On the retino-cortical mapping

STEWART W. WILSON

Research Laboratories, Polaroid Corporation, Cambridge, Massachusetts 02139, U.S.A.

Based on Hubel & Wiesel's physiological findings on the projection from retina to cortex, a schematic model of that stage of visual processing is constructed and its properties investigated. The projection or mapping appears to carry out an automatic "normalization of description" for the same object independent of retinal image size. This property suggests new concepts regarding (1) contrast sensitivity, (2) the nature and role of indirect vision, (3) the role of eye movements and (4) the recognition of patterns and the analysis of scenes.

Introduction

This article explores implications for human visual perception of the results of Hubel & Wiesel (1974), in macaque monkeys, on the relationship between receptive field size, magnification and eccentricity. A schematic model of the retino-cortical projection is constructed which seems roughly to include their results, and leads to an hypothesis about the general purpose of that stage of visual processing. The model introduces the notion of "data field" as a generalization of Hubel & Wiesel's "aggregate receptive field", and the notion of "message sending unit" (MSU), as a dedicated cell sub-assembly of which their "hypercolumn" is the prototypical example. The hypothesis about the retino-cortical projection is that it is arranged so that, in the presence of certain world constancies whose retinal images change with viewing distance, the output signals—in the model's terms, messages—of the primary visual cortex are constant and independent of viewing distance. The model is applied to an explanation of the contrast sensitivity function; to Aubert & Foerster's (in Helmholtz, 1962) classic experiment, and Lettvin's (1976) observations, on peripheral visibility; and to some basic questions of pattern recognition.

Physiology and model

In the view of Hubel & Wiesel (1979), the primary visual cortex is both anatomically and physiologically a regular array in which a standard unit of neuronal processing machinery, the hypercolumn, is repeated over and over. Each hypercolumn has some thousands of input fibers which project backwards via the lateral geniculate body to ganglion cells subserving a delimited region of the retina. Each cortical cell measured in a perpendicular penetration of the hypercolumn will respond to a characteristic stimulus over a locale within that delimited region; the precise response locale is termed that cell's receptive field. The fields vary in size for different cells in the hypercolumn by a factor of two or three. The variation is probably because different cortical cells compute different things: for example, the so-called "complex" cells by

definition have a larger field than the "simple" cells. The receptive fields in one hypercolumn also tend to vary somewhat in position (the positions of their centers); the variation is about the same as the variation in (linear) field size; this is termed "scatter". Hubel & Wiesel call the "pile of superimposed fields that are mapped in a penetration beginning at any point on the cortex the 'aggregate field' of that point". In effect, each hypercolumn may be said to have an aggregate field.

In the model which will now be developed, the hypercolumn is regarded as a "message sending unit" (or MSU) and its aggregate field is regarded as an experimentally observed manifestation of the MSU's "data field", i.e. the retinal region over which the MSU collects stimulus information. Hubel & Wiesel say the hypercolumn has "perhaps 50,000" output fibers. Incorporating this, the model's MSU is an entity which receives input on some thousands of input fibers from its data field, computes simple overall properties of the stimulus (related to, but not necessarily the same as, those that Hubel & Wiesel have identified by probing single cells within the hypercolumn-MSU) and outputs the presence or absence of these properties—"the messages"—as a code on the 50,000 output fibers. In the ensuing discussion, the reader may certainly still think of "hypercolumn" for "MSU" and "aggregate field" for "data field". New terminology has been introduced because the model and the physiology are distinct, and to emphasize suggested function.

In developing the model further, let us go back to Hubel & Wiesel (1974) and examine particularly their fig. 6A (reproduced here as Fig. 1). We want to know more
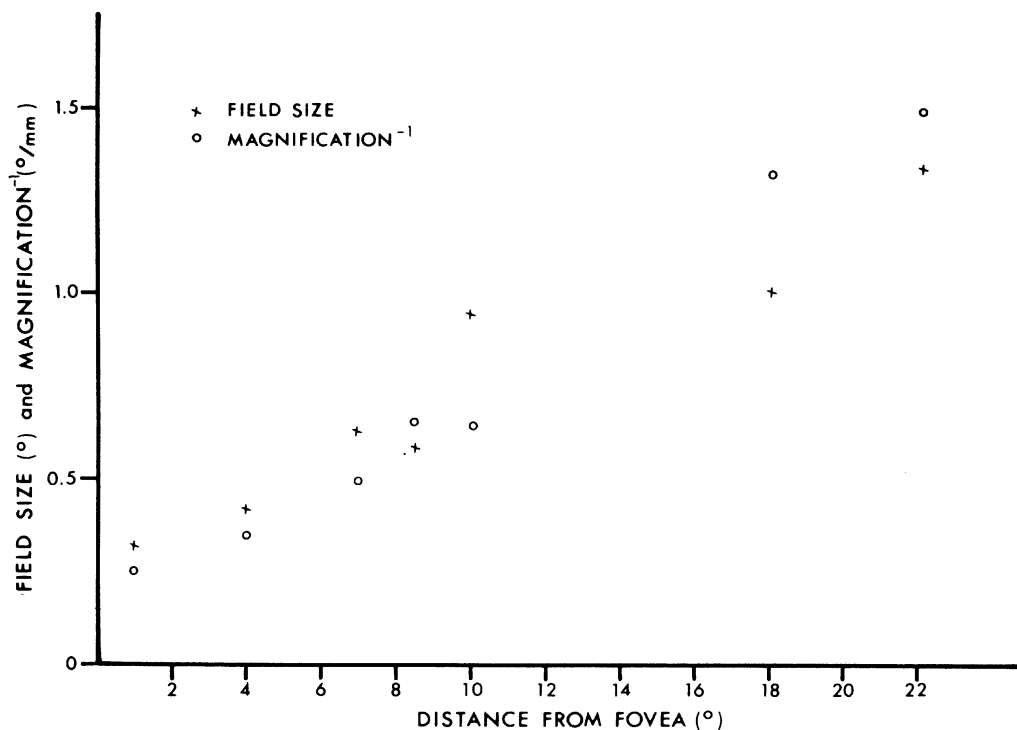

Fig. 6A   Graph of average field size (crosses) and magnification^{-1} (open circles) against eccentricity, for five cortical locations. Points for 4°, 8°, 18° and 22° were from one monkey; for 1°, from a second. Field size was determined by averaging the fields at each eccentricity, estimating size from (length × width)$^{0.5}$.

FIG. 1. Field size and magnification^{-1} vs distance from fovea in macaque monkey. [From Hubel & Wiesel (1974), with original caption. Reprinted by permission of Alan R. Liss, Inc., copyright owners.]

precisely the relationship between retina and primary visual cortex. At a given eccentricity, how big is the aggregate field of a hypercolumn, and to what extent do the aggregate fields of adjacent hypercolumns overlap?

Figure 1, while containing just a few hard-won points, gives an answer. (1) Field size increases as a linear function of eccentricity, with a slope of about 0·05 deg/deg, and the field size close to the fovea is about 0·25°. ("Field size" is the average size of the receptive fields comprising the aggregate field at that point.) (2) The parallelism between field size and "magnification^{-1}" means that the spacing between aggregate fields for adjacent hypercolumns is such that the diameters of the two aggregate fields overlap by about 50%.

For the model, it will be assumed that field size vs distance from fovea in Fig. 1 really is a straight line; also that the overlap percentage is constant, independent of eccentricity. As noted, field size in Fig. 1 is about 0·25° at 0° "distance from fovea". Whether this is distance from the fovea's boundary or center is not made clear. The model is neater if one assumes that the slope of the curve outside the fovea would allow straight-line extrapolation to zero field size at exact fovea center.

Suppose now that one wants to diagram the input domain of the retino-cortical mapping on a piece of paper. For this, place the fovea center at the center of the paper and let radial distance from there represent angular distance to a given point on the retina. To show the mapping, one might place dots on the paper, one for each hypercolumn, wherever the center of the hypercolumn's aggregate field falls. But how should the dots be arranged? What is the correct pattern of their spacing?

The above interpretation of Fig. 1, plus the assumption that the arrangement is radially symmetric, gives the answer. One curve says field size (diameter, say) is proportional to distance. The other curve says overlap is a constant percentage of field size. Since overlap is linearly related to both field size and field separation, constancy of percentage overlap requires that the dots be laid out so that *the distance between adjacent dots is also proportional to distance from the center.*

Figure 2 shows the resulting arrangement. Imagine each dot to be encircled by a field which extends more-or-less up to the nearest neighboring dots. One can see that (1) field size is proportional to eccentricity; (2) the fields have constant percentage overlap; and (3) (a consequence of the first two properties) field *spacing* is proportional to eccentricity.

The pattern of Fig. 2 has a fundamental property which we can establish by placing an object on the pattern. In Fig. 3, imagine that the inner square is the outline of some general object which has been imaged on the retina. The object happens to be fixated at its center in this case. The various parts of the object are being examined, so to speak, by the data fields (aggregate fields) of the MSUs (hypercolumns) whose representative dots fall on the object. (It is assumed, as before, that the data field of a dot is a circle extending more-or-less up to the nearest neighboring dots.) The outer parts of the object are examined by fewer data fields per unit area than the inner parts. We can suppose that this object has surface detail of some kind, so that the various data fields see various things "under" them and the associated MSUs send out various messages.

Now, let us magnify the object on the retina somewhat so it now is represented by the outer square. The diagram reveals that for every data field now receiving input from some portion of the object, there was, prior to magnification, another data field
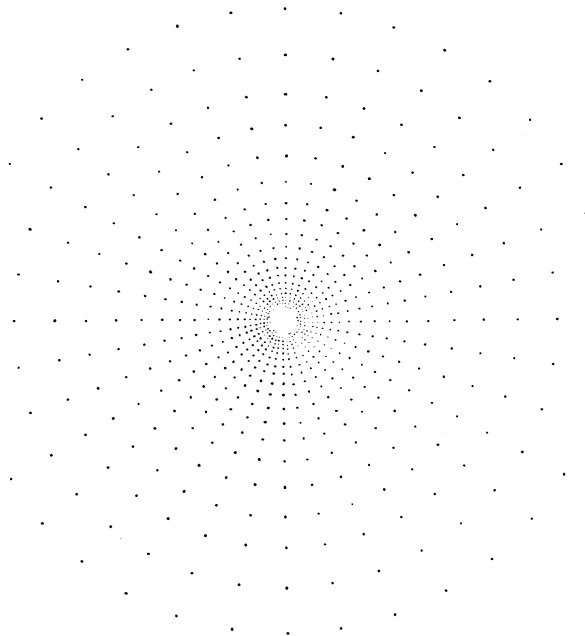
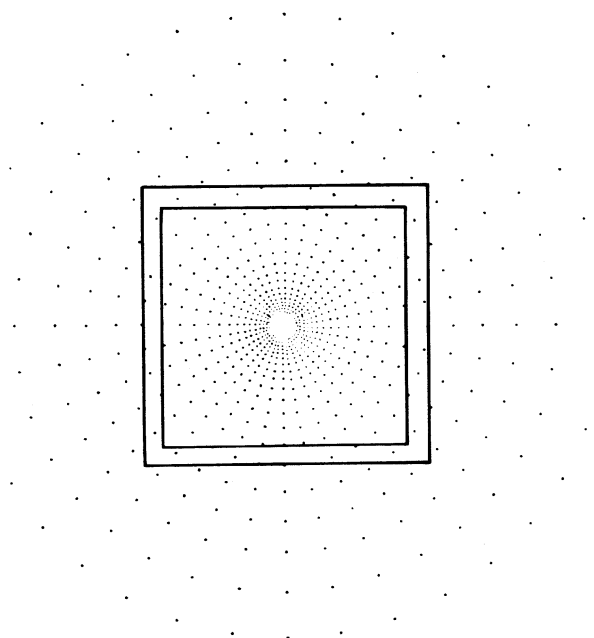FIG. 2. Arrangement of MSU data field centers in retinal space.

FIG. 3. Array of Fig. 2 with superimposed, centered images of a "generalized object" at two magnifications.

which received its input from precisely the same object portion. Thus, if all MSUs are similar (Hubel & Weisel's assumption for hypercolumns), the *ensemble of messages* output by the MSUs will not differ between the two magnifications. The particular set of MSUs *sending* the messages will shift slightly, but not the messages themselves.

It therefore appears conceivable that the retino-cortical mapping is a system designed to produce a similar output or, so to speak, to "tell the same story", whatever the

size of an object's image on the retina. The output message ensemble may be likened to a set of reports that speak about the various portions of an object, where "portion" has units something like "percent of overall object size". The effect is to extract information about an object's visual pattern, independently of its size in space or the distance at which it is seen. Such a property is consistent, of course, with daily experience. The appearance of an object is substantially independent of its accidental size on the retina.

Indirect vision

In the *Physiological Optics*, Helmholtz (1962) described experiments by Aubert & Foerster on "the precision of vision in the peripheral parts of the retina". In one experiment, they investigated the extent to which, during a brief flash, subjects could identify letters and numerals which occurred at an angle to the direction of vision at the moment of the flash. That is, they asked: what is the maximum off-axis (horizontal, in this case) angle at which a letter of a given size will still be just recognizable? Aubert & Foerster's main result was that, for letters of a given actual size, the maximum angle was proportional to the letter's visual angle, the ratio being approximately five to one.

It may be helpful to put this important result somewhat differently. Imagine the following special kind of eye chart which an eye doctor might have. On a large white card there is a small x on which one fixates, and then, somewhat to the side of the x, a single letter such as "E". Suppose that when you look at the x, you can indirectly see the E just well enough to tell that it is an E. What Aubert & Foerster discovered is that if you can just barely tell the E at this distance from the chart, you can move forward and backward (within fairly wide limits, always fixating the x), and the E will *still* be just barely identifiable.

Some flavor of the experiment may be gained by looking at Fig. 4. If the x is fixated, the appearance of the E will be substantially independent of the distance at which you hold the paper.

x   E

FIG. 4. "Eye chart" for indirect vision.

These phenomena are very important for the model of the retino-cortical mapping. In the "expanding diagram" of Fig. 2, suppose the fixated x to be at the center; the image of the E will then fall somewhere away from the center. It will be "under" the data fields of some number of MSUs—more if the E is close to the center, fewer if farther away—but, in any case, some definite number. Now, what will happen as, say, our patient approaches the eye chart, or we move Fig. 4 nearer? The answer is clearly that the E will move farther out from the center; at the same time, it will also grow larger. In fact, its distance and its size will grow by exactly the same factor.

But if this is true, a brief consideration of the properties of Fig. 2 will persuade one that the E, in its new position, will fall under the same number of MSU data fields as before. Put more generally, as long as the point of fixation does not change,

the number of MSU data fields "subtended" by an indirectly seen object is independent of distance.

Now, from Aubert & Foerster's experiment we can draw two important conclusions. First, since they found that recognizability in this situation is independent of distance, one can infer that the MSUs, as information processors, are functionally identical. For, if equal recognizability at all distances goes together with an unchanging number of "inspecting MSUs", then, since the actual MSUs doing the inspecting *do* change, it must be concluded that one MSU is just as good as any other MSU—i.e. they are functionally alike.

There is thus independent support for Hubel & Wiesel's assumption of the similarity of the hypercolumns. But from Aubert & Foerster one can infer something more. Since the "E" on the eye chart may be drawn at such a distance from the fixation "x" that it is *no longer* recognizable, it must be the case that each MSU has a finite and limited information processing capacity. Can we say anything about this capacity, that is, about the specific computations which the MSU performs?

In an article entitled "On seeing sidelong", Lettvin (1976) drew attention to the way in which the visibility of an indirectly seen object depends strongly on the presence or absence of other objects in its immediate neighborhood. Figure 5 gives examples
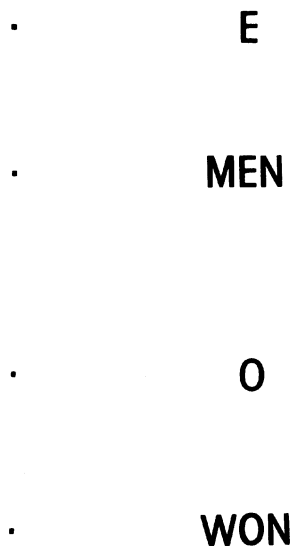
·   E

·   MEN

·   O

·   WON

FIG. 5. "Eye chart" for indirect vision with isolated and embedded letters.

of an E by itself and between M and N, likewise an isolated and an embedded O. In each case the associated dot should be used as a point of fixation. After inspecting the figure, notice that the indistinctness of the embedded E and O is not a matter of insufficient basic visual acuity; for, if one fixates the letters instead, the dots are sharp.

The present model seems to shed some light on these observations. In particular, the notion of finite processing capacity in the MSU suggests that the MSUs that are "speaking" about an embedded letter and its environs may be unable to describe "cleanly" a situation of this degree of complexity; whereas, when the letter exists in isolation, the situation is inherently less complicated and the processing machinery may be more adequate to the task of description.

The above is rather vague; but the problem is subtle. Up to a point one might say it is just a matter of "resolving power": the embedded letter is harder to separate from its neighbors; without neighbors, we "see it better". But a more powerful notion, that the MSU is a primitive pattern recognizer, is suggested by the work of Hubel & Wiesel. In the hypercolumn they found cells sensitive to spots, cells sensitive to oriented bars, cells sensitive to a properly oriented bar anywhere within a relatively large area, etc. Let us assume that these stereotyped single-cell responses represent aspects of the MSU's calculation of its output message, and that the message itself is drawn from a quite limited vocabulary. Then, if the MSU sees a stimulus which has simple properties of the appropriate type, its output message will be strong and clear: e.g. "there is a corner", "there is a round patch", etc. But, if the stimulus is not easily summarized in the terms the MSU is designed to detect and express, its message will be indecisive—an outcome whose conscious consequence is the belief that we "can't see the form clearly".

The concept of the MSU as a primitive pattern recognizer implies that it is an information-reducer: the MSU operates on a stimulus so as to describe it in certain simplified or stylized terms. This leads to a further question. If, by detecting and talking in a language of simple forms, the MSU acts as an information-reducer, then for every output message there must be a set of physically distinct stimuli which will all produce the same output. Put another way, are there ways to vary a stimulus seen indirectly such that one would not notice any change? And, what variations produce pronounced changes? Experiments to answer these questions would give further clues as to the forms for which the MSU is looking.

Luminance gradients

To test and extend the ideas developed so far, we shall consider the major phenomena in a rather different area, that of the visibility of sinusoidal luminance gradients. Here, it is *not* always the case that the appearance of an object is independent of distance or subtended visual angle. Instead, there appear to be two regions: for objects of "low" spatial frequency, appearance and threshold visibility are independent of retinal image size; while for objects of "high" spatial frequency, retinal image size matters. Quotation marks have been used because the dividing line is not well defined in the literature; an alternative viewpoint will be offered here in which "spatial frequency" turns out not to be the central variable. In the process, the investigation will identify a widely useful characteristic parameter of the mapping.

A number of people (see McCann, Savoy, Hall & Scarpetti, 1974) have investigated the "high" frequency region. McCann and his co-workers (McCann *et al.*, 1974; McCann, 1978) extensively studied the "low" region and showed that it had to be treated quite differently. There are many experiments and the subject is conceptually tricky to navigate. For simplicity we shall describe a typical experiment and use it in explaining both the phenomena and the new viewpoint.

In general, subjects look at displays or "targets", usually square, in which the luminance varies as a sinusoidal function across the target in, say, the horizontal direction. Figure 6 shows the luminance profile of a typical target. The target may have a uniformly luminous surround: sometimes the surround is black, sometimes it is equal in luminance to the average of the sinusoid, etc. "Contrast" in the display is
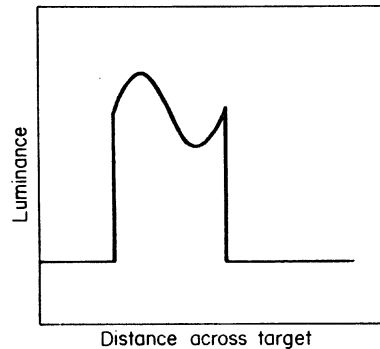
FIG. 6. Luminance profile of a sinusoidal target.

defined by McCann as $(L_{max} - L_{min})/L_{max}$, where the Ls are luminances. "Spatial frequency" means the rate of variation of the target luminance, in cycles per degree measured at the eye. "Retinal gradient" is defined as contrast divided by the retinal angle between L_{max} and L_{min}. "Number of cycles" is just the number of cycles through which the luminance varies in going across the display; it need not be an integer.

Now, suppose in our experiment we use a fixed target size and average luminance, but allow ourselves to change the target's contrast, its number of cycles and the distance at which we view it (angular size on the retina). We will begin with a low number of cycles, say 0·5, set the contrast to zero, and then gradually increase contrast until we just see the luminance pattern. This contrast value is the "threshold contrast" and its reciprocal is "contrast sensitivity".

We repeat the experiment at a different viewing distance and, perhaps surprisingly, find no difference in threshold contrast. We realize that we might have guessed this result: the ensemble of messages output by the MSUs is concerned with the pattern of an object, and is independent of the object's size. Certainly this target, once it is visible above threshold, *looks* the same at any distance. It would be natural to expect the threshold value to be independent of distance, as well.

We now increase the number of cycles in the target, using 1, 2, 3 and 4 cycles. In each case we find a threshold independent of viewing distance, but the threshold becomes lower as the number of cycles increases. Contrast sensitivity is increasing with number of cycles. However, as we proceed above about 4 cycles, contrast sensitivity stops increasing; instead, it plateaus for a while and then begins a steady decline as number of cycles goes higher and higher. Furthermore, we find a point somewhere above 4 cycles where, for any given number of cycles, contrast sensitivity falls as we move away from the target, instead of staying constant as before. Figure 7 shows the contrast sensitivity curve we would typically have obtained in the experiment. The widening band at the upper end shows the effect of distance; the lower lines correspond to the observer being farther from the target.

How, in the light of the retino-cortical mapping, can we explain this overall curve? Two new concepts are needed. The first is "object frequency", which will be represented by f_o. Object frequency means number of cycles per object and is intended to indicate the amount of detail of an object independent of object size, viewing distance, etc. For these targets, object frequency is equivalent to number of cycles. (A more complex target or object would have an "object frequency spectrum", based on the Fourier spectrum of a geometrically similar object whose longest dimension was unity.)