Perception, 1981, volume 10, pages 469-474

# Size invariance: reply to Schwartz

Patrick Cavanagh
Département de Psychologie, Université de Montréal, Montréal, Québec
Received 27 March 1979, in revised form 29 July 1980

## 1 Introduction

There are several important points on which Schwartz and I are in agreement:
(i) The spatial representation of a stimulus in the brain may be of direct functional significance to perception.
(ii) The Fourier transform itself is not physiologically realizable in the visual system (Cavanagh 1978, p 174).
(iii) The Fourier transform does not possess size and rotation invariances (Cavanagh 1978, p 167).
(iv) Phase information must be retained in one form or another (Cavanagh 1978, p 171).
(v) Local analyses must be combined at some level to provide a global representation (Cavanagh 1978, p 169).

We differ, however, in that I proposed that a Fourier transform can be combined with a log polar mapping to provide local transform fields that are physiologically plausible in the striate cortex and that can be integrated at a later stage to provide size invariance. Schwartz, on the other hand, proposes a radically different basis for size invariance: the retinotopic mapping of the entire striate surface. I shall first present arguments showing that this topography does not possess functional size invariance and then elaborate on several points relevant to the log polar frequency proposal: spatial inhomogeneity, phase encoding, transform integration, and overall processing architecture.

## 2 Log polar mapping

Schwartz (1977, 1981) proposes that the log polar mapping of the retinal input seen on the striate cortex is the basis of the size and rotation invariances obtained in perception. This proposal is subject to four important restrictions. First, the only size and rotation changes for which the log polar mapping alone is invariant are those centred at the origin [the fovea in the case of the visual system (Schwartz 1981; Chaikin and Weiman 1979)]. Second, patterns represented on the log polar mapping have no position (translation) invariance. Changing the location of the input pattern radically alters its representation (figure 1). Third, the cortical surface cannot, in fact, be laid out flat as Schwartz proposes (Daniel and Whitteridge 1961). A planar mapping is at best an approximation that varies with species in goodness-of-fit. Fourth, the transform that Schwartz has fitted to the experimental data is only log polar in the periphery—beyond 5 deg eccentricity (Schwartz 1977, figure 1c) in the owl monkey and not at all in the lower hemifield of the cat. There is, however, no evidence of the deficiencies in size invariance for the fovea or for the lower hemifield of the cat that would be expected if Schwartz's proposal were correct. The article by Ross et al (1980) cited by Schwartz as support for foveal deficiencies is, in fact, totally irrelevant. Ross et al investigated size constancy, the judgment of size at a distance, and not size invariance, the ability to recognize objects independently of

size. The one possible functional niche for Schwartz's proposal is the one he outlines (Schwartz 1981, p 16). That is, for an observer moving through an environment in linear motion, objects in the periphery will retain a fixed (although not stationary) representation on the cortical surface. This, however, remains true only as long as the observer does not change his direction of motion and as long as his axis of gaze is colinear with his axis of motion. These restrictions are so limiting that it is highly unlikely that this process could subserve size invariance even in the most primitive of senses.

It seems clear then that pattern analysis would be only complicated, not simplified, if the striate surface projection were the base data of the recognition process. It is more likely that the intriguing log polar nature of the striate topology reflects simply the optimum packing solution for a high-resolution fovea and a low-resolution periphery.

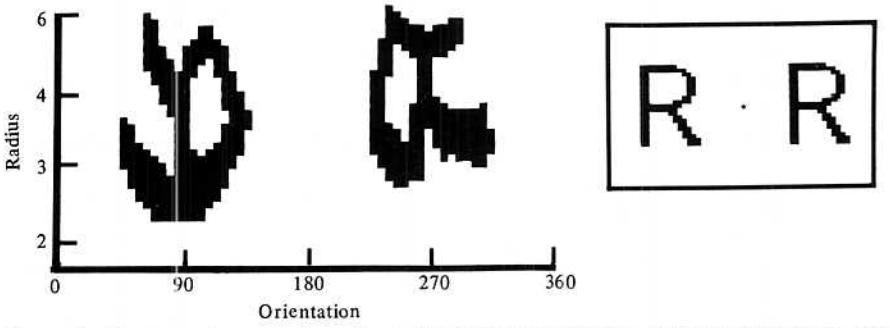


**Figure 1.** The log polar mapping of two identical upright letters R situated on either side of the origin of the input plane, as shown in the inset. Orientation is clockwise from a line drawn downwards from the fixation point, radius is in arbitrary units. Position invariance is not obtained.

### 3 Log polar frequency mapping

The processing sequence that I (Cavanagh 1974, 1978) and others (Brousil and Smith 1967; Casasent and Psaltis 1976, 1977) have described differs principally from that proposed by Schwartz in that the log polar mapping is preceded by a Fourier amplitude transform. This has two advantages. First, the essential requirement for obtaining general size invariance with the log polar mapping is that all size changes—expansions or contractions—must be centred at the origin. This requirement is always satisfied if the log polar mapping is preceded by a Fourier transform. Rotation and size changes of a stimulus pattern at any location produce corresponding size and rotation changes of its Fourier representation but these changes are always centred at the *origin* of the transform. A log polar mapping of the transform domain itself therefore obtains invariances for all possible size changes of an arbitrarily located object. Second, the amplitude portion of the Fourier transform is invariant to position and so the result of a log polar mapping of the amplitude transform is identical for all stimulus positions.

The transform sequence therefore has very important pattern-recognition properties. Owing to the use of the Fourier amplitude transform it also has, as Schwartz points out, two distinct disadvantages. First, the lack of phase information in the amplitude transform leads to a number of ambiguities in pattern recognition. For example, positive and negative images are indistinguishable, as are 180° rotations of images. Second, it is quite unlikely that an amplitude transform is being computed by the visual system (although see Tyler and Sutter 1979).

Note, however, that the use of the amplitude transform was clearly stated to be for purposes of demonstrating the existence of a size- and position-independent encoding based on spatial frequency information (Cavanagh 1978, p 168). It was

not proposed that the Fourier amplitude transform itself is computed by the visual system and, in fact, it was specifically pointed out that this is not the case (Cavanagh 178, p 174).

What was interesting about the transform sequence was the similarity between the intermediate stage of the log polar frequency representation and the organization of local areas of the striate cortex as revealed by Maffei and Fiorentini (1977). They have reported what appear to be local transforms with preferred orientation of cells varying along an axis parallel to the surface of the cortex, and preferred spatial frequency varying orthogonally along an axis perpendicular to the surface of the cortex. The local transforms are delineated by an abrupt reversal in the direction of change of preferred orientation. There is direct evidence that the preferred orientation varies linearly with distance (Hubel and Wiesel 1974; Albus 1975) and I presented indirect evidence that the frequency axis may be logarithmic.

This combined evidence points to the possibility of local log polar frequency transforms arrayed like file cards in the striate cortex surface. Each local transform encodes the visual input centred at its particular retinotopic locus. A single pattern may then be encoded by several thousand overlapping local transforms.

Note that, within each local transform, size and rotation changes of the input are already transformed into simple shifts of the transformed pattern. Changing the locus of the input pattern will change the locations of the local transforms within which these shifts occur but will not change the nature of the encoded representation. Thus perhaps the most essential step to obtaining size invariance in the manner described by the Fourier-log polar-Fourier transform sequence may already be achieved at the level of the striate cortex.

## 4 Some questions

If these local log polar frequency plots are functionally involved in a size-invariant encoding in the visual system, then several questions should be considered.

(i) *Spatial inhomogeneity.* The range of preferred spatial frequencies of cortical cells changes with eccentricity as do the receptive field sizes. This simply implies that pattern resolution will change as a function of retinal location, as less high-frequency information will be encoded in the peripheral regions (Berkley et al 1975). The inhomogeneity complicates the modelling of the visual system but there is no reason for it to rule out a spatial-frequency-based encoding. It certainly rules out the notion that a mathematically pure Fourier transform could be computed, but this never was a viable theory in any case.

(ii) *Local spatial inhomogeneity.* Within a single local transform, even if it is assumed that the receptive fields of all cells are centred at the same retinal location, the size of the receptive fields will vary greatly. A low-frequency cell may be responding to a stimulus element which is totally outside the receptive field of a higher-frequency cell. This certainly implies that a single local transform is not very meaningful by itself, and we must look to the process that integrates the local transforms to evaluate the importance of the variations in receptive field size.

(iii) *Integration of local transforms.* There is no possibility of considering a size- and position-invariant encoding unless the local transforms are combined in some way. Certainly the simplest proposal is just to sum them all together to form a single final transform. If the sums are formed with respect to each frequency, orientation, and position, the true Fourier coefficients could theoretically be generated[1], subject to

[1] $G(f, \theta, \phi) = \sum_i g_i(f, \theta, \phi)$ where $f$ and $\theta$ are the frequency and orientation, respectively, $\phi$ is a phase dimension having at least two values assuring orthogonality, $G$ is the global transform, and $g_i$ is the $i$th local transform.

the limitations of channel bandwidth and of the spatial inhomogeneity mentioned above. It is possible to consider a simpler analysis, however, where the local transforms are added with respect to frequency and orientation of each but with the position ignored[2]. In the case of simple cells, cell output appears to represent a half-wave rectification of the spatial input (Movshon et al 1978; Andrews and Pollen 1979); a straightforward sum of these rectified outputs would produce an average-absolute-value transform with properties similar to those of the amplitude transform—including the loss of phase information. Other possibilities might be considered, but this simplest of options brings up the importance of preserving phase or position information in some manner.

(iv) *Absolute and relative phase.* In terms of pattern recognition, absolute phase or position information is not of great importance. More important is the relative phase of the frequency components of a given pattern, as the relative phase offsets between components together with their amplitudes are sufficient to specify the pattern uniquely. Some cells in the cortex do appear to respond to two or more harmonically related components (Pollen et al 1978; Glezer et al 1976). In fact, any cell whose receptive field is nonsinusoidal will provide relative phase information (Cavanagh et al 1980). Rather than looking on these departures from the standard orthogonal base set of the Fourier transform as a drawback, the introduction of relative phase into the analysis may be very productive. There is, however, a wide variety of cell types and receptive field profiles, and this diversity requires a matching degree of complexity in the processes which interpret the encoding. The possibility of a rather large set of encoding primitives is a problem for any model of visual analysis and is not particular to the transform approach.

(v) *Preferred stimuli of the transform sequence.* I have outlined responses to a number of problems raised by the proposal of functional local log polar frequency transforms. Assuming an integration of the local transforms is accomplished, the transform sequence requires a final position-invariant transformation to achieve true size invariance. These two steps—integration and transformation—could be accomplished simultaneously, and I suggested that any speculation concerning the location of the final transform result should include the inferotemporal cortex. Schwartz notes that periodic grating patterns have not been found to be particularly effective stimuli in areas where cells with extremely large receptive fields are found. This is exactly as would be expected if the proposed transform sequence were in effect. Figure 2 shows that, for pure Fourier amplitude transforms for the first and last steps, the stimuli that translate into single points on the last level are the family of logarithmic spirals given by

$$h(r, \theta) = \sin(n\theta + a \log r) \qquad (1)$$

where $r$ is the radius, $\theta$ the polar angle, and $n$ an integer; $n$ and $a$ are defining parameters.

All of these patterns have the property that changes in size or orientation have either no effect or cause periodic repetitions of the original patterns. Changes of the properties of the first and last transform steps to reflect the physiology of the visual system would of course change the particular preferred stimuli (ie those mapping to a single point) but should not change the cyclical nature of their response to size and/or rotation changes. Another family of patterns that is cyclical to size changes is that of the fractal patterns (Mandelbrot 1977).

[2] $G(f, \theta) = \sum_i \sum_\phi g_i(f, \theta, \phi)$ terms as before.

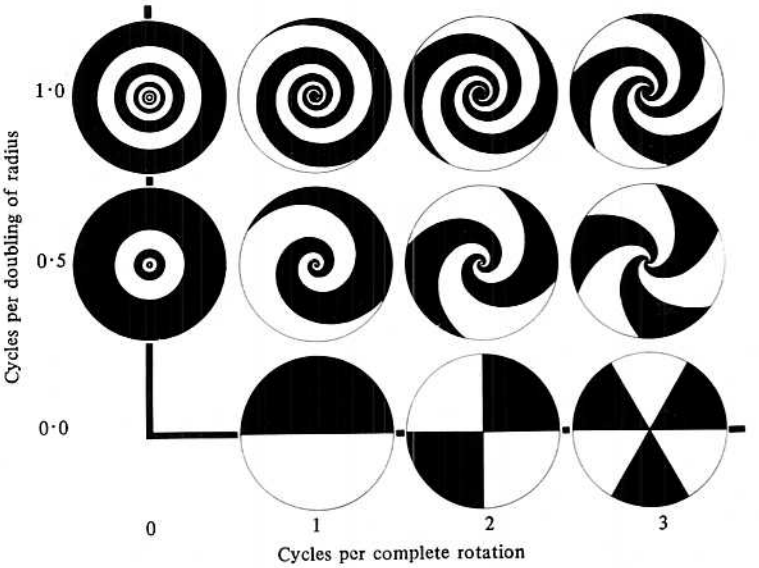


**Figure 2.** Preferred stimuli of cells at various locations in a Fourier-log polar-Fourier transform representation. Each cell should respond maximally to its preferred stimulus independently of its position in the visual field. 'Cycles per complete rotation' reflects the parameter $n$ of equation (1) and 'cycles per doubling of radius', reflects parameter $a$. For simplicity, the luminance profiles are shown in black and white and as size-limited, but are actually sinusoidal and of unlimited extent as given in equation (1).

(vi) *Templates of visual scenes.* On this final level it would be possible to store pattern representations and then use these again later as recognition filters (see Anderson et al 1977; Cavanagh 1976; Kohonen 1977) that now possess size and to some extent position and rotation invariances, depending on the exact encoding transform. What can be accomplished with such templates? Recognizing unfamiliar handwriting, distinguishing shadows from objects, classifying an object according to function, these are things that cannot be done with templates. That is, the essential work of interpreting a visual scene is certainly well beyond the capacities of any template system anyone might wish to devise. Powerful contextual scene analysis programs have been developed in recent years (cf Winston 1975; Hanson and Riseman 1978) and the brain may well be performing analogous analyses.

The scene analysis programs generally start from a line and angle description and must deduce each object from its contour structure [primal sketch (Marr 1976)]. Such deduction is a complex, typically serial task. It would be of great and obvious advantage if, instead of a base level encoding of lines, the scene analysis were able to start from higher-level patterns as primitives—simple forms, letters, familiar faces—whatever might already have been stored as size-invariant filters. Whenever patterns arise for which no higher-level filters are appropriate, the encoding falls to lower-level primitives.

The transform sequence that I have proposed is thus one possible method of bringing high-level encoding primitives into the first level of scene analysis. The physiology of the visual system seems appropriate for effecting direct transforms of visual input such as these, although the actual transforms postulated here are quite speculative. The two immediate concerns in evaluating the transform approach are the interpretation of the possible encoding primitives provided by the diverse receptive-field and spatial-frequency properties of the striate cells, and the integration of the local transforms into a global representation.

**References**

Albus K, 1975 "A quantitative study of the projection area of the central and paracentral visual field in area 17 of cat. II The spatial organization of the orientation domain" *Experimental Brain Research* **24** 181-202

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

Berkley M A, Kitterle F, Watkins D W, 1975 "Grating visibility as a function of orientation and retinal eccentricity" *Vision Research* **15** 239-244

Brousil J K, Smith D R, 1967 "A threshold logic network for shape invariance" *IEEE Transactions on Electronic Computers* **EC-16** 818-828

Casasent D, Psaltis D, 1976 "Position, rotation and scale invariant optical correlation" *Applied Optics* **15** 1795-1799

Casasent D, Psaltis D, 1977 "New optical transforms for pattern recognition" *Proceedings of the IEEE* **65** 77-87

Cavanagh P, 1974 "A two-dimensional position, size and rotation invariant pattern transform: an electrooptical process and a neural analogue" Unpublished manuscript, Département de Psychologie, Université de Montréal, Canada

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, Brussell E M, Coupland S, 1980 "Spatial frequency characteristics of discrimination mechanisms" *Investigative Opthalmology and Visual Science, ARVO Abstracts Supplement* **8** 44

Chaikin G, Weiman C, 1979 "Logarithmic spiral grids for image processing" in *Proceedings of the IEEE Computer Society Conference on Pattern Recognition and Image Processing,* Chicago, pp 25-31

Daniel P M, Whitteridge D, 1961 "The representation of the visual field on the cerebral cortex in monkeys" *Journal of Physiology* **159** 203-221

Glezer V D, Cooperman A M, Ivanov V A, Tsherbach 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

Hanson A R, Riseman E M, 1978 *Computer Vision Systems* (New York: Academic Press)

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

Kohonen T, 1977 *Associative Memory* (Berlin: Springer)

Maffei L, Fiorentini A, 1977 "Spatial frequency rows in the striate visual cortex" *Vision Research* **17** 257-264

Mandelbrot B B, 1977 *Fractals: Form, Chance, and Dimension* (San Francisco, Calif.: W H Freeman)

Marr D, 1976 "Early processing of visual information" *Philosophical Transactions of the Royal Society of London B* **275** (942) 483-524

Movshon J A, Thompson I D, Tolhurst D J, 1978 "Spatial summation in the receptive fields of simple cells in the cat's striate cortex" *Journal of Physiology* **283** 53-77

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

Ross J, Jenkins B, Johnstone J R, 1980 "Size constancy fails below half a degree" *Nature (London)* **283** 473-474

Schwartz E L, 1977 "Spatial mapping in primate visual cortex: analytic structure and relevance to perception" *Biological Cybernetics* **25** 181-194

Schwartz E L, 1981 "Cortical anatomy, size invariance and spatial frequency analysis" *Perception* **10** 455-468

Tyler C W, Sutter E E, 1979 "Depth from spatial frequency difference: An old kind of stereopsis?" *Vision Research* **19** 859-865

Winston R H, 1975 *The Psychology of Computer Vision* (New York: McGraw Hill)