computer simulations of (A) brightness constancy and (B) brightness contrast. The rows labeled "Feature" depict the response profile of the FC system which discounts the illuminant. The rows labeled "Boundary" show the output of the BC system. Rows on top present the final filled-in pattern.

enable the model to provide a unified explanation of how different images generate brightness constancy, contrast, or assimilation when the activation patterns in which illumination is discounted activate subsequent model interactions between boundary segmentation and filling-in.

For example, Figure 22A depicts the one-dimensional cross section of the luminance distribution of an image illuminated from the right-hand side ("Stimulus" traces). The image contains two objects of equal reflectance on a less reflective background. Because of the uneven illumination, the right-hand patch has a higher luminance than the left-hand patch. The "Feature" traces represent the reaction of Level 2 neurons, center–surround retinal units with shunting excitation and inhibition that discounts the illuminant and consequently enhances high contrasts while attenuating low contrasts. The "Boundary" traces of Fig. 22A present the output of the BCS, consisting of sharp peaks corresponding to abrupt luminance steps of both polarities in the image. The "Output" traces are the model's filled-in output. The model correctly predicts brightness constancy. The filling-in process of the FCS generates the homogeneous regions with different brightness levels in the output. The importance of filling-in is also demonstrated in the model's simulation of brightness contrast in Fig. 22B.

In a simulation of a Mondrian-like image resembling Fig. 12, consisting of a number of homogeneous patches of differing luminance levels, Grossberg's model correctly predicted that local contrast would dominate over reflectance or luminance. In addition, the model has been used to simulate several classical and new variants of brightness contrast, brightness assimilation, the Craik–O'Brien–Cornsweet effect, the Koffka–Benussi ring, and the Hermann grid (see Chapter 7, this volume). The model handles a far greater variety of effects than competing computational approaches (Horn, 1974; Land, 1983; Hurlbert, 1986). These latter models are explicitly geared for reflectance recovery and must therefore fail to account for human vision on the basis of the evidence presented in Figs. 11 and 12. Also, the Retinex models alluded to have not been applied consistently to the larger selection of brightness phenomena needed to understand form from luminance adequately. In summary, neural models for cortical BCS and FCS interactions have begun to be able to account for and predict a far-reaching set of interdisciplinary data as manifestations of basic design principles, notably how the cortex achieves a resolution of uncertainties through its parallel and hierarchical interactions.

VIII. CONCLUSIONS

In order to perform the effortless act of visual perception, neural networks of substantial complexity must exist. Features, either fixed or adaptive, must be recognized by such networks. Objects are recognized and represented internally for recollection. Such tasks require spatial filters, memories, spatial organization of activity distributions in different populations of neurons, and complex competitive and cooperative interactions to regulate the network's responses. Furthermore, these interactions have a certain spatial scale intrinsic to their function that determines the stability and the characteristics of visual perception. The simplest visual tasks, such as perceiving colors and recognizing familiar faces, require elaborate computations and more neural circuitry than we have yet imagined. In the pursuit of adequate theories of vision we have learned much about the difficulty and computational requirements of the task as we have progressed toward characterizing how the task is accomplished. Theoretical progress toward the goal of understanding the visual process and how it is accomplished by nervous systems will require such a multifaceted approach for some time to come.

ACKNOWLEDGMENTS

Preparation of this chapter and the research reported by the authors were supported from many sponsors that we here acknowledge with gratitude. Robert Shapley was supported by grants from the National Eye Institute (EY1472) and National Science Foundation (BNS 8708606) and by the Sloan Foundation and MacArthur Foundation. Terry Caelli acknowledges the support of the National Research Council of Canada (Grant No. A2568). Stephen Grossberg received the support of the U.S. Air Force Office of Scientific Research (Grants F49620-86-C-0037 and F49620-87-C-0018), the U.S. Army Research Office (Grant No. DAAG-29-85-K-0095), and the National Science Foundation (IRI-84-17756). Ingo Rentschler acknowledges the Deutsche Forschungsgemeinschaft for Grant No. PO 121/13 1 Project 5.