Edited by
Peter C. Dodwell
Queen's University
Terry Caelli
University of Alberta

FIGURAL SYNTHESIS

LEA
1984 LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS
Hillsdale, New Jersey London

6 Image Transforms in the Visual System

Patrick Cavanagh
Université de Montréal

1. INTRODUCTION

The set of retinal receptors that responds to a given object in the visual field varies directly with the size, orientation and position of the object. The fact that this object can be recognized as the same object in spite of these variations implies that, at some point, the brain reduces this variable retinal input into a single pattern of neural activity that defines the object. Various approaches to the extraction of an invariant encoding have been discussed since Helmholtz first presented the problem (Cutting, 1983; Dodwell, 1982; Foster, 1977; Hu, 1962—to name a few of the articles on the subject), and most current models can be characterized as one of two types: an abstract, structural representation of the object or a transformation of the form into an analogue representation that is itself invariant to input size, position, and orientation. Although both approaches are often seen as competing models, I will argue that they are actually similar in several respects and may both be at work in the visual system.

A structural representation reduces the form to be identified to a set of primitive elements and the structural relations between them (Fig. 6.1). In the case where the primitives are, for example, lines (contours) and angles, the first stage requires the extraction of edge information and the second a determination of the relations between the edges.

Size and position are not essential attributes of the primitive elements of this type of encoding as only the presence or absence of the elements and the relations among them need be specified. As a result, the encoding is by nature invariant to the size and position of the overall pattern. The computer vision literature offers several examples of different possible primitives (e.g., lakes and bays—Fig. 6.2,

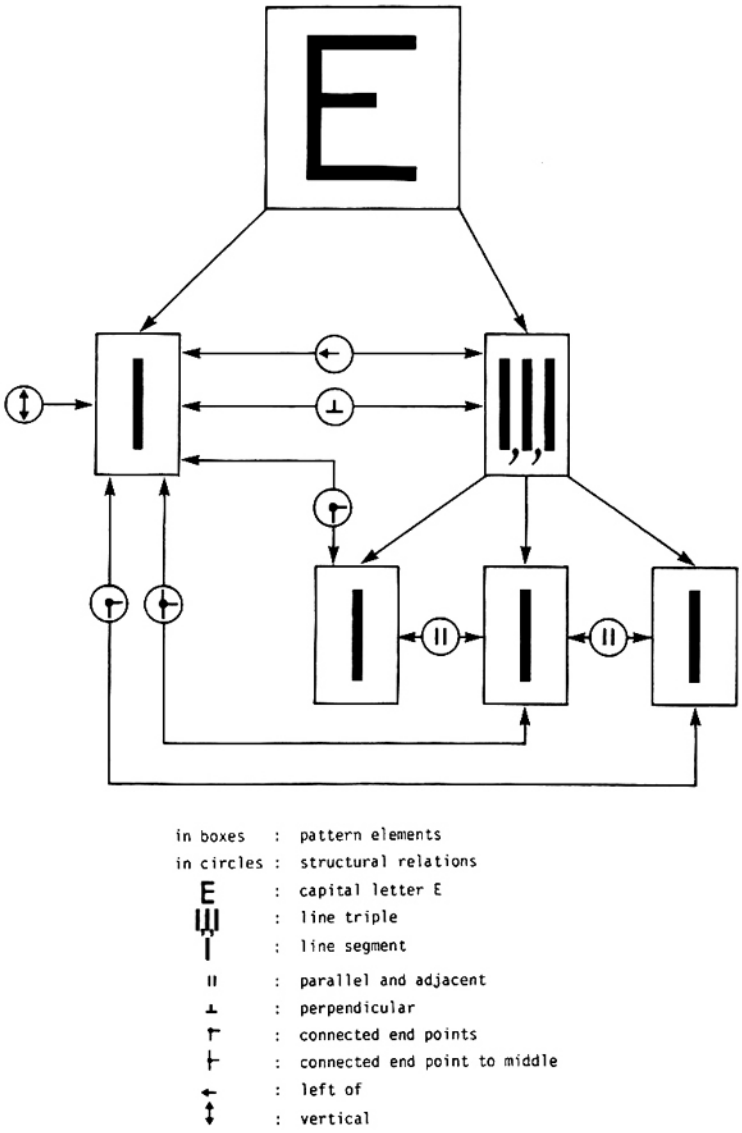


FIG. 6.1. A structural representation of the letter E. Lines, their orientations, and their positions with respect to each other form the basic encoded information. The representation is not size specific but is orientation dependent in the relations "left-of" and "vertical." These relations can be reformulated to provide orientation invariance. The tolerance of the representation to variations in the letter E depends on the lowest level of encoding: what is accepted as a line, as parallel and adjacent, as perpendicular, etc.

Duda & Hart, 1973) and notational schemes (networks, Oden, 1979; pattern grammars, You & Fu, 1979; trees, Cunningham, 1980; graphs, Shapiro, 1980; etc.). Since these representations as a whole encode the structure of the object, they are insensitive not only to size and position changes but also to large variations in style. For example, the many styles of printing a letter all, in general, retain the same basic structure but vary along several dimensions such as angle, size, aspect ratio and line thickness (Fig. 6.3). Thus, there are compelling

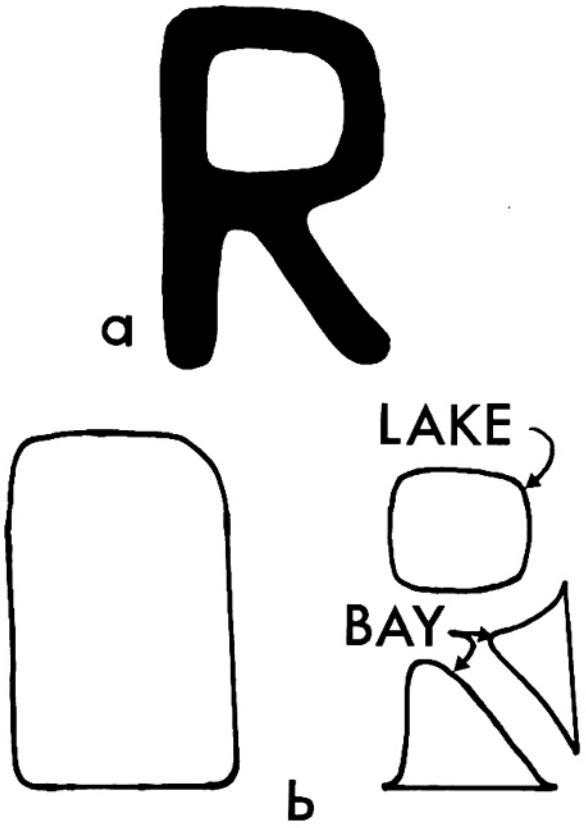


FIG. 6.2. Lakes and bays. A topological decomposition of letter shapes is produced by collapsing a rubber band around the outer perimeter of the letter, forming its convex hull. A "lake" is an open area entirely enclosed by the figure and a "bay" is an area bounded by the figure and by the convex hull. (a) Letter R. (b) Convex hull, bays and lakes. This representation is size invariant.

theoretical and practical reasons for proposing that the brain may use structural representations for encoding and recognizing visual patterns. The possible mechanisms by which the brain might implement these structural encoding procedures have only been partially sketched out in the literature however. The initial stage of primitive extraction—in the case of edge descriptions—can be related in a straightforward fashion to the action of the oriented line detectors in the striate cortex (Hubel & Wiesel, 1962, 1968; see Barlow, Narasimhan, & Rosenfeld, 1972; Marr, 1982). On the other hand, the analysis of the structural relations between the elements and the encoding of the overall representation finds little in the way of candidate neural mechanisms in the current literature (although see the chapters in this book by Caelli, Grossberg and Hoffman).

The alternate approach to pattern encoding involves transformations of the stimulus patterns into representations that do not vary with the size and position of the input. The fibre bundle that connects the retinal array, through the lateral geniculate body to the striate cortex performs a particular, at present only partially understood, transformation of the input pattern. The connections from the striate array to the prestriate surfaces and from there to the inferotemporal cortices perform additional unknown remappings of the visual representation. The nature of the repeated remapping of the information of the visual cortices appears to be well suited to the possibility of transforming visual patterns into domains where recognition and classification are more easily accomplished.

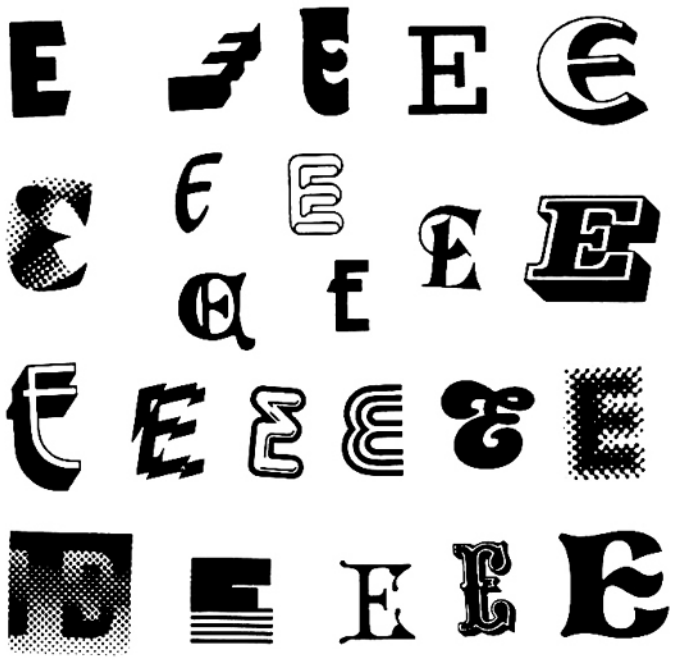


FIG. 6.3. Variations on the letter E. Most of these examples can be identified in isolation by human observers. A simple encoding of lines and angles would, however, be highly misleading in several of these examples due to the addition of depth information and extraneous, stylistic contours. Some preprocessing such as blurring might be useful in order to ignore irrelevant details.

There have been several proposals for particular transformations occurring in the visual system: compensatory (Foster & Mason, 1979; Marko, 1973), Fourier (Campbell & Robson, 1968; Pollen, Lee, & Taylor, 1971), densely connected (Kabrisky, Hall, Goble, & Gill, 1971), autocorrelation (Uttal, 1975), log polar (Chaikin & Weiman, 1979; Schwartz, 1977, 1980), log polar frequency (Cavanagh, 1978a), contour spectrum (Desimone, Schwartz, Albright, & Gross, 1982), and composite pseudo Wigner (Jacobson & Wechsler, 1982), among others.

The proposal of Fourier analysis in the visual system followed the work of Schade (1956) who applied Fourier analysis to the transmission quality of the visual system and Campbell and Robson (1968) who proposed that the visual system itself performed a Fourier analysis on the stimulus pattern. The hypothesis of a global Fourier transform at the level of the striate cortex is quickly ruled out, however, as the receptive fields are restricted to small local areas and do not cover the entire visual field as would be required. Various piecewise Fourier analyses have been proposed (Glezer & Cooperman, 1977; Pollen, Lee, & Taylor, 1971; Robson, 1975) but even if the Fourier transform were realized at the striate level this transform obtains neither size nor rotation invariances (Casasent & Psaltis, 1976) and so would be of little use on its own.

Schwartz (1977) has proposed that the log polar organization of the retinotopic mapping on the striate cortex can support size invariance. However, this

mapping because of its dependence on position is of marginal value in pattern analysis (Cavanagh, 1981, 1982).

If, rather than looking at the overall organization of the striate cortex, we look at the local structure, a very striking and potentially useful organization is revealed—that of a log polar frequency transform (Berardi, Bisti, Cattaneo, Fiorentini, & Maffei, 1982; Cavanagh, 1978a; Maffei & Fiorentini, 1977). The goal of this chapter is to show how this organization might contribute to the pattern analysis process. The next section demonstrates the useful properties of a log polar frequency transform and the subsequent section outlines the physiology that may underly this process.

A brief note concerning the use of the terms "spatial frequency" and "spatial frequency detector" should be added here. The receptive fields of striate cells cover only local areas and so do not encode true Fourier spatial frequency components. They are, however, more selective to spatial frequency content than to other more naturalistic characteristics such as bar width (Albrecht, De Valois, & Thorell, 1980). For simplicity, the term "spatial frequency detector", as opposed to bar width detector or size detector or orientation detector, is used throughout to describe the encoding operation of simple and complex cells. This can be viewed as a convenient label for a process that in fact only approximates spatial frequency analysis.

2. LOG POLAR FREQUENCY TRANSFORM

This section demonstrates the size invariance properties of a transform sequence based on the Fourier amplitude representation. It should be stressed that this sequence has drawbacks both in terms of its pattern recognition capabilities and its physiological realizability. These drawbacks are outlined and it will be seen that in modifying the sequence to be more physiologically appropriate, the pattern recognition capabilities are simultaneously improved.

A log polar frequency transform (Brousil & Smith, 1967; Casasent & Psaltis, 1976; Cavanagh, 1974, 1978a) arrays the spatial frequency components of the input pattern along orthogonal axes of orientation and the logarithm of the spatial frequency (in Fig. 6.4, the letter F is shown at various sizes and orientations, a G and an E are included for comparison). To generate the demonstration, the frequency components used are the Fourier amplitude coefficients of the input (Goodman, 1968) but are arrayed on orientation and log spatial frequency axes rather than Cartesian frequency axes (x and y). The transform has three important properties.

Position Invariance. Because the information is in terms of spatial frequencies, the values can be represented as the amplitude and the phase of the sinusoidal components present in the input. The amplitude reflects the strength of a

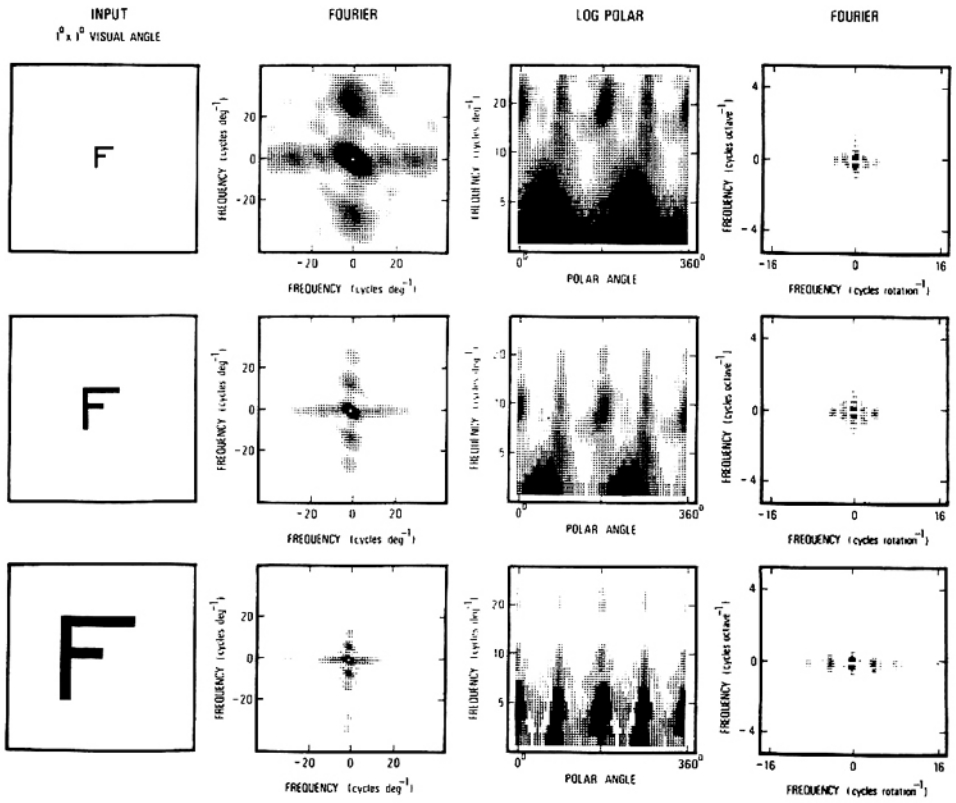
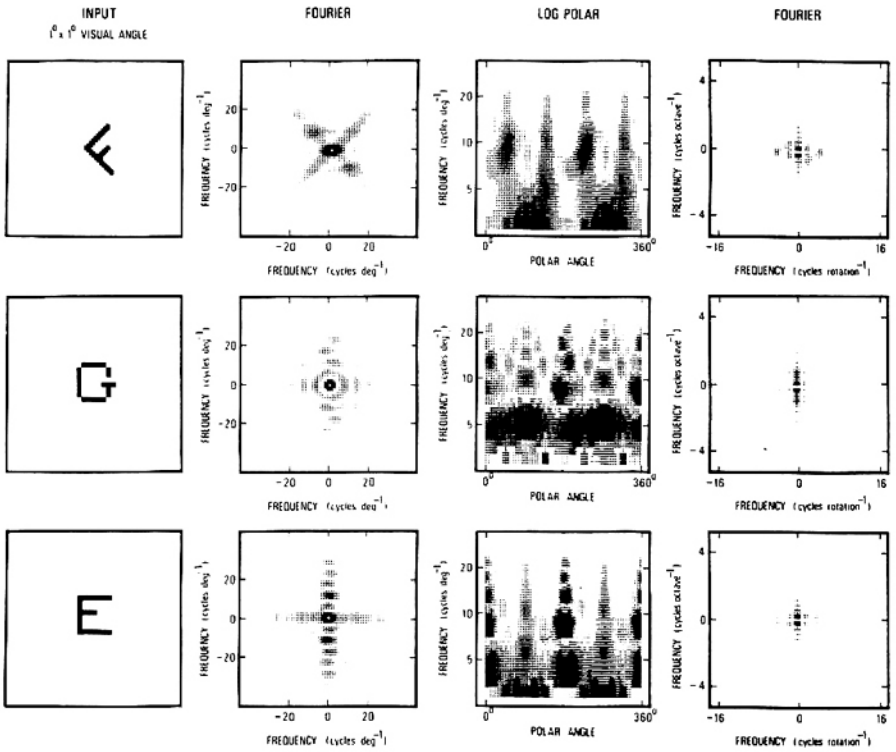


FIG. 6.4. A size and position invariant encoding sequence. The letter F is presented in three different sizes (7.5, 15, and 30 minutes of visual angle) and at a nonupright orientation (315 degrees, size—15 minutes of visual angle). Opposite, a G and an E are also presented for comparison. Each input is shown at three stages of transformation: a Fourier amplitude transform, a log polar mapping of the Fourier amplitude transform and a Fourier amplitude transform of the log polar

component and the phase its position. The key attribute of this encoding is the position invariance that is obtained when amplitudes are considered: The amplitude values are the same no matter where in the input field the pattern lies.

The basic patterns encoded by the Fourier transform, the primitive features, are sinusoidal grids extending over the entire input field and these features are size and orientation specific. As a result, the transform will be affected by scale and orientation changes. Other basis features that might be considered as alternatives to the sinusoidal grids—oriented bar detectors (Hubel & Wiesel, 1962, 1968), Gabor signals (Kulikowski, Marcelja, & Bishop, 1982; Marcelja, 1980)—are also specific to size and orientation. The set of encoded features would therefore change with size and orientation for these functions as well.

Although these representations would be affected by size and orientation factors, it is possible to arrange the dimensions of the encoding so that size and orientation invariances can be obtained. For the log spatial frequency and orientation dimensions described here, the overall pattern of the encoded features remains the same and its position in the transform space simply shifts as a



mapping. The images are presented on matrices of 64 × 64 resolution with amplitude being represented in 30 equal intervals between the minimum and maximum of each image by the density of dots. The final Fourier amplitude transform is performed on the log polar representation after it is reduced to 32 × 32 resolution and zero mean amplitude with the rest of the 64 × 64 array set to zero. This minimizes aliasing and aperature effects (Goodman, 1968).

function of the size and orientation of the stimulus. Note that it is the arrangement of the axes that provides the essential size and orientation invariances. The critical requirement of the basic features themselves, whether bar width, Gabor signals, sinusoidal grids or other, is the capacity to provide a position invariant encoding—as does the amplitude value of the spatial frequency representation.

Orientation Axis. Since one axis of the two dimensional transform space (see Fig. 6.4, third column) is orientation, a change in the orientation of the input shape changes all the orientations of the component features by the same amount and the whole pattern simply shifts along the orientation axis. Notice in Fig. 6.4 that the rotation of the input letter has only shifted the feature pattern (Fig. 6.4, third column, F at 45 degrees versus upright F of same size). The pattern as a whole is unchanged with the exception that the part shifted off the right border reappears on the left border and vice versa.

Log Size Axis. When the input shape changes size by a factor x, all the component features are scaled by the same factor. If a stimulus were halved in

size, a feature that had a size 10 (in arbitrary units) would become a size 5, a size 5 would become a 2.5 and so on. (Since a smaller feature has a higher spatial frequency content, the changes in terms of spatial frequencies would be in the opposite direction, 10 cycles per degree becoming 20 cycles per degree, etc.) In order for these features to shift by a constant amount along the size axis, the axis must have a logarithmic scale. A constant multiplicative factor then becomes a constant linear shift. This constant linear shift of all components ensures that the overall pattern of the transform is retained (Fig. 6.4, third column for the three Fs of different sizes) and only shifted as a whole (with the exception of information being lost or being added at the transform borders for a size axis of finite length).

These three aspects combine to provide an encoding that is invariant to position of the input pattern and that shifts for changes in size and orientation. This encoding represents an explicit mechanism embodying the suggestions by Milner (1974) concerning angle and length-ratio feature detectors and those by Blakemore and Campbell (1969) concerning spatial frequency ratios.

In order to obtain an encoding that is strictly form specific, a final stage of processing must be assumed that can extract the constant pattern at the log polar level, independently of its position. A Fourier amplitude transform of the log polar representation is used to demonstrate this (Fig. 6.4, fourth column) although other transforms or feature maps with position invariance would certainly be sufficient. At this final level, the similarities of the different size, position and orientations of the input are captured by a fixed pattern. (Size, orientation and position information are no longer represented at this level and must be processed by other means.) Such fixed patterns could then be used for identifying future instances of the same pattern at new locations, sizes and orientations. Such encoding schemes make the use of correlational (Anderson, Silverstein, Ritz, & Jones, 1977; Kohonen, 1977) and holographic filter (Cavanagh, 1975, 1976) memories a practical suggestion for the visual system. Table 6.1 shows the correlations of the final transforms of the variously sized and oriented letters F against the transforms for the letters F, E, and G (point by point Pearson correlations of two 64 × 64 value matrices at a time). The input is correctly recognized

TABLE 6.1.
Correlations between the final transform representations of the E, G and F at 15 minutes of visual angle and the final transforms of the variously sized and oriented Fs.

	F	F	F	F
F	.89	1.00	.94	.99
G	.86	.83	.87	.86
E	.88	.93	.91	.93

(has the highest correlation) as an F for all input conditions. Note, however, that the discrimination between the E and the F is only moderate. Even though these are two highly confusable letters, the small degree of discriminability is a reflection of the incomplete nature of pattern information in the Fourier amplitude domain. It is shown in a subsequent section that when additional relative phase information is combined with the amplitude representation discrimination is dramatically improved.

For simplicity in generating this demonstration, the Fourier amplitude components were assumed as the basis of the initial encoding. Although sufficient for these demonstrations, there are three specific problems with the Fourier amplitude transform as a model of visual encoding.

1. It is clear that the cells of the striate cortex do not act as spatial frequency detectors. The principal difference is that the receptive fields of these cells are responsive over only local areas rather than over the entire visual field. If a cell were to be a true spatial frequency "detector" it would have to have a receptive field that covered the entire visual field and whose sensitivity varied sinusoidally along one direction and not at all along the orthogonal direction. The local nature of the striate receptive fields is in direct contradiction to this requirement. A few authors have suggested (Kulikowski et al., 1982; Marcelja, 1980; Pollen, Nagler, Daugman, Kronauer, & Cavanagh, 1983) that the receptive fields that are seen are actually small patches of a sinusoidal grid—Gabor profiles. This shape has the advantage of providing both spatial (location) and frequency (size) information in an optimal way. No effort has yet been made to show how this optimal encoding might fit into an overall form-encoding process, however, although local analysis of texture (Robson, 1980) is one possibility. Whatever the specific characterization of the receptive field, its local nature renders the general position invariance that is a prerequisite to size invariance unavailable at the striate level.

2. The spatial frequency amplitude transform is not only physiologically unrealistic but is also a rather poor choice for a pattern recognition process. In particular, it is insensitive to the position of the sinusoidal components and so is essentially a feature list with no indication of the interfeature relations. An infinity of patterns can be made from the same set of sine waves just by varying their relative positions and these variants are not reflected in the amplitude transforms. Two particular variants underline this deficiency: 180 degree rotations and negative images are indistinguishable on the basis of the amplitude transforms.

3. Nonlinearity. For a linear system, the sum of the responses to several stimuli is equal to the response to the sum of the stimuli. This principle holds true for the Fourier transform itself (with real and imaginary components, the Fourier transform of the sum of two patterns is the sum of their individual transforms) but not for the AMPLITUDE component in isolation. Because position is not repre-

sented by the amplitude components, the sum of the amplitude transforms of two patterns will not reflect the position-dependent interactions between their shared frequency components. This is an unavoidable characteristic of any position invariant transform. If the encoding is invariant to the pattern's position, the sum of two transforms cannot reflect interpattern relations. The encoding of the sum of the two patterns will, however, be affected by interpattern spacing.

This nonlinearity interferes significantly with the ability to recognize a given stimulus encoded in the presence of others. Using the full, linear Fourier transform, it is possible to identify a letter A in a field of other letters no matter where the letter is, how many other letters there are, and whether or not it overlaps with other letters (Vander Lugt, 1964). The nonlinearity of the Fourier amplitude transform makes this task increasingly difficult as the number of nontarget letters increases. Moreover, since the amplitude transform is position invariant over the entire input field, the interference created by distractor letters is just as severe for letters at opposite sides of the input field as for overlapping letters. The visual system itself has difficulty in identifying overlapping figures, but the interference decreases with increasing separation between a target and a distractor. Thus, although the visual system may also suffer some nonlinear interference due to position independent encoding, it would appear to be less severe and probably limited to cases of overlap (as in hidden figures and masking paradigms) and adjacency (metacontrast).

To summarize, the Fourier amplitude transform is not an appropriate choice for pattern recognition and is not a realistic representation of the encoding in the visual system, certainly not that of the striate cortex. Evidently, to achieve a general pattern recognition capability, the local information available at the striate level must be integrated in some manner and the basis set must be something other than the amplitude of the sinusoidal components of the Fourier transform. Specifically, the encoding of interfeature relationships would be important. The following section describes in more detail the nature of the local receptive fields, how they may be organized into local log polar frequency transforms that avoid the problems of Fourier amplitude encoding, and how these might further be processed to provide the necessary invariance properties.

3. PHYSIOLOGICAL CONSTRAINTS

Receptive Fields

If we consider the general classifications of the cells of the striate cortex, four types have been reported (Hubel & Wiesel, 1962; 1968): nonoriented, simple, complex, and hypercomplex. The nonoriented cells provide no more advanced coding than that available at the retinal level—their role in a form encoding

process is therefore unclear. Simple and complex cells respond to oriented bars of a particular width. These dimensions are the basis of the form encoding process described here. The hypercomplex cells prefer oriented bars of a particular width and LENGTH. Although Julesz (1981) has proposed that these cells could underlie his terminator textons, they play no role in the encoding discussed here.

The receptive field sensitivity profiles of simple cells typically show two or more adjacent, elongated subfields having antagonistic excitatory and inhibitory influences. A bright bar aligned with the excitatory subfield or, if there is more than one, a grating (several parallel bars spaced to fall on each excitatory region) gives the maximum response for this cell (Albrecht, De Valois & Thorell, 1980). The output from simple cells therefore depends directly on the position of the stimulus, consequently, no position invariance is available at the level of the simple cells.

Complex cells respond best to moving bars and exhibit no spatially distinct excitatory or inhibitory regions, responding to a drifting bar uniformly over the entire receptive field as long as it is appropriately oriented and of the preferred width (Glezer, Tscherbach, Gauselman, & Bondarko, 1980; Heggelund, 1981; Hubel & Wiesel, 1962, 1968).

Some authors (Hubel & Wiesel, 1962, 1968; Pollen & Ronner, 1981) have suggested that the simple cells may be feeding into the complex cells to produce the position independent nature of the latter's response. That is, if a set of simple cells tuned to the same frequency and orientation are staggered by ½ cycle or less each (i.e., the width of the excitatory subregion) along an axis perpendicular to the preferred orientation are fed into a complex cell, the complex cell could respond to a bar independently of its position within the field. The bar would always be falling on the excitatory region of at least one of the simple cells. If the receptive field profile of the *i*th simple cell of the set feeding the complex cell were given by $R_i(x,y)$ and the stimulus intensity distribution given by $S(x,y)$ then C, the output of the complex cell would be given by

$$C = \sum_i \text{MAX}\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R_i(x,y)\cdot S(x,y)dxdy,\ 0\right] \tag{1}$$

Because the responses of the simple cells are rectified (represented by the MAX function), that is, cannot become negative when the effect of the stimulus is strongly inhibitory, the sum will be fairly independent of the stimulus position if the receptive fields of the set are sufficiently closely spaced. According to Pollen and Ronner (1981), simple cells that are physically adjacent respond to similar frequencies and orientations and are offset by 90 degrees (¼ cycle) along a direction perpendicular to their preferred orientation. The staggering of simple cells feeding a complex cell may then be in ¼ cycle steps. Pollen and Ronner (1981) have presented a possible organization for this summation that differs

slightly from that given here but that achieves the same result. Some authors have suggested that complex and simple cells are not hierarchically organized but rather receive their inputs in parallel from lateral geniculate cells (Stone, Dreher & Leventhal, 1979). Whether the organization underlying the position independence of the complex cell is hierarchical or parallel has no direct impact on the model being described here as either organization can produce equivalent receptive field substructures and the essential property of position independence.

Movshon, Thompson and Tolhurst (1978a) have probed complex cells' receptive fields with line pairs and have shown a number of discrete but spatially overlapping subunits that individually seem to act in a more or less linear way. The overall response amplitude of a complex cell is, in fact, fairly linearly related to input contrast (Tolhurst, Movshon, & Thompson, 1981).

Because the position invariance of the complex cells within their receptive fields is an important step towards the overall position invariant features necessary for the form encoding process, it is assumed that the output of the complex cells conveys the essential pattern information being passed on to subsequent stages and that the simple cells are not further implicated in the processes described here although they may participate in other parallel analyses (Burr, Morrone, & Maffei, 1981).

The cortical layers in which complex cells predominate are layers II and III (Hubel & Wiesel, 1962). These are also the layers that, in monkey and therefore most probably in man, project to the prestriate cortex (Lund, Lund, Hendrickson, Bunt, & Fuchs, 1975; Spatz, Tigges, & Tigges, 1970), while other layers project principally to subcortical centres.

For the purposes of the processes described here the essential characteristics of the complex cells of the striate cortex are as follows:

1. They are selective for the size and orientation of visual stimuli.
2. Their receptive fields are local. Global position invariance must be achieved through some subsequent processing.
3. The complex cells are, within their receptive fields, invariant to the stimulus position. When different components of the same stimulus stimulate two different complex cells, for example, the relative positions of the pattern's components are not directly reflected in the output of those two cells. Some process that overcomes this apparent loss of relative position information must be identified.

Local and Global Architecture

The visual input is remapped onto the striate cortex in a retinotopic fashion (Tootell, Silverman, Switkes, & De Valois, 1982). Schwartz (1977), among others, has suggested that this GLOBAL striate representation is functionally involved in a size invariant form encoding. The representation of any stimulus on

the striate surface varies nonrigidly as a function of its position on the retina, however, ruling out any direct functional role in form encoding (Cavanagh, 1981, 1982).

Several researchers have documented various LOCALLY organized domains in the striate cortex: orderly strips of ocular dominance (Hubel, Wiesel, & LeVay, 1976), color specificity (Michael, 1981) and orientation (Hubel & Wiesel, 1974). Of most significance is the work of Maffei and Fiorentini (1977), Tootell et al. (1982), and Berardi et al. (1982) who all claim that size (or spatial frequency) and orientation form the two orthogonal dimensions of local representations of the visual stimulus. Although these authors agree on the axes of the representations they do not agree on the orientation of the local transforms with respect to the striate surface. Maffei and Fiorentini (1977) and Berardi et al. (1982), using microelectrode techniques, claim that preferred orientation varies along a direction parallel to the cortical surface and that size or spatial frequency varies along a direction perpendicular to the cortical surface. Tootell et al. (1982) using deoxyglucose labeling techniques claim that both size and orientation axes are parallel to the surface and perpendicular to each other. (Tolhurst & Thompson, 1982, using microelectrode techniques, were unable to discriminate between the two alternatives.)

Independently of how the axes may be oriented with respect to the cortical surface, both research groups are suggesting that the visual information is organized in small local transforms having orthogonal axes of orientation and size. Figure 6.5 depicts how this organization might occur for the transforms described by Maffei and Fiorentini (1977). According to these authors, the preferred spatial frequency of cells first increases as layers II and III of the cortex are traversed and then decreases again crossing layers IV, V, and VI. Only the first of these representations can be seriously considered as a possible basis for the local transform. First, the complex cells having local position invariance predominate in these layers (Hubel & Wiesel, 1962) and second, these layers project to the prestriate cortex and from there to the inferotemporal cortex, possible sites of further processing while the other layers, IV, V and VI, project mainly to subcortical areas (Lund et al., 1975; Spatz et al., 1970).

These local transforms are the essential pieces of the overall transform sequence described in the previous section. Their important properties are as follows.

1. Although the organization within the striate cortex is controversial, the local transforms appear to have axes of size and orientation. The data of Berardi et al. (1982) show that while the orientation axis is clearly well ordered, advancing about 50 degrees for every 250 μm along the direction where spatial frequency is constant, the spatial frequency axis is not as well ordered. The average increase in preferred spatial frequency for a fixed change (250 μm) in position along a penetration for which preferred orientation is constant is about ½ octave,

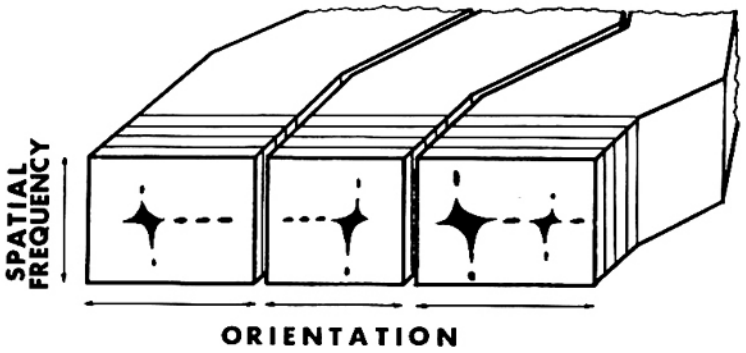


FIG. 6.5. View of a slab of striate cortex showing a cross-section perpendicular to the cortical surface (the top plane) and parallel to the axis of orientation. The cross-sections are broken down into local transforms (shown with exaggerated separations between them here), each sufficiently wide to cover the full range of orientations (approximately 1 mm, Hubel & Wiesel, 1974), and sufficiently deep to include a single, ordered range of spatial frequencies. The retinal area encoded by each local transform changes with cortical position in a retinotopic fashion.

but the actual change is quite variable. It is not possible from their data to determine whether it is valid to describe the size axis as a logarithmic axis as required for the form encoding discussed here. Psychophysical data do imply that the axis is logarithmic (Cavanagh, 1978a). Campbell, Nachmias and Jukes (1970), for example, report that the just discriminable change of a grating's spatial frequency is a fixed percentage (3%) of its current value, classic evidence of an underlying logarithmic scale.

2. Assuming the size axis is logarithmic, the stimulus encoding (considering only the complex cells) will respond to any size and orientation changes of the stimulus within their receptive area by simple shifts in the pattern encoded in the local transform.

3. Each local transform encodes the visual stimulus from a given spatial area. Because receptive field size is generally inversely related to preferred spatial frequency (Movshon, Thompson, & Tolhurst, 1978b), the receptive fields of the cells responding to large features extend beyond the area covered by the receptive fields of cells preferring high spatial frequencies. The local transforms can therefore not be considered individually as encodings of a given spatial region, they are only meaningful when all are considered together. Some integration process is essential for any useful information to be gained from the local striate representation.

4. The range of spatial frequencies covered in each local transform decreases with eccentricity (Berkley, Kitterle, & Watkins, 1975; Movshon et al., 1978c). This spatial inhomogeneity is similar in effect to vignetting in photography and limits the high-frequency content in the periphery. This is a characteristic of the visual system but does not impose any fundamental limitations on the form encoding process.

In sum, the local transforms have the necessary axes of size and orientation although evidence that the size axis is logarithmic is indirect. For the partial representations in the local transforms to be useful they must be integrated in some further processing step.

Integration

The striate representations project to the prestriate cortex from layers II through III while the lower layers project to the subcortical areas (Lund et al., 1975; Spatz et al., 1970). The local arrangement of size, orientation and position information appears to be similar in the prestriate (area 18) and striate cortices—a retinotopic organization with local transforms having size and orientation axes (Berardi et al., 1982). Multiple representations appear in area 19 of the prestriate cortex (Zeki, 1978), but little is known of their local organization.

At the level of the inferotemporal cortex, the receptive fields of the cells are extremely large, up to 90 by 90 degrees, always include the fovea and typically extend into both visual hemifields (Gross, 1973). These cells must be receiving input from several cells in the striate cortex, integrating their responses across the various visual fields involved. The inferotemporal cortex is certainly a candidate area for the integration process necessary for the pattern transform sequence described here.

Several studies have shown that the inferotemporal cortex plays an important role in form perception. Lesions in this area have been shown to directly affect pattern vision while not producing any sensory loss (Mishkin, 1972). Gross and Mishkin (1977) have claimed that the inferotemporal cortex is the basis of the ability to recognize a stimulus independently of its position. Sato, Kawamura, and Iwai (1980) have recorded from some cells in the inferotemporal cortex that respond selectively to given shapes independently of their size, color or brightness. Perrett, Rolls, and Caan (1982) have reported cells in inferotemporal cortex responding to face stimuli independently of orientation, color or size. Several authors have suggested that the inferotemporal cortex is the site of a categorization process wherein perceptual information becomes symbolically coded for general purpose manipulation (Dean, 1982; Wilson & DeBauche, 1981). Other studies have implicated attentional factors in the role of the inferotemporal cortex (Gross, Bender, & Gerstein, 1979).

Desimone, Schwartz, Albright and Gross (1982) attempted to show that the cells in the inferotemporal cortex were selective to the texture of an object's contour, an aspect related to a boundary-specific, size-invariant scheme used in pattern recognition (Zahn & Roskies, 1972). Pollen et al. (1983) have found in the posterior area of IT that cells responded very much like complex cells of area 17 but their receptive fields were much larger. Orientation and spatial frequency tuning were similar to that of area 17 and no differences in orientation or size

tuning could be found across the subregions of the receptive fields. They concluded that the cells in this area of the IT cortex may simply be integrating over several receptive fields of area 17 neurons.

Pollen et al.'s (1983) data imply as well that the summation across individual receptive fields may not maintain phase coherence across fields. In particular, the selectivity to spatial frequency and orientation was no narrower in the inferotemporal cortex than in area 17; if the summation had been performed coherently, the tuning functions would have been considerably sharpened. To some extent it appears that the outputs of the individual area 17 cells may simply be averaged independently of the position of the spatial elements within each subfield. This phase-independent integration throws away a good deal of pattern information and the relevant information must be retained by some other process for functional pattern encoding to occur. The possible solutions to this problem are discussed in detail in the following sections.

The cells studied by Pollen et al. (1983) may or may not represent the required integration of area 17 local information. If the cells studied by Pollen et al. were, in fact, participating in an integration process, the overall organization of the size and orientation axes would have to be retained as well, now becoming the global structure, rather than the local as was the case in area 17. Unfortunately, nothing is known about the local organization of cells in the inferotemporal cortex, nor of their global organization, other than the fact that the region is the only visual structure which is not retinotopically organized (Gross, 1973).

Following an integration step, a position-independent GLOBAL log polar transform would be available as in Fig. 6.4, third column. (Note that the higher spatial frequencies could only be integrated over progressively smaller areas around the fovea as few high-frequency detectors exist in the periphery, Movshon et al., 1978c.) As described previously, size and rotation changes of an object simply shift its representation in this log polar frequency plane without affecting its shape.

A final step of a further position-independent transform (where position now reflects the size and orientation of the input) is then necessary to derive a truly size invariant, form specific representation. Because the inferotemporal cortex is the only nonretinotopically organized visual cortex, speculation concerning the location of these global encodings is limited to this area as well. (Note that rather than assuming separate and sequential steps of integration and final transformation, the combination of the two operations into a single step could be considered as an alternative.)

A variety of transforms could be suggested for this final position invariant step. For computational convenience in the original demonstration (Fig. 6.4, fourth column), it was assumed to be a Fourier amplitude transform. Insofar as the visual system seems to be able to compute some type of frequency analysis on the retinal array, it is not unreasonable to assume that it might be able to perform a similar analysis on some higher order representation as well.

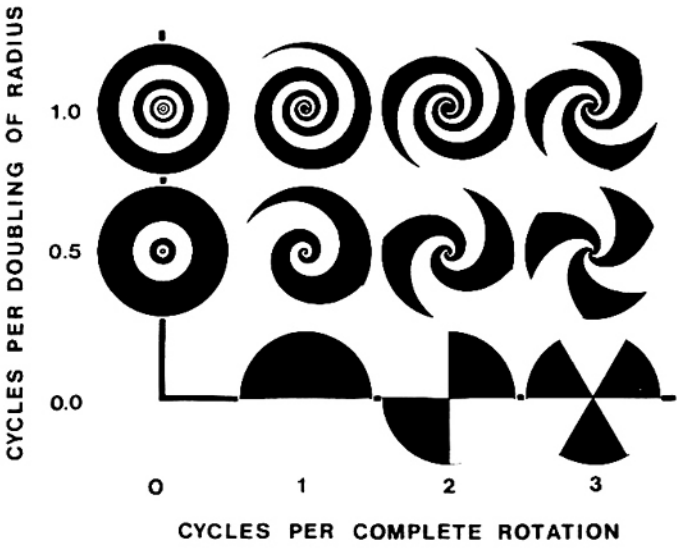


FIG. 6.6. Preferred stimuli of cells at various locations in a Fourier log polar Fourier transform representation. Each cell should respond maximally to its preferred stimulus (or portions thereof) independently of its position in the visual field. "Cycles per complete rotation" reflects the parameter n of Eq. 2 and "cycles per doubling of radius" reflects parameter a. For simplicity, the luminance profiles are shown in black and white and size limited but are actually sinusoidal and of unlimited extent as given in Eq. 2.

Figure 6.6 depicts the stimuli that would optimally stimulate cells at this final level—if the first and the final transforms were both Fourier amplitude transforms. Other transforms at the final level would be sufficient to achieve position invariance and some will undoubtably turn out to be better choices. These stimuli are therefore not predictions in any strict sense but only demonstrations. Note that along one axis the encoded dimension is angle and along the other it is relative size. The maxima along the horizontal axis of Fig. 6.4, fourth column, in fact represent the 90 degree angles present in the stimuli. The stimuli that translate into single points on the final level are the family of logarithmic spirals given by

$$h(r,\theta) = \sin(n\,\theta + a \log r) \qquad (2)$$

where

r is the radius,
θ the polar angle,
n the number of spiral arms, and
a the rate of expansion.

In sum, integration and final transformation would most likely take place in the inferotemporal cortex although the evidence concerning this possibility is incomplete and far from convincing at present. The possibility that integration may be phase incoherent requires that some attention be paid to possible mechanisms that would counteract the resulting information loss.

Loss of Position Information Across Locations

In area 17, the complex cells will respond to a stimulus independently of its position within the receptive field. The precise location of the stimulus is therefore not known. Accentuating this problem is the possibility, just discussed, that the integration of size or spatial frequency information across receptive fields may be effected without respect to the relative locations in the different fields—that is, in a phase incoherent manner. For example, two short, parallel line segments falling on two separate receptive fields would produce the same output from those fields no matter how the two segments line up. If the output of all cells tuned to similar orientations and spatial frequencies are simply summed irrespectively of the relative positions of elements within the various fields (i.e., in a phase incoherent manner), then the information of the relative positions of these two line segments appears to be irretrievably lost.

Although the position information has been lost from these two particular receptive fields, other receptive fields, particularly of lower preferred spatial frequency, will cover parts or all of both line segments. The relative positions of the two line segments will therefore be uniquely encoded by cells that respond to the overall orientation and spatial frequency of the configuration of the two line segments.

This mechanism will provide for the encoding of the relative position information as long as there are receptive fields large enough to cover the separate elements being encoded. This implies a lower bound of approximately 6 degrees separation, in cats, as this is the largest receptive field size reported for the foveal area of cat striate cortex (Movshon, Thompson, & Tolhurst, 1978c, although in area 18 the lowest preferred frequency is about .1 cpd, i.e., a receptive field width of approximately 18 degrees).

Loss of Position Information Between Frequency Components at the Same Location

Stimuli generally have a range of spatial frequency components at each orientation and those frequency components that are sufficiently separated (about $\pm$ 1 octave) will stimulate different complex cells. Because of the loss of position information within each receptive field, no information would be available concerning the RELATIVE locations of the individual components detected by different cells—for example, whether first and third harmonics are in peaks-add or peaks-subtract relative phase. This is a fundamental deficiency of the Fourier amplitude transform used here to model the first position-independent level of encoding. For this transform, no information is available concerning the positions of the individual frequency components, only their degree of presence. If this were also true for the visual system, we should not be able to distinguish between images with the same frequency content but altered phase (position)

content—a single dot and a field of random noise, for example—so this phase or position information concerning frequency components must be encoded in some manner.

Relative Phase Encoding

For shape recognition, the relative phase relations between frequency components as well as the amplitudes of the components are together sufficient to describe the shape. The absolute phase value only determines the position of the shape.

In effect, the Fourier amplitude spectrum used in demonstrations here (Fig. 6.4) throws out all phase information including the intercomponent phase relations and this is more than is necessary to produce position invariance. A transform based on relative phase information would retain position invariance but avoid the ambiguities of the Fourier amplitude encoding.

The use of relative phase information to overcome the loss of position information previously mentioned is in fact a potential function of the physiology of the receptive fields. Simple and complex cells respond to a broad band of frequencies (about ± 1 octave, Maffei & Fiorentini, 1973; Movshon et al., 1978b) and as a result may be sensitive to the relative positions or phases of the frequencies within that band. The argument is clear for simple cells. They respond optimally for a stimulus whose intensity distribution matches the cell's receptive field profile—brightest in the excitatory regions and darkest in the inhibitory regions. The simple cell is therefore not only selective for the spatial frequency content of the stimulus but also for the relative phase offsets of the frequency components within its passband. Both of these types of information are necessary to specify the optimal stimulus. Any variation in receptive field profiles for cells having the same spatial frequency passband indicates that relative phase can be differentiated by these cells. For example, the symmetric and antisymmetric receptive fields (Stromeyer & Klein, 1974) for simple cells would appear to classical electrophysiological techniques as a center strip with two surrounding inhibitory flanks in the symmetric case and as adjacent excitatory and inhibitory strips for the antisymmetric case. Both these organizations are widely reported (Andrews & Pollen, 1979; Movshon et al., 1978a) and a pair of such cells that has similar frequency characteristics would have different relative phase preferences: all components in cosine phase for the symmetrical receptive field, and all components in sine phase for the antisymmetrical field (Fig. 6.7).

Whether or not complex cells are selective for relative phase is not so easily determined. Complex cells have been suggested here as the basis for invariant form encoding because of their insensitivity to the stimulus position but this insensitivity also makes it difficult to directly map effective receptive field sensitivity profiles. Using advanced techniques, Movshon et al. (1978b) and Heggelund (1981) have demonstrated overlapping, antagonistic subfield structures

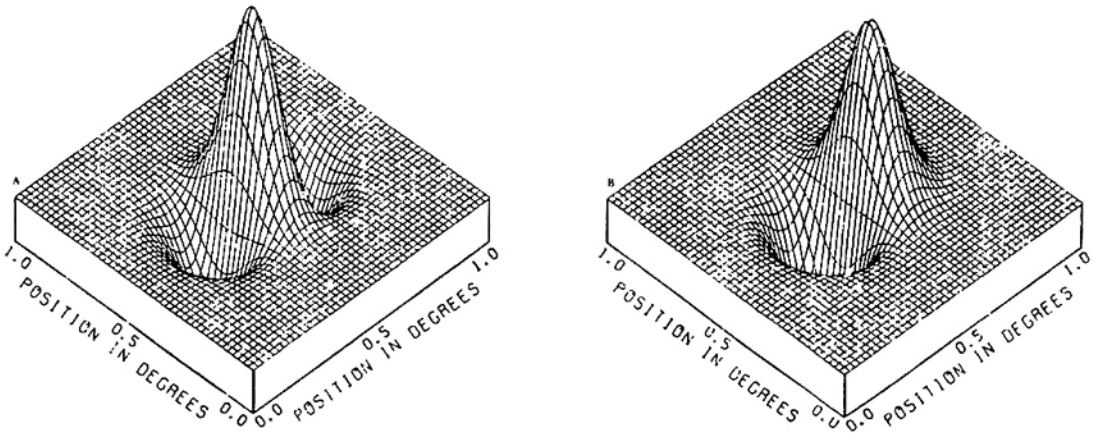


FIG. 6.7. Simple cell receptive field sensitivity profiles. (a) Symmetrical receptive field profile. (b) Antisymmetrical receptive field profile. (Figures courtesy of John Daugman)

for complex cells but it is not currently known whether these cells have relative phase (e.g., symmetric versus antisymmetric) preferences. If complex cells receive their input from sets of simple cells as suggested by Hubel and Wiesel (1962) and Pollen and Ronner (1981), and all the simple cells involved have the SAME relative phase preferences then each complex cell could retain the same relative phase specificity as the simple cells feeding into it. Studies of complex cell response to drifting antisymmetric or symmetric brightness profiles could clarify the situation.

To retain sufficient pattern information, the visual system must use at least two different relative phase sensitive detectors. It might possibly use more than two, which would give the encoding more the flavor of a feature or profile encoding. That is, there could be several profiles being analyzed at each size and orientation.

In addition to the possibility of different phase spectra across the relatively broad spatial frequency bandwidths of cells in the striate cortex, there is also the possibility of multiple band spatial frequency spectra (i.e., multiple narrow peaks rather than one broad one). In particular, cells or channels responding to first and second or first and third harmonics have been described (Cavanagh, 1978b; Cavanagh, Brussell, & Stober, 1981; Glezer, Cooperman, Ivanov, & Tscherbach, 1976; Pollen, Andrews, & Feldon, 1978). All of these alternatives simply point to a variety of effective receptive field profiles of NONSINUSOIDAL shape. Several psychophysical studies have supported the possibility of relative phase specificity in the spatial frequency channels (Arend & Lange, 1979; Burr, 1980; Sansbury, Distelhorst, & Moore, 1978).

A pair of broad bandwidth complex cells can encode relative phase only over a restricted range. However, the overall relative phase information across the spectrum could be uniquely encoded if each range overlaps with those of cells tuned to neighboring frequencies and orientations.

In summary, the fact that the basic encoding elements in the visual system—the receptive field profiles—are far from the sinusoidal form of the Fourier transform implies that relative phase information can be directly encoded. That is, complex cells with a spatial frequency bandwidth of ± 1 octave and with at least two different preferred phase patterns (e.g., symmetric and antisymmetric) would be able to encode the relative phase information that ensures an unambiguous, position independent representation.

The broad bandwidth of the spatial frequency detectors may therefore have a significant functional role. But what particular advantages of the Fourier transform are lost because of this broad bandwidth? The principal loss is the orthogonality of the encoding set. This implies that the activity of a given detector does not ensure that a particular pattern is present in the input. The activity of a given detector implies only that a pattern having components within a certain frequency and orientation range is present. In the visual system, however, the orientation and frequency content of the stimulus may be determined from the distribution of detectors responding. Since a unique encoding is produced for each given input pattern, no ambiguity arises as a result of the loss of orthogonality. Relative phase encoding therefore appears to be a feasible and functional aspect of visual encoding.

Nonlinearities

A basic property of a linear system is that of superposition: The sum of the responses to several stimuli is the same as the response to the sum of the stimuli. In addition to basic cell nonlinearities (threshold and saturation), the position invariance of the complex cell output produces a significant encoding nonlinearity. The responses of a complex cell to a drifting grating of a given frequency is an almost unmodulated increase in the firing rate (i.e., a position invariant "amplitude" response), if the grating is within the passband of the cell. If two different frequency components are present within the frequency and orientation passbands of a cell, the response to the sum of the two components is not necessarily, nor even likely to be, the sum of the individual responses to each component in isolation. It is assumed here (Eq. 2) that the resulting response depends as well on the relative phase of the two gratings, a factor that is not reflected in the output for each grating separately.

The interaction between these components of different frequency (i.e., the departure from additivity) is in fact the relative phase information discussed in the previous section. This information is essential pattern information if the two frequency components arise from a single stimulus to be identified: a letter, a face, etc. If the two components come from two separate patterns that are adjacent or overlapping then their interaction is representative of their relative positions and will interfere with the identification of either one. That is, the

response to two patterns will be the response to each individually plus an interaction term that depends on the relative positions of the two patterns.

This interaction term may produce noise problems that would be especially deleterious if the initial encoding were performed over the entire input field. For example, as previously discussed, the presence of a second pattern would seriously interfere with the identification of the first if identification were based on the Fourier amplitude transform, a transform that encodes the entire input field. This interference would occur independently of where the two patterns were situated, either overlapping or widely sparated. One way to reduce this interaction is in fact to analyze only local areas for amplitude terms and then sum across local areas. If the two patterns are not within the same receptive field then their amplitude terms can be summed linearly. This is a distinct advantage of the local nature of the initial encoding in the striate cortex, if it is assumed, as has been discussed, that the outputs of the complex cells in area 17 are simply summated across the visual field to produce a global transform. The lack of superposition will therefore apply only within the initial receptive field, producing interference effects with similar characteristics (overlap or adjacency) to those seen in masking and hidden figures studies of human vision.

On the other hand, the interaction terms among patterns (the deviations from additivity) can become important if the arrangement of the patterns is meaningful, e.g., the arrangement of the letters in a word. Each of the different possible arrangements of the letters of a word will produce different interaction terms thus distinguishing between POST and SPOT, for example.

Although the nonlinearity can be useful for encoding meaningful combinations of patterns it represents only noise, as mentioned previously, when one pattern is to be identified in a field of distractors. It is important to determine to what extent this noise would impair or prohibit the identification of a stimulus. The simulation of relative phase encoding in the next section permits a rough evaluation of the interference produced.

4. A RELATIVE PHASE SENSITIVE LOG POLAR FREQUENCY ENCODING

This section demonstrates some of the pattern recognition properties of the log polar frequency transform when relative phase information is included. Several simplifications are made for computational convenience and no claim is made that this represents a physiological process. Several properties of transforms that may be used in the visual system may be extrapolated from this simulation nonetheless.

The advantage of the relative phase information inherent in the broad bandwidth complex cell output is the discrimination of patterns sharing the same frequency components but differing in relative phase relations. To simplify the

calculations, only a single relative phase sensitive profile—antisymmetrical—is used. Without the complementary symmetric information, phase ambiguities still exist: a pattern and its black/white reversal cannot be distinguished, for example. However, antisymmetrical detectors that respond to the same frequency components but at orientations 180 degrees apart can differentiate left-facing versus right-facing combinations of the same frequency components and thus a pattern versus its 180 degree rotation, a distinction that is impossible in the Fourier amplitude representation.

As an additional simplification, the possibility of interactions between frequencies within the passband is represented by using only two spatial frequencies rather than a broad band of frequencies. A comparison of the frequency and spatial characteristics of actual and simulated cells is shown in Fig. 6.8. The sensitivity profile is basically that of an antisymmetrical detector encoding the first and second harmonics. The encoded dimensions are thus frequency (periodicity of the sensitivity profile) and orientation (from 0 to 360 degrees).

The output of each detector is given by the integral of the product of the input brightness profile by the detector sensitivity profile (Eq. 3) at the position for which this integral is maximum. This is taken to be a realistic representation of the position invariant aspect of the output of a complex cell (Heggelund, 1981; Pollen & Ronner, 1982). For a detector of frequency, f, at orientation, θ, where

$$f_x = f \cos\theta$$
$$f_y = f \sin\theta$$

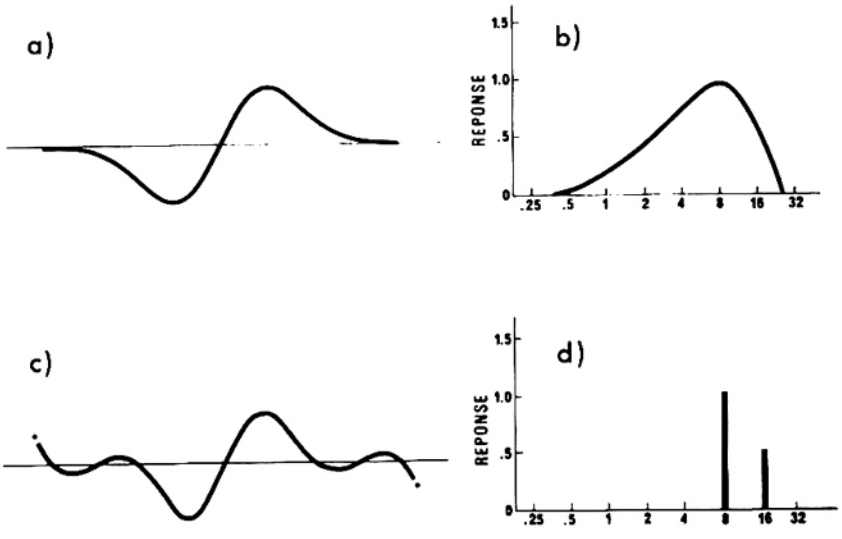


FIG. 6.8. (a) The antisymmetrical receptive field profile for simple cells assumed to be feeding into the complex cells. (b) The spatial frequency response curve for this cell. (c) The sensitivity profile for the simulated antisymmetrical detector. (d) The spatial frequency response curve of the simulated detector.

and an input intensity profile S(x,y), the integral G(f, θ, γ) represents the response of a particular subfield at position, or phase angle γ.

$$G(f,\theta,\gamma) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} S(x,y)\cdot\{\sin[2\pi(f_x\ x + f_y y) + \gamma] + .5\ \sin\ [4\pi(f_x x + f_y y) + 2\gamma + \pi/2]\}dxdy \quad (3)$$

The overall response, H(f,θ), is taken as that of the subfield giving the maximum output, that is the maximum value of the integral over the range 0 to 2 π for γ

$$H(f,\theta) = \text{MAX}\left\{G(f,\theta,\gamma)\Big|_{\gamma=0}^{\gamma=2\pi}\right\} \text{ if } H(f,\theta) > H(f,\theta + \pi)$$

$$= 0 \qquad \text{otherwise} \quad (4)$$

An additional nonlinearity (Eq. 4) is added to accentuate the relative phase information. Klein and Stromeyer (1980) and Cavanagh, Brussell and Stober (1981) have reported suggestive evidence that detectors of different phase sensitivities but similar frequency and orientation preferences are mutually inhibitory in order to enhance dominant local features. In this simulation then, a given detector's output will be its maximum integrated product with the input if this value is greater than the corresponding value for its relative phase pair (the detector at the same frequency but with a 180 degree difference in orientation), and zero otherwise. Discrimination of the orientation of the test letters was in fact poor without this mutual inhibition.

The transform is identical in all other respects to that described for Fig. 6.4.

Classification of the tests is again based on the correlations of their final amplitude transforms with the final transforms of the prototypes: an F, an E, and a P. The discrimination of the orientation of the tests is based on the crosscorrelation of the intermediate transforms (Fig. 6.9, second column) with the intermediate transforms of the prototypes. Since at the intermediate level, the orientation of the test is reflected by the shifting left or right of the encoded pattern, the position of maximum correlation should reflect the test orientation directly. If the intermediate encoding had been based on a simple amplitude transform, as was the case in Fig. 6.4, then there would be no difference between encoded patterns for 180 degree rotations of the stimulus—note the repetition of the encoded pattern from 0 to 180 degrees in the 180 to 360 degree range for all log polar frequency transforms of Fig. 6.4.

It can be seen in Fig. 6.9 that this 180 degree ambiguity is gone, the differences greatly accentuated by the suppression of the weaker of each pair of components differing by 180 degrees.

Each orientation of the input thus produces a unique encoding at this level and the orientation can be derived by sliding the prototype pattern over the test until the maximum correlation is obtained. Note that because the orientation axis is

TABLE 6.2.
Correlations between the final relative phase sensitive representations of E, P and F at 15 minutes of visual angle and 0 degrees orientation and the final transfoms of the variously oriented F's and the double F presentation.

	F	⊓	ⅎ	⊓F
F	1.00	.87	.91	.71
E	.79	.73	.73	.63
P	.77	.75	.73	.63

discontinuous in a log polar representation, the maximum obtainable correlation will decrease as more of the pattern drops off one border and returns on the other.

This crosscorrelation measure is used as a demonstration only. The operation was performed directly on the final transform level by taking the product of the input final transform and the prototype final transform (both real and imaginary values, not simply the amplitude component) and inverse Fourier transforming it. This will produce a peak at the coordinates of the maximum correlation and the orientation and size (x and y coordinates of the intermediate transform) can be read directly from this. This is not to imply that the brain might proceed in the same way. If the brain were to encode size and orientation as phase information at the final transform level the optimum method for decoding size and orientation would depend on how this information was to be used.

Table 6.2 shows that all inputs were correctly classified, the discrimination between F and E being even greater than in the amplitude transform examples of Fig. 6.4. This improvement brings the quality of this recognition operation—simple correlation between the final transforms—close to the optimum obtainable by correlating the original letters, normalized for size, orientation and position in the space domain. Significantly, the inclusion of a second F in the input did decrease the correlation with the prototype as a result of nonlinearities (amplitude and inhibition) in the input encoding (Eq. 2). Classification remained clearcut nevertheless. A more extensive evaluation might point out more serious instances of interitem interference but recall that the examples here are under the worst case conditions of a single input field rather than a mosaic of individual receptive fields. Interference from nonlinearities in the encoding discussed previously would be greatly reduced by local nonlinear encoding followed by linear summation across local regions.

In addition, the input orientation was successfully decoded in every case, the crosscorrelation maxima occurred approximately at 0, 90, and 180 degrees. No ambiguity was found for 180 rotations of the stimuli.

In sum, the inclusion of relative phase information, as demonstrated in a limited way here, not only renders the modeling physiologically more realistic but also improves the pattern recognition capabilities of the encoding.

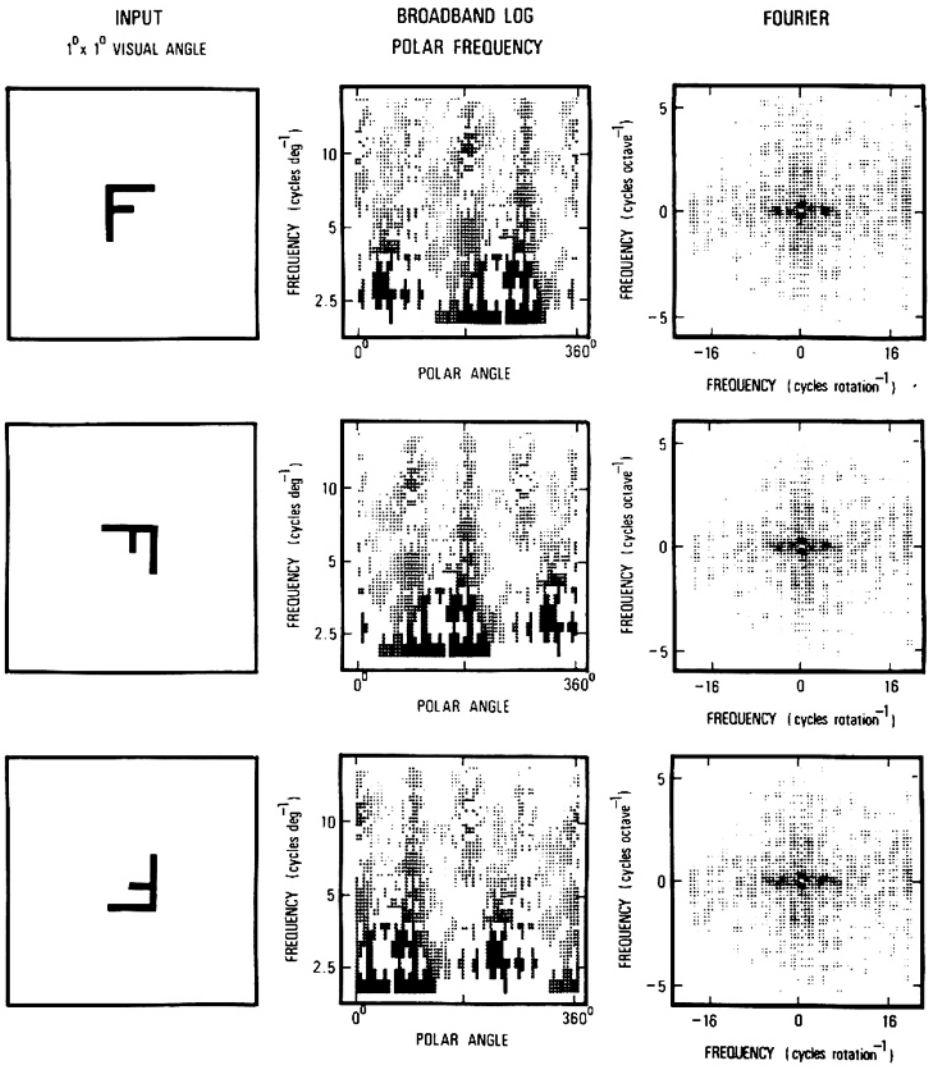
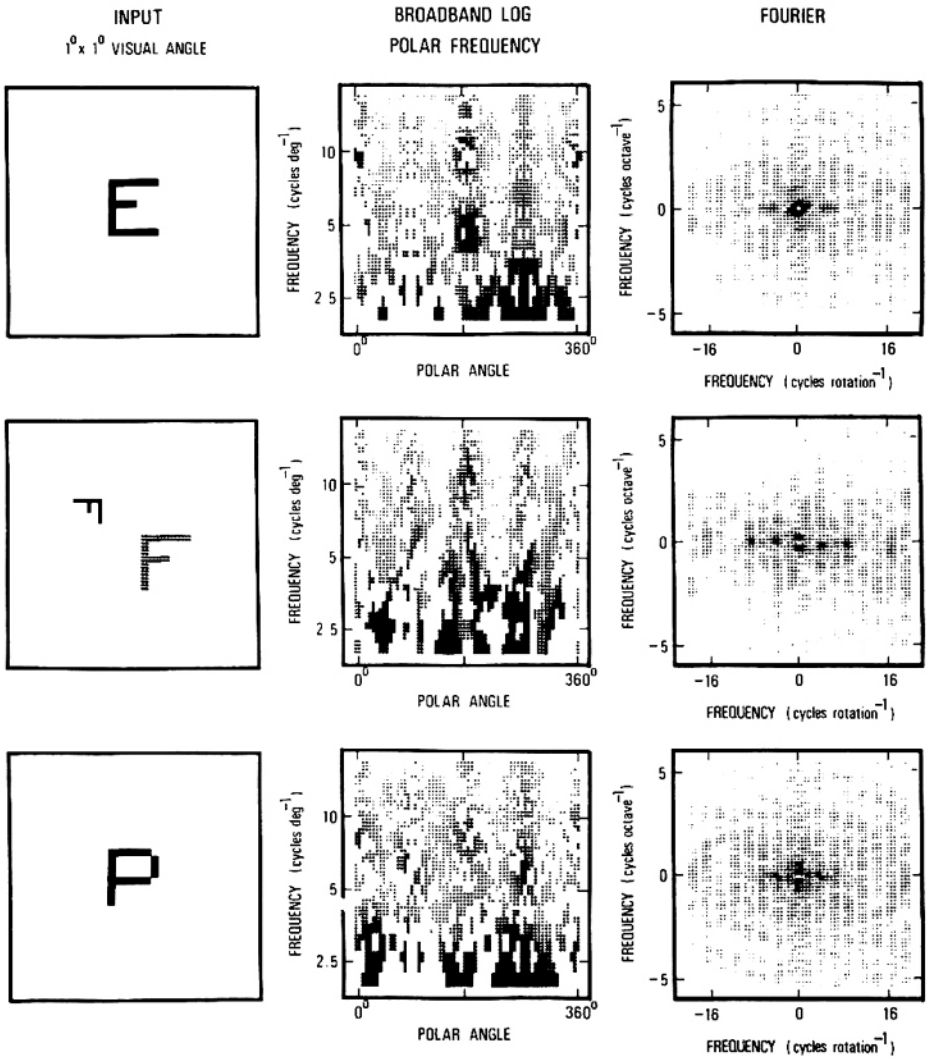


FIG. 6.9. A relative phase sensitive size and position invariant encoding sequence. The letter F is presented at three different orientations (0, 90, and 180 degrees) as well as presented twice within the same field at 0 and 90 degrees and 15 and 7.5 minutes of visual angle respectively. Opposite, an E and a P are also presented for comparison. Each input is shown at two stages of transformation: a broadband log polar frequency representation that is sensitive to relative phase and the Fourier amplitude transform of this representation. The images are presented on matrices of 64 × 64 resolution with amplitude being represented in 30 equal intervals between the minimum and maximum of each image by the density of dots. The final Fourier amplitude transform is performed on the log polar representation after it is reduced to 32 × 32 resolution and zero mean amplitude with the rest of the 64 × 64 array set to zero. This minimizes aliasing and aperture effects (Goodman, 1968).



5. TEMPLATES

The end result of the encoding sequence proposed here is a form-specific representation that is independent of size and position (and, to some extent, orientation). This permits the classification of stimuli using previously stored templates—final transforms of prototype patterns: letters, familiar faces, common forms and symbols, and familiar words might be a few examples. The input transform would have to be matched in parallel against all relevant prototypes for this approach to achieve significant efficiency. Such parallel matching processes have been described in the literature (Anderson, Silverstein, Ritz, & Jones, 1977; Cavanagh, 1975, 1976; Eich, 1982; Kohonen, 1977; Murdoch, 1982).

Earlier work with the confusability of letters (Cavanagh, 1983) has shown that interletter similarity is well predicted by such correlational matches. Several other studies have claimed to predict significant proportions of the letter confusion variance using feature (Keren & Baggen, 1981), cluster (Shepard & Arabie,

1979), or choice models (Keren & Baggen, 1981) but these models use several free parameters (from 29 to 350). Thirty arbitrary parameters are probably sufficient to predict 95% of the variance of the order of the interletter confusions. With no free parameters at all, a correlational model can predict 78% of the available variance. In addition, the analysis showed that no structural features (e.g., closure, symmetry, curvature) could explain any significant proportion of the variance in the confusion data beyond that already attributable to the correlation match predictions.

The data analyzed (Cavanagh, 1983) showed convincingly that the observed interletter confusions are best explained by a position independent encoding (for example, I and J are as confusable as I and T even though the center stroke is overlapping only for I and T, indicating a position independent process) but the available data could not test the size or rotation properties of the transform sequence proposed here as the size and orientation of the letters were never varied.

These findings indicate that for situations where the form of the tests is known (the letter confusion studies used a fixed font) an analogue match operation is a reasonable model for identification and recognition. For situations where the test is in an unpredictable form—handwriting, for example—it is likely that a structural or topological representation would be more appropriate.

For analyzing real world scenes involving shadows, partially hidden objects and objects recognized by function as opposed to shape (e.g., chairs), a more sophisticated analysis than that available from simple template matches—even if size and position invariant—is certainly essential. It is unlikely that the visual system would switch back and forth from one mode to the other depending on the situation but it is not unreasonable to assume that a more sophisticated structural analysis could take as its base data, pattern elements identified by a transformational encoding. Rather than having to encode patterns and scenes as structures of simple lines and angles, the structural encoding could start after the analogue encoding had matched all identifiable elements in the scene—geometrical shapes, letters, familiar objects—all elements for which stored representations were available. When the scene is totally unfamiliar with no previously stored shapes of any sort present, the primitives are simply reduced to the lines and angles extracted by the receptive field profiles of the initial encoding level.

The analogue and structural approaches to pattern recognition may therefore be simply two levels of a more complex process. Stored prototypes could provide a rich, high-level set of size and position invariant primitives to serve as the basis set for an intelligent structural analysis. Tasks that require identification or classification of familiar patterns should reveal principally the properties of the analogue, transformational encoding. Tasks requiring an identification of unfamiliar representations of known objects or understanding of novel scenes should reveal the properties of the structural levels of analysis. Examples of high level structural encodings have been given by Marr (1982), for example, in the

cylinder representations of animal shapes. He assumed that the basic body parts are identified by contour extraction rather than by template extraction, a far simpler procedure.

It remains to be seen whether a viable model integrating a high level set of template extracted patterns with a structural analysis can be constructed. It would appear, however, that there is good evidence for both levels of processing.

6. CONCLUSIONS

An encoding transform has been described that obtains a size invariant representation of form. The first of two essential steps in the sequence is a position invariant encoding of the input that arrays the pattern information along axes of orientation and log size. A log polar Fourier amplitude transform was used to demonstrate this level. A representation of the stimulus in this transform will shift along the orientation axis for rotations of the stimulus and will shift along the log size axis for size changes. The encoded pattern itself is unchanged except for these shifts. Because of the position invariance, the encoding is affected neither by the location of the stimulus nor by the centre of its rotation or size change.

The second step is a position independent encoding of the first representation. The shifts caused by size and orientation changes are now ignored and the representation is a true, size invariant, form specific encoding. A Fourier amplitude transform was used to demonstrate this stage.

The possibility of a size and position invariant representation makes the use of correlational memories a realistic proposition. There are, however, a number of drawbacks in the simplified sequence presented here for demonstrating the principles of the encoding. First, the Fourier amplitude transform is not well suited for pattern encoding because of its phase ambiguity. Second, cortical cells only respond to small, local areas of the visual field rather than the entire visual field as does the Fourier amplitude transform. Third, the amplitude transform has significant encoding nonlinearities due specifically to its position invariance property.

In considering the physiology of the striate cortex to determine a more realistic representation, it is noted that the complex cells do provide a local, position independent encoding response to size and orientation and that, according to Maffei and Fiorentini (1977), Berardi et al. (1982) and Tootell et al. (1982), cells in the visual cortex are organized locally with orthogonal axes of size and orientation. It is also noted that the cells respond to a broad range of spatial frequencies and thus are sensitive to the relative phase content of the stimuli. This information is sufficient to remove the ambiguities that accompany the Fourier amplitude encoding if there are at least two different relative phase sensitivity profiles for the complex cells. The complex cells are, as well, located

principally in the striate layers that project to the prestriate and then to the inferotemporal cortices. Because of its nonretinotopic organization, the inferotemporal cortex is suggested as the site for integrating across receptive fields and for the final, position invariant transform to obtain the size invariance. Thus the various shortcomings of the Fourier encoding are overcome: the relative phase sensitivity of the striate cells removes any phase ambiguity; the summation across receptive fields implies that although the encoding at each receptive field is still nonlinear due to position invariance (the response for two components is not the sum of the individual responses but also reflects intercomponent spacing), it is only locally nonlinear. The response to components falling on separate receptive fields is linear assuming linear summation. The properties of the cortical encoding therefore appear to provide the potential for an encoding of significantly greater utility than that described in the original demonstration.

It remains to be seen whether there is, in fact, a range of phase sensitivities for complex cells as required by this model. If all complex cells show similar phase spectra then the encoding is again troubled by phase ambiguities and the suggestion of Robson (1980) that the spatial frequency encoding serves a texture segmentation process becomes the more plausible model.

If the brain is actually using an encoding sequence similar to that described here, it must be remembered that the results of such an encoding would probably be integrated as the base data into a structural analysis of the stimulus. Only in a task requiring direct recognition of familiar objects—letters in a known font familiar faces, etc.—will the results of the size and position invariant encoding be the final level of analysis.

In conclusion, a physiologically plausible transform sequence capable of producing a size and position invariant encoding has been described. Whether or not the visual system makes use of this process appears to be testable in a number of ways but a final decision is not possible until the structure and function of the inferotemporal cortex is better understood. This encoding process would probably operate as one of several parallel analyses of the visual input. Information from color, depth and motion channels, as well as the brightness-based form encoding described here, would all flow into higher order structural analyses in order to build an overall representation of the visual input.

ACKNOWLEDGMENTS

This research was supported by NSERC grant A8606 and by the Ministère d'Education du Québec. The helpful comments of Dan Pollen and the technical assistance of John Boeglin, Claude Charbonneau and Pierre Poirier are gratefully acknowledged.

REFERENCES

Albrecht, D. G., De Valois, R. L., & Thorell, L. G. (1980). Visual cortical neurons: Are bars or gratings the optimal stimuli? *Science, 207*, 88–90.

Anderson, J. A., Silverstein, J. W., Ritz, S. A., & Jones, R. S. (1977). Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological Review, 84,* 413–451.

Andrews, B. W., & Pollen, D. A. (1979). Relationship between spatial frequency selectivity and receptive field profile of simple cells. *Journal of Physiology, 287,* 163–176.

Arend, L. E., & Lange, R. V. (1979). Phase-dependent interaction of widely separated spatial frequencies in pattern discrimination. *Vision Research, 19,* 1089–1092.

Barlow, H. B., Narasimhan, R., & Rosenfeld, A. (1972). Visual pattern analysis in machines and man. *Science, 177,* 567–575.

Berardi, N., Bisti, S., Cattaneo, A., Fiorentini, A., & Maffei, L. (1982). Correlation between preferred orientation and spatial frequency of neurones in visual areas 17 and 18 of the cat. *Journal of Physiology, 323,* 603–618.

Berkley, M. A., Kitterle, F., & Watkins, D. W. (1975). Grating visibility as a function of orientation and retinal eccentricity. *Vision Research, 15,* 239–244.

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

Brousil, J. K., & Smith, D. R. (1967). A threshold logic network for shape invariance. *I E E E Transactions on Computers, EC-16,* 818–828.

Burr, D. C. (1980). Sensitivity to spatial phase. *Vision Research, 20,* 391–396.

Burr, D., Morrone, C., & Maffei, L. (1981). Intra-cortical inhibition prevents simple cells from responding to textured visual patterns. *Experimental Brain Research, 43,* 455–458.

Campbell, F. W., Nachmias, J., & Jukes, J. (1970). Spatial frequency discrimination in human vision. *Journal of the Optical Society of America, 60,* 555–559.

Campbell, F. W., & Robson, J. G. (1968). Application of Fourier analysis to the visibility of gratings. *Journal of Physiology, 197,* 551–566.

Casasent, D., & Psaltis, D. (1976). Position, rotation, and scale invariant optical correlations. *Applied Optics, 15,* 1793–1799.

Cavanagh, P. (1974). A two dimensional position, size, and rotation invariant pattern transform: An electro-optical process and a neural analogue. *Technical report,* Département de Psychologie, Université de Montréal.

Cavanagh, P. (1975). Two classes of holographic processes realizable in the neural realm. In T. Storer & D. Winter (Eds.), *Formal aspects of cognitive processes.* Berlin: Springer-Verlag, 14–40.

Cavanagh, P. (1976). Holographic and trace strength models of rehearsal effects in the item recognition task. *Memory & Cognition, 4,* 186–199.

Cavanagh, P. (1978a). Size and position invariance in the visual system. *Perception, 7,* 167–177.

Cavanagh, P. (1978b). Subharmonics in adaptation to sine wave gratings. *Vision Research, 18,* 741–742.

Cavanagh, P. (1981). Size invariance: Reply to Schwartz. *Perception, 10,* 469–474.

Cavanagh, P. (1982). Functional size invariance is not provided by the cortical magnification factor. *Vision Research, 22,* 1409–1412.

Cavanagh, P. (1983). Visual encoding processes revealed by alphabetic confusions. In preparation.

Cavanagh, P., Brussell, E. M., & Stober, S. R. (1981). Evidence against independent processing of black and white pattern features. *Perception & Psychophysics, 29,* 423–428.

Chaikin, G., & Weiman, C. (1979). Logarithmic spiral grids for image processing. Proceedings of the *IEEE Computer Society Conference on Pattern Recognition and Image Processing,* 25–31.

Cunningham, J. P. (1980). Trees as memory representations for simple visual patterns. *Memory & Cognition, 8,* 593–605.

Cutting, J. E. (1983). Four assumptions about invariance in perception. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 310–317.

Dean, P. (1982). Visual behavior in monkeys with inferotemporal lesions. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behavior*. Cambridge, Mass.: MIT Press, 587–628.

Desimone, R., Schwartz, E. L., Albright, T. D., & Gross, C. G. (1982). Inferior temporal neurons selective for stimulus shape. *Supplement to Investigative Ophthalmology and Visual Science, 22*, 238.

Dodwell, P. C. (1982). Geometrical approaches to visual processing. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behavior*. Cambridge, Mass.: MIT Press, 801–825.

Duda, R. O., & Hart, P. E. (1973). *Pattern classification and scene analysis*. New York: Wiley.

Eich, J. M. (1982). A composite holographic associative recall model. *Psychological Review, 89*, 609–626.

Foster, D. H. (1977). Visual pattern recognition by assignment of invariant features and feature-relations. *Optica Acta, 24*, 147–157.

Foster, D. H., & Mason, R. J. (1979). Transformation and relational-structure schemes for visual pattern recognition. *Biological Cybernetics, 32*, 85–93.

Glezer, V. D., & Cooperman, A. M. (1977). Local spectral analysis in the visual cortex. *Biological Cybernetics, 28*, 101–108.

Glezer, V. D., Cooperman, A. M., Ivanov, A., Tscherbach, T. A. (1976). An investigation of the spatial frequency characteristics of the complex fields of the visual cortex of the cat. *Vision Research, 16*, 789–797.

Glezer, V. D., Tsherbach, T. A., Gauselman, V. E., & Bondarko, V. M. (1980). Linear and non-linear properties of simple and complex receptive fields in area 17 of the cat visual cortex. *Biological Cynbernetics, 37*, 195–208.

Goodman, J. W. (1968). *Introduction to Fourier optics*. New York: McGraw-Hill.

Gross, C. G. (1973). Visual function of inferotemporal cortex. In R. Jung (Ed.), *Handbook of sensory physiology Vol VIII/3 Part B*. Berlin: Springer-Verlag, 451–482.

Gross, C. G., Bender, D. B., & Gerstein, G. L. (1979). Activity of inferior temporal neurons in behaving monkeys. *Neuropsychologica, 17*, 215–229.

Gross, C. G., & Mishkin, M. (1977). The neural basis of stimulus equivalence across retinal translation. In S. Harnad et al. (Eds.), *Lateralization in the nervous system*. New York: Academic Press, 109–122.

Heggelund, P. (1981). Receptive field organization of complex cells in cat striate cortex. *Experimental Brain Research, 42*, 99–107.

Hu, M.-K. (1962). Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory, IT-8*, 179–187.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology, 160*, 106–154.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology, 195*, 215–243.

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

Hubel, D. H., Wiesel, T. H., & LeVay, S. (1976). Functional architecture of area 17 in normal and monocularly deprived macaque monkeys. *Cold Spring Harbor Symposium on Quantitative Biology, 40*, 581–589.

Jacobson, L., & Wechsler, H. (1982). A new paradigm for computational vision based on the Wigner distribution. *Technical Report*, Electrical Engineering Department, University of Minnesota, Minneapolis.

Julesz, B. (1981). A theory of preattentive texture discrimination based on first-order statistics of textons. *Biological Cybernetics, 41*, 131–138.

Kabrisky, M., Hall, C. F., Goble, L., & Gill, R. A. (1971). Realization of a data independent

pattern recognition system based on human physiology. *IEEE Systems, Man and Cybernetics Annual Symposium Record*, 233–240.

Keren, G., & Baggen, S. (1981). Recognition models of alphanumeric characters. *Perception & Psychophysics, 29*, 234–246.

Klein, S., & Stromeyer, C. F., III (1980). On inhibition between spatial frequency channels: adaptation to complex gratings. *Vision Research, 20*, 459–466.

Kohonen, T. (1977). *Associative memory*. Berlin: Springer-Verlag.

Kulikowski, J. J., Marcelja, S., & Bishop, P. O. (1982). Theory of spatial position and spatial frequency relations in the receptive fields of simple cells in the visual cortex. *Biological Cybernetics, 43*, 187–198.

Lund, J. S., Lund, R. D., Hendrickson, A. E., Bunt, A. H., & Fuchs, A. F. (1975). The origin of efferent pathways from the primary visual cortex, area 17, of the macaque monkey as shown by retrograde transport of horseradish peroxidase. *Journal of Comparative Neurology, 164*, 287–304.

Maffei, L., & Fiorentini, A. (1973). The visual cortex as a spatial frequency analyser. *Vision Research, 13*, 1255–1267.

Maffei, L., & Fiorentini, A. (1977). Spatial frequency rows in the striate visual cortex. *Vision Research, 17*, 257–264.

Marcelja, S. (1980). Mathematical description of the responses of simple cortical cells. *Journal of the Optical Society of America, 70*, 1297–1300.

Marko, H. (1973). Space distortion and decomposition theory: A new approach to pattern recognition by vision. *Kybernetik, 13*, 132–143.

Marr, D. (1982). *Vision*. San Francisco: W. H. Freeman.

Michael, C. R. (1981). *Journal of Neurophysiology, 46*, 587–604.

Milner, P. M. (1974). A model for visual shape recognition. *Psychological Research, 81*, 521–535.

Mishkin, M. (1972). Cortical visual areas and their interactions. In A. G. Karczmar & J. C. Eccles (Eds.), *Brain and human behavior*. New York: Springer-Verlag, 187–208.

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

Movshon, J. A., Thompson, I. D., & Tolhurst, D. J. (1978b). Receptive field organization of complex cells in the cat's striate cortex. *Journal of Physiology, 283*, 101–120.

Movshon, J. A., Thompson, I. D., & Tolhurst, D. J. (1978c). Spatial and temporal contrast sensitivity of neurones in areas 17 and 18 of the cat's visual cortex. *Journal of Physiology, 283*, 101–120.

Murdoch, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review, 89*, 609–626.

Oden, G. C. (1979). A fuzzy logical model of letter identification. *Journal of Experimental Psychology: Human Perception and Performance, 5*, 336–352.

Perrett, D. I., Rolls, E. T., & Caan, W. (1982). Visual neurones responsive to faces in the monkey inferotemporal cortex. *Experimental Brain Research, 47*, 329–342.

Pollen, D. A., Andrews, B. W., & Feldon, S. E. (1978). Spatial frequency selectivity of periodic complex cells in the visual cortex of the cat. *Vision Research, 18*, 665–697.

Pollen, D. A., Lee, J. R., & Taylor, J. H. (1971). How does the striate cortex begin the reconstruction of the visual world? *Science, 173*, 74–77.

Pollen, D. A., Nagler, M., Daugman, J., Kronauer, R., & Cavanagh, P. (1983). Use of Gabor elementary functions to probe receptive field substructure of posterior inferotemporal neurons in the owl monkey. *Vision Research*, in press.

Pollen, D. A., & Ronner, S. F. (1981). Phase relationships between adjacent simple and complex cells in the visual cortex. *Science, 212*, 1409–1411.

Pollen, D. A., & Ronner, S. F. (1982). Spatial computation performed by simple and complex cells in the visual cortex of the cat. *Vision Research, 22,* 101–118.

Robson, J. (1975). Receptive fields: Neural representation of the spatial and intensive attributes of the visual image. In E. D. Carterette & M. D. Friedman (Eds.), *Handbook of perception. Vol. 5: Seeing.* New York: Academic Press, 81–117.

Robson, J. (1980). Neural images: The physiological basis of spatial vision. In C. S. Harris (Ed.), *Visual coding and adaptability.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 177–214.

Sansbury, R. V., Distelhorst, J., & Moore, S. (1978). A phase-specific adaptation effect of the square wave grating. *Investigative Ophthalmology and Visual Science, 17,* 442–448.

Sato, T., Kawamura, T., & Iwai, E. (1980). Responsiveness of inferotemporal single units to visual pattern stimuli in monkeys performing discrimination. *Experimental Brain Research, 38,* 313–319.

Schade, O. H. (1956). Optical and photoelectric analog of the eye. *Journal of the Optical Society of America, 46,* 721–739.

Schwartz, E. L. (1977). Spatial mapping in the primate visual cortex: Analytic structure and relevance to perception. *Biological Cybernetics, 25,* 181–194.

Schwartz, E. L. (1980). Computational anatomy and functional architecture of striate cortex: A spatial mapping approach to perceptual coding. *Vision Research, 20,* 645–669.

Shapiro, L. (1980). A structural model of shape. IEEE Transactions on *Pattern Analysis and Machine Intelligence, PAMI-2,* 111–126.

Shepard, R. N., & Arabie, P. (1979). Additive clustering: Representation of similarities as combinations of discrete overlapping properties. *Psychological Review, 86,* 87–123.

Spatz, W. B., Tigges, J., & Tigges, M. (1970). Subcortical projections, cortical associations and some intrinsic interlaminar connections of the striate cortex in the squirrel monkey (Saimiri). *Journal of Comparative Neurology, 140,* 155–174.

Stone, J., Dreher, B., & Leventhal, A. (1979). Hierarchical and parallel mechanisms in the organization of visual cortex. *Brain Research Reviews, 1,* 345–394.

Stromeyer, C. F. III, & Klein, S. (1974). Spatial frequency channels in human vision as asymmetric (edge) mechanisms. *Vision Research, 14,* 1409–1420.

Tolhurst, D. J., Movshon, J. A., & Thompson, I. D. (1981). The dependence of response amplitude and variance of cat visual cortical neurones on stimulus contrast. *Experimental Brain Research, 41,* 414–419.

Tolhurst, D. J., & Thompson, I. D. (1982). Organization of neurones preferring similar spatial frequencies in cat striate cortex. *Experimental Brain Research, 48,* 217–227.

Tootell, R. B., Silverman, M. S., Switkes, E., & De Valois, R. L. (1982). Deoxyglucose analysis of retinotopic organization in primate striate cortex. *Science, 218,* 902–904.

Uttal, W. R. (1975). *An autocorrelation theory of form detection.* Hillsdale, N.J.: Lawrence Erlbaum Associates.

Vander Lugt, A. B. (1964). Signal detection by complex spatial filtering. *IEEE Transactions on Information Theory, 10,* 2–7.

Wilson, M., & DeBauche, B. A. (1981). Inferotemporal cortex and categorical perception of visual stimuli by monkeys. *Neuropsychologia, 19,* 29–41.

You, K. C., & Fu, K.-S. (1979). A syntactic approach to shape recognition using attributed grammars. *IEEE Transactions on Systems Man and Cybernetics, 9,* 334–345.

Zahn, C. T., & Roskies, R. Z. (1972). Fourier descriptors for plane closed curves. *IEEE Transactions on Computers, C-21,* 269–281.

Zeki, S. M. (1978). Uniformity and diversity of structure and function in rhesus monkey prestriate visual cortex. *Journal of Physiology, 277,* 273–290.