Models of the Visual Cortex
Edited by D. Rose and V. G. Dobson

CHAPTER 9

Local log polar frequency analysis in the striate cortex as a basis for size and orientation invariance

PATRICK CAVANAGH
Département de Psychologie, Université de Montréal, Montréal, Québec, Canada, H3C 3J7

When an object moves across the visual field its image stimulates an ever-changing array of retinal receptors. If the observer is to recognize it as the same object in spite of these variations, this variable retinal input must be transformed into a unique pattern of neural activity that defines the object. There are two current models of form encoding: (a) an abstract, structural or propositional representation of the object, typically a list of the object's features and their interrelations, or (b) an analogue representation, typically a transformation onto a new set of dimensions where form information is invariant to changes in input size, position and orientation (for a further discussion of analogue versus structural models see Sutherland, 1973; Pylyshyn, 1973; Kosslyn, 1980).

A structural or propositional representation reduces the form to be identified to a list of primitive elements and the structural relations between them. If the primitives are, for example, lines (contours) and angles, the edges must first be extracted and then the relations between them determined. Pattern classification is based on structure and the descriptions of the primitive features do not necessarily specify their size, orientation or position. As a result, this type of encoding permits representations that are, in a very simple way, invariant with respect to the size, orientation and position of the overall pattern.

In the other model, stimulus patterns are transformed into analogue representations that do not vary with the size, orientation or position of the input. Following the original work of Schade (1956), Campbell and Robson (1968) proposed that the visual system performs a Fourier analysis on the

spatial attributes of the retinal input. The amplitude portion of the Fourier transform is invariant with respect to the position of the input. However, the striate cortex cannot carry out a true Fourier transform because the receptive fields at this level 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 (Pollen, Lee and Taylor, 1971; Robson, 1975 Glezer and Cooperman, 1977) but even if a Fourier transform were computed, this transform has neither size nor rotation invariance (Casasent and Psaltis, 1976).

If we look at the local structure of the striate cortex, however, a very striking and potentially useful organization is revealed—that of a log polar frequency transform (Cavanagh, 1978; Maffei and Fiorentini, 1977, Berardi *et al.*, 1982). The goal of this paper is to show how this organization might help in the analysis of patterns. I will describe first the receptive fields of simple and complex cells in the striate cortex and the organization of the cortical cells into local transforms of the retinal pattern. I will show how these local transforms change the rotation and magnification of a stimulus into simple translations of an invariant pattern of cortical activity. Finally, I will describe how these local transforms might be integrated or summed to give a global transform and how a final, translation-invariant transformation will then yield a position, size, and orientation-invariant encoding.

SPATIAL FREQUENCY AND POSITION INVARIANCE

In the striate cortex, simple and complex cells respond to bars of a particular width and orientation (Hubel and Wiesel, 1962, 1968). These dimensions, size and orientation, are the basis of the form-encoding process described here.

The sensitivity profile of the receptive field of a typical simple cell shows two or more parallel, elongated excitatory and inhibitory subfields. Although the optimum stimulus for the cell is a bar, or set of bars, aligned with the excitatory subfields, the cell will actually respond to a wide variety of stimuli. In general, the cell's output can be roughly predicted from integrating the product of the stimulus intensity and the receptive field sensitivity over the entire receptive field. The responses of the set of simple cells that respond to a stimulus can be thought of as a decomposition of the stimulus pattern into localized size and orientation-specific features (Marcelja, 1980). Theoretically, the stimulus pattern can be reconstructed from the outputs of the simple cells if their receptive field locations and sensitivity profiles are known. This decomposition possesses no invariances, however. Size, orientation and position will all influence the set of cells responding.

The receptive fields of complex cells, on the other hand, exhibit no spatially distinct excitatory or inhibitory regions. These cells respond uniformly to a drifting bar whatever its position in the receptive field, as long as it has the

optimal orientation and width (Hubel and Wiesel, 1962; Glezer *et al.*, 1980; Heggelund, 1981). These cells therefore show a specificity for pattern information (orientation and width of a bar or the spacing of the bars in a grating) but an indifference to position. Because the position invariance of the complex cells within their receptive fields is an important first step for the form-encoding process, it will be assumed that the output of the complex cells conveys the essential pattern information which is passed on to subsequent stages.

The position independence of the complex cell response implies that the decomposition of a stimulus by the complex cells may not be unique. Stimuli generally have a range of spatial frequency components at each orientation and those components that are sufficiently separated in frequency (about ± 1 octave) will stimulate different complex cells. Since positional information is lost within each receptive field, it is also lost for the relative locations of the frequency components detected by different cells. This implies that we should not be able to distinguish between images with the same frequency content but altered phase (position) content—e.g. a single dot would be indistinguishable from a field of random noise and a square wave (first and third harmonics in the peaks-subtract phase) indistinguishable from a triangle wave (first and third harmonics in the peaks-add phase). Since we can make these distinctions easily, phase or position information for frequency components must be encoded in some manner. In fact, the *relative* positioning or spatial phase relations between frequency components as well as the strengths of the components are together sufficient for a unique encoding of shape.

The broad bandwidth of the spatial frequency detectors may, therefore, play a part in encoding relative phase information. Simple and complex cells respond to a broad band of frequencies (about ± 1 octave; see Maffei and Fiorentini, 1973; Movshon, Thompson and Tolhurst, 1978) and may be sensitive to the relative positions or phases of the frequencies within that band. For example, simple cells with the same preferred spatial frequency but with symmetric or antisymmetric receptive fields (Stromeyer and Klein, 1974; Andrews and Pollen, 1979; Movshon, Thompson and Tolhurst, 1978) respond to similar ranges of spatial frequency but differ in their relative phase sensitivity. For a symmetrical receptive field, the frequency components of the stimulus within the range to which the cell responds must all be in the cosine phase (peaks aligned at the receptive field centre) to stimulate the cell optimally; however, for the antisymmetrical receptive field, they must be in the sine phase. Whether or not complex cells are selective for the relative phase is not so easily determined. DeValois and Tootell (1983) have shown that complex cells in cats are not sensitive to the relative phase but no work has yet been done in primates. Studies of complex cell response to drifting antisymmetric or symmetric brightness profiles could clarify the situation.

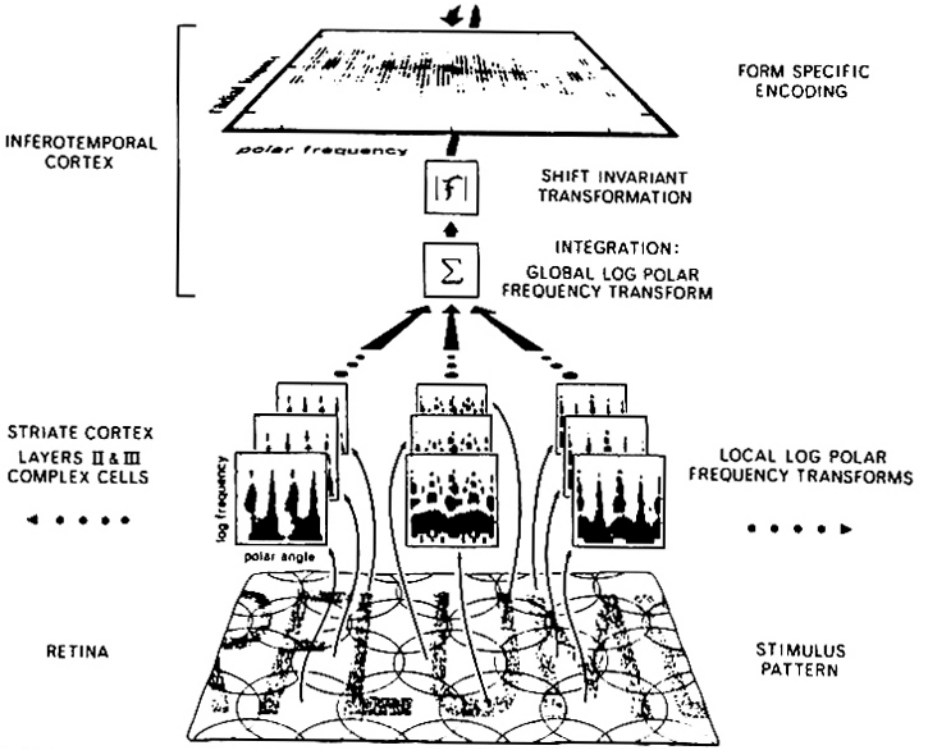

FIGURE 1 A schematic representation of the proposed transform sequence. The retinal image is encoded into an array of local log polar frequency transforms in the striate cortex. These representations are then summed together and retransformed (Fourier amplitude transform) to produce a position, size and orientation-invariant encoding, most likely in the inferotemporal cortex. The axes of the final transform, radial and angular frequency, are described in more detail elsewhere (Cavanagh, 1978, 1981, 1982, 1984).

LOCAL LOG POLAR FREQUENCY TRANSFORMS

Maffei and Fiorentini (1977) and Berardi *et al.* (1982) suggest that within each local region the cells in the striate cortex are organized into a two-dimensional matrix with the preferred orientation of the cells varying smoothly across the columns of the matrix and the preferred spatial frequency varying across the rows (see Fig. 1). This organization is remarkably similar to that of the log polar frequency transform. In this transform (Brousil and Smith, 1967; Casasent and Psaltis, 1976; Cavanagh, 1974, 1978; Zwicke, 1983), the spatial frequency information of the input pattern is organized in a two-dimensional array with orthogonal axes of orientation and the logarithm of the spatial frequency. This arrangement has two important properties:

(a) *Orientation*. Since one axis of the two-dimensional representation is orientation, a change in the orientation of the input shape changes the orientation of all the component features by the same amount and the whole transform pattern simply shifts left or right along the orientation axis. The shape of the transform pattern (the various blobs at the striate level of Fig. 1) is unchanged.

(b) *Log size*. When the input shape changes size by a factor x, all the component features are scaled by the same factor 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 and merely shifts up or down as a whole. A similar size invariance for a log frequency scale is seen in the uniform octave arrangement of a piano keyboard or the log frequency mapping of sounds in the auditory cortex (Tunturi, 1952). Both transform a given melody or chord into a pattern which is invariant to its absolute pitch (Altes, 1978).

This transformation converts rotation or magnification of the stimulus into translations of the activity patterns in the local transforms. However, any appreciable rotation, magnification or change in position of the stimulus will move the striate representation to a new set of local transforms. A *global* transform is essential if true position, size and orientation invariances are to be obtained. For example, a global transform could be produced by simply summing over all the local transforms. This summed transform would have the same dimensions as the local transforms but each cell's response would now represent the sum of activity at the preferred frequency and orientation over the entire visual field. This representation would (a) shift its position for size or orientation changes of the stimulus and (b) be indifferent to the stimulus position in the visual field (other than the loss of high spatial frequency information at more peripheral locations). If a further transformation were performed that was indifferent to the position shifts of the activity

pattern within the global transform (Fig. 1), a final representation would be generated that would be specific to the shape of the stimulus and invariant with respect to its position, size and orientation. Local transforms in the striate cortex have been reported in various recent articles; however, the proposed integration and retransformation of these local transforms are purely hypothetical. It is quite intriguing, though, that the axes of the local transforms in the striate cortex are just those required for size and orientation invariance.

STRIATE ARCHITECTURE

Maffei and Fiorentini (1977), Tootell *et al.* (1982) and Berardi *et al.* (1982) all claim that size (or spatial frequency) and orientation comprise two orthagonal dimensions for the local representations of the visual stimulus in the striate cortex. However, they do not agree on how the local transforms are arrayed in the striate cortex. According to Maffei and Fiorentini (1977) and Berardi *et al.* (1982), the preferred orientation of cells varies along an axis parallel to the surface of the cortex while the preferred spatial frequency of cells first increases as layer II and III of the cortex are traversed and then decreases again across layers IV, V and VI. However, Tootell *et al.* (1982) claim that frequency varies parallel to the surface rather than in depth. In either case, the physical orientation of the local transforms will not change their size-invariance property. Each local transform would be sufficiently wide to cover the full range of orientations (approximately 1 mm; see Hubel and Wiesel, 1974) and sufficiently deep or broad to include a single, ordered range of spatial frequencies. There is direct evidence that preferred orientation varies linearly with distance parallel to the cortical surface (Hubel and Wiesel, 1974; Maffei and Fiorentini, 1977, Berardi *et al.*, 1982), and I have presented indirect evidence that the change of preferred frequency with cortical distance is scaled logarithmically (Cavanagh, 1978, 1984). Each local transform represents the two-dimensional frequency by orientation encoding of the pattern for the particular region of the visual field that is covered by the receptive fields of the constituent cells (see Fig. 1). The upper two cortical layers (II and III) are most suitable as the site for the local transforms for two reasons. First, the complex cells having local position invariance predominate in these layers (Hubel and 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).

INTEGRATION AND FINAL TRANSFORMATION

Size, orientation and position information appear to have similar local organizaton in the prestriate (area 18) and striate cortices—a retinotopic organiz-

ation 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.

The receptive fields of the cells in the inferotemporal cortex 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 receive inputs from several cells in the striate cortex, effectively integrating 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 (Mishkin, 1972; Wilson and DeBauche, 1981, Dean, 1982).

The simplest way to integrate would be to sum together the local transforms of the striate cortex. Thus a cell at the global level would merely add the outputs of all complex cells preferring the same orientation, spatial frequency and relative phase. This representation is sufficient to code a stimulus pattern uniquely (Cavanagh, 1984). The resulting invariant pattern will shift in position in the log polar transform plane as a function of changes in the size and orientation of the stimulus anywhere in the visual field. (Note, however, that because of these shifts, some information will be lost or added at the borders of the transform plane. The result is a gradual drop-off in recognition performance when target and test differ in size or orientation, and this drop-off mimics that seen for human subjects under these conditions; see Cavanagh, 1978.)

To achieve a form-specific encoding, the invariant pattern on the log polar plane must be extracted and its position on that plane, which is determined by its size and orientation, ignored. The Fourier amplitude transform is, for example, position invariant. A Fourier amplitude transform of the log polar frequency plane will therefore encode the pattern of activity on that plane and ignore its position. The demonstration in Fig. 1 has used the Fourier amplitude transform at the final step but, as mentioned previously, the Fourier amplitude transform does not unambiguously encode patterns and so some other position-invariant transform would be, in fact, preferable.

Following this shift-invariant step applied to the global log polar representation, a given input shape at various different sizes, positions and orientations always produces a fixed, unchanging pattern of cell firing rates. (Size, orientation and position information are no longer represented at this level and must be processed by other means, either a more elaborate encoding transform or a parallel analysis.) Such fixed patterns could then be used as templates to identify future instances of the same pattern at new locations, sizes and orientations. Since the inferotemporal cortex is the only non-retinotopically organized visual cortex, these proposed size and orientation-invariant encodings must be located in this area as well. (Note that instead

of separate and sequential steps of integration and final transformation, the two operations could be combined into a single step.)

CONCLUSIONS

An encoding transform has been described that can produce a position, size and orientation-invariant representation of form. The two essential steps in the sequence are (a) a position invariant encoding of the input that arrays the pattern information along axes of orientation and log size and (b) a position invariant encoding of this log polar representation.

Since the proposed encoding sequence represents form independently of position, size and orientation, stimuli may be classified by matching their transforms against previously stored templates—final transforms of prototype patterns such as letters, familiar faces, common shapes and familiar words. The transformed input would have to be matched in parallel against all prototypes (see Cavanagh, 1975, 1976; Anderson *et al.*, 1977, Kohonen, 1977; Murdock, 1982; Eich, 1982).

No template matching scheme could ever analyse real world scenes involving shadows, partially hidden objects and objects recognized by function (e.g. chairs). However, it seems improbable that the visual system would develop a position, size and orientation-invariant template mechanism just for a few specialized tasks. What then could be the role of such a mechanism? One possibility is that a structural analysis (e.g. Marr, 1982) might be able to take, as its data base, 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 process could start after the template mechanism had matched all elements in the scene for which stored representations were available. When the scene is totally unfamiliar, the available 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 position, size and orientation-invariant primitives to serve as the basis for an intelligent structural analysis.

The proposed encoding process, if it is actually used by the visual system, would probably operate as only one of several parallel analyses of the visual input. Information from colour, 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. Finally, the encoding sequence described here requires at least two distinct relative phase sensitivities for complex cells in order to eliminate phase ambiguities. If these ambiguities cannot be corrected, then the role of the size and orientation detectors of the striate cortex may be one of texture analysis (Robson, 1980) rather than form encoding.

ACKNOWLEDGEMENTS

This research was supported by NSERC grant A8606 and by the Ministère d'Education du Québec. The helpful comments of Stuart Anstis and Ian Howard are gratefully acknowledged.

REFERENCES

Altes, R. A. (1978). The Fourier-Mellin transform and mammalian hearing. *J. Acoust. Soc. Amer.*, **63**, 174–183.

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

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

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

Brousil, J. K., and Smith, D. R. (1967). A threshold logic network for shape invariance. *IEEE Trans. Computers*, **EC-16**, 818–828.

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

Casasent, D., and Psaltis, D. (1976). Position, rotation, and scale invariant optical correlation. *App. 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 *Formal Aspects of Cognitive Processes*, (Eds. T. Storer and D. Winter), Springer-Verlag, Berlin, pp. 14–40.

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

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

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 Res.*, **22**, 1409–1412.

Cavanagh, P. (1984). Image transforms in the visual system. In *Figural Synthesis* (Eds. P. C. Dodwell and T. M. Caelli), Lawrence Erlbaum Associates, Hillsdale, New Jersey, pp. 185–218.

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

DeValois, K. K., and Tootell, R. B. H. (1983). Spatial-frequency-specific inhibition in cat striate cortex cells. *J. Physiol.*, **336**, 359–376.

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

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

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

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

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

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

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

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

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

Kosslyn, S. M. (1980). *Image and Mind*, Harvard University Press, Cambridge, Mass.

Lund, J. S., Lund, R. D., Hendrickson, A. E., Bunt, A. H., and 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. *J. Comparative Neurol.*, **164**, 287–304.

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

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

Marcelja, S. (1980). Mathematical description of the responses of simple cortical cells. *J. Optical Soc. Amer.*, **70**, 1297–1300.

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

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

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

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

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

Pylyshyn, Z. W. (1973). What the mind's eye tells the mind's brain: a critique of mental imagery. *Psycholog. Bull.*, **80**, 1–24.

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

Robson, J. (1980). Neural images: the physiological basis of spatial vision. In *Visual Coding and Adaptability* (Ed. C. S. Harris), Lawrence Erlbaum Associates, Hillsdale, New Jersey, pp. 177–214.

Schade, O. H. (1956). Optical and photoelectric analog of the eye. *J. Optical Soc. Amer.*, **46**, 721–739.

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

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

Sutherland, N. S. (1973). Object recognition. In *Handbook of Perception*, Vol. 3, Biology of Perceptual Systems (Eds. E. D. Carterette and M. P. Friedman), Academic Press, New York, pp. 157–206.

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

Tunturi, A. R. (1952). A difference in the representation of auditory signals for left and right ears in the iso-frequency contours of tight middle ectosylvian auditory cortex of the dog. *Amer. J. Physiol.*, **68**, 712–727.

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

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

Zwicke, P. E. (1983). A new implementation of the Mellin transform and its application to radar classification of ships. *IEEE Trans. Pattern Analysis and Machine Intelligence*, **PAMI-5**, 191–199.