MAPPINGS OF THE CEREBRAL CORTEX

Thesis by

George John Carman

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1990

(Submitted May 23, 1990)

ACKNOWLEDGEMENTS

This thesis describes research conducted under the auspices of David Van Essen of the Division of Biology at Caltech. His confidence in and support of my work made the completion of this research possible, and his understanding of the visual system has provided a strong foundation upon which to build my own. He has also given me the freedom to pursue as many as four separate lines of research while at Caltech, providing me with an exceptionally broad and diverse graduate education in neurobiology and computational theory. Several members of my Thesis Committee have also promoted the development of this work, and I wish to thank Mark Konishi, Jim Bower, and John Allman of the Division of Biology for their respective contributions on three unrelated projects. Likewise, Karl Herrup of the E.K. Shriver Center, Waltham, Massachusetts, has provided both the substance and the inspiration for a fourth project. However, none of this would have been possible without the perseverence of several colleagues who joined me in these investigations. It was my priviledge to work with the follwing individuals (in chronological order): Michelle A. Mahowald, Michael M. Walker, Andy K. Lee, Jessica Hopfield, and Bassem Mora. Thanks are also due to David Bilitch and Kathy Tazumi for their respective programming and technical contributions to this work. Finally, I would like to thank the colleagues at Caltech and elsewhere who provided important understanding and intuition, and a critical audience for the development of ideas. These people include: Edgar DeYoe, Harry Orbach, Walter Nadler, Matthew Wilson, Ralph Adolphs, Jaime Olivarria, and Jack Gallant. I extend my appreciation to all the aforementioned, with the hope that they learned as much through me as I did through them.

ABSTRACT

The mammalian cerebral cortex is organized into a variety of two-dimensional areas whose anatomical and physiological organization is obscured by the folding of the cortical mantle in three dimensions. This organization can be revealed by unfolding the cortex so as to produce a two-dimensional representation or *map* of its surface. To produce such mappings, we have developed algorithms for computational cartography which maximally preserve intrinsic geometry while unfolding the cortical surface. Our computational algorithms were used to produce the first computational maps of the entire primary visual cortex of the macaque monkey (Carman and Van Essen, 1985), and the first completely noninvasive mapping of *in vivo* human visual cortex (Carman and Mora, 1989).

In order to measure the geometry of the region of cortex to be mapped, a reconstruction of a surface or layer of cortex must be obtained from a typically sparse sample of contours of section. We obtain a solution to this reconstruction problem by computing a *flow* which fuses pairs of images containing successive contours of the surface. These flows are governed by a pair of complex harmonic potentials which represent translations, rotations, and scalings which combine to produce a conformal mapping of the two images onto a third fused and interpolated image. Since these potentials are harmonic, their values over a region of the images can be computed from samples taken only along the boundary of that region by solution of the Dirchlet problem. Thus, a coarse to fine series of samples on concentric annuli, similar to the sampling of the primate retina, can be used to compute such flows at a continuum of spatial scales. A number of visual problems arising in the analysis of motion, stereo, and shape information are formally equivalent to this reconstruction problem and can therefore also be solved by computing such flows. Remarkably, the equations which determine these flows can reproduce many aspects of the topography of the first stages of the primate visual pathway, suggesting that such flows may also be computed by the mappings of the cerebral cortex.

TABLE OF CONTENTS

1. Mapping the Cerebral Cortex

1.1 Introduction

In the second century, Ptolemy described terrestrial cartography as the "survey of the whole in its just proportions" and the representation "in pictures the whole known world together with the phenomena contained therein" (Wilford, 1982). This is an apt description of the goals of cortical mapping, with the notable exception that the "world" of the cerebral cartographer consists of the cortical structures together with the sensory, associative, and motor images which they generate, and which are themselves representations of the world of our subjective perception.[1] Indeed, contemporary efforts to map the cerebral cortex are but part of a larger project which seeks an explanation of how the various levels of cortical organization all combine to provide a substrate for the representations of our conscious experience. These various levels range from the cortex as a whole, to the *areas* into which the cortex is subdivided, and finally to the neurons that comprise each area. Accordingly, these efforts rely on both anatomical and physiological measures taken at various scales, such as the extent of the different cortical areas and the topology of their interconnections, combined with such physiological measures as the topography of the projection from the periphery onto the cortex.

---

[1] The mappings which we seek to construct are of mappings that the cortex elaborates of the sensory, associative, or motor fields. Although the term "cortical map" has been used to describe either of these in separate contexts, we will attempt to keep the distinction clear by referring to the former as a *mapping* of the cortex, and the latter as a *projection* or *image* in the sensory, associative, or motor domains.

Although the first mappings of the visual cortices were made 50 years ago (Talbot and Marshall, 1941, 1942), a rigorous theory describing the organization of the cortex still does not exist. This effectively limits researchers to phenomenological interpretation of their anatomical and physiological observations, and has precluded any significant progress in the project of emulation of biological vision. This is partially a result of empirical limitations on the number and accuracy of the observations that can be made, akin to limitations that distorted the continents on the maps of the earliest terrestrial cartographers. But this is also a consequence of a more significant limitation on our understanding of the function of cortical areas, and how these functions relate to the anatomical and physiological organization we observe. Such limitations can result in substantial misinterpretation of observations, akin to the depiction by early mapmakers of the Earth as flat and situated at the center of the universe.

The work which will be described first began with the objective of overcoming some of the methodological limitations on the construction of maps based on empirical observations of the cortex. However, we soon found it necessary to consider the theoretical functions of the very cortex which we sought to map. We were compelled to do so by the isomorphism between computational problems which arose in the course of the work and certain "unconscious processes" of vision. Analysis of this isomorphism led to elucidation of hypothetical functions for the earliest visual areas, and for a link between these functions and the projection of the visual field through various stages of the visual pathway. These theoretical considerations suggest how the solution to the aforementioned computational problems should proceed, and are presented as such in the context of the aforementioned mapping endeavor. This hypothesis for the relation between the function of cortical areas and their anatomical and physiological organization is further discussed at the end of this work, in order to make clear how it leads to a unified view of many fundamental aspects of cortical organization, such as the characteristic two-dimensional laminar organization seen throughout the cortical mantle.

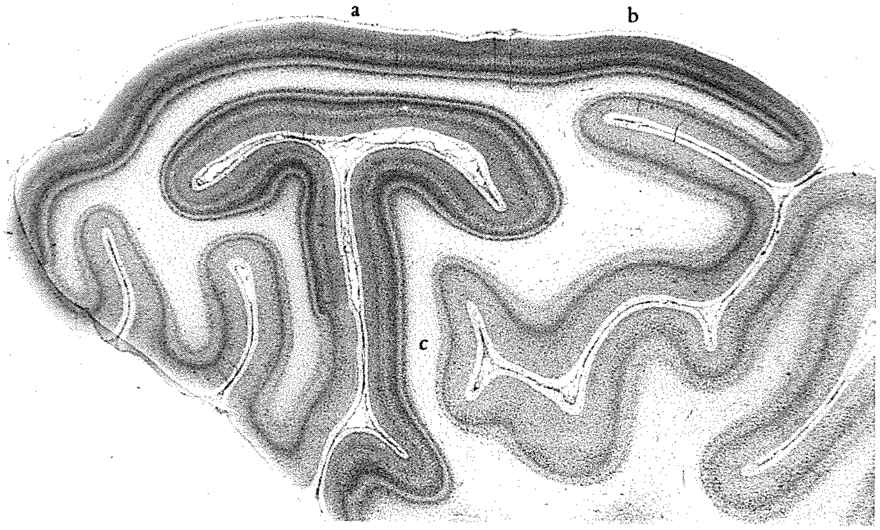


Figure 1. Laminar organization of the cortex as seen in a parasaggital section through macaque cortex.

It has been known for some time that the cortex consists of a series of laminae defined on the basis of anatomical criteria (Figure 1). These laminae are considered either as separate layers, or collectively as a single thick layer which overlies the white matter and subcortical structures. From this perspective, the layer(s) of interest are intrinsically two-dimensional surfaces which happen to be embedded within a three-dimensional cortex. In some species, the cortex is highly folded into a series of sulci and gyri, presumably reflecting constraints of packing and connectivity which determine its shape during ontogeny and phylogeny (Richman *et al.*, 1975). In such species, which include most primates and humans, the organization of the various cortical subdivisions is obscured by these convolutions. For example, when anatomical or physiological data are combined to delineate the extent of a cortical area, these areas are frequently represented as the shaded portions of a standard series of histological sections. The spatial distribution of the cortical area cannot be determined directly from such an image. The cortical surface must first be mentally "reconstructed" in order to appreciate the extent and location of the area on the brain surface (Figure 2). Similarly, the need to generate this reconstruction when anatomical or physiological data are presented on such parallel sections obscures the anatomical and physiological organization both within and between cortical areas.

These difficulties can be circumvented by "unfolding" the cortex so as to produce a map of the cortical surface of interest upon which such anatomical or physiological observations can be projected. Perhaps the simplest way to obtain such maps is to literally unfold the cortex by removing the underlying white matter, leaving a flexible sheet of gray matter which can be flattened and then sectioned tangentially (Woolsey and Van der Loos, 1970; Imig and Brugge, 1978; Tootell and Silverman, 1985; Olavarria and Van Sluyters, 1985). While such physical unfolding has been achieved for small regions of the cortex, such as individual areas of the macaque cortex, it is not well suited for mapping extensive regions such as the collection of cortical areas comprising the visual

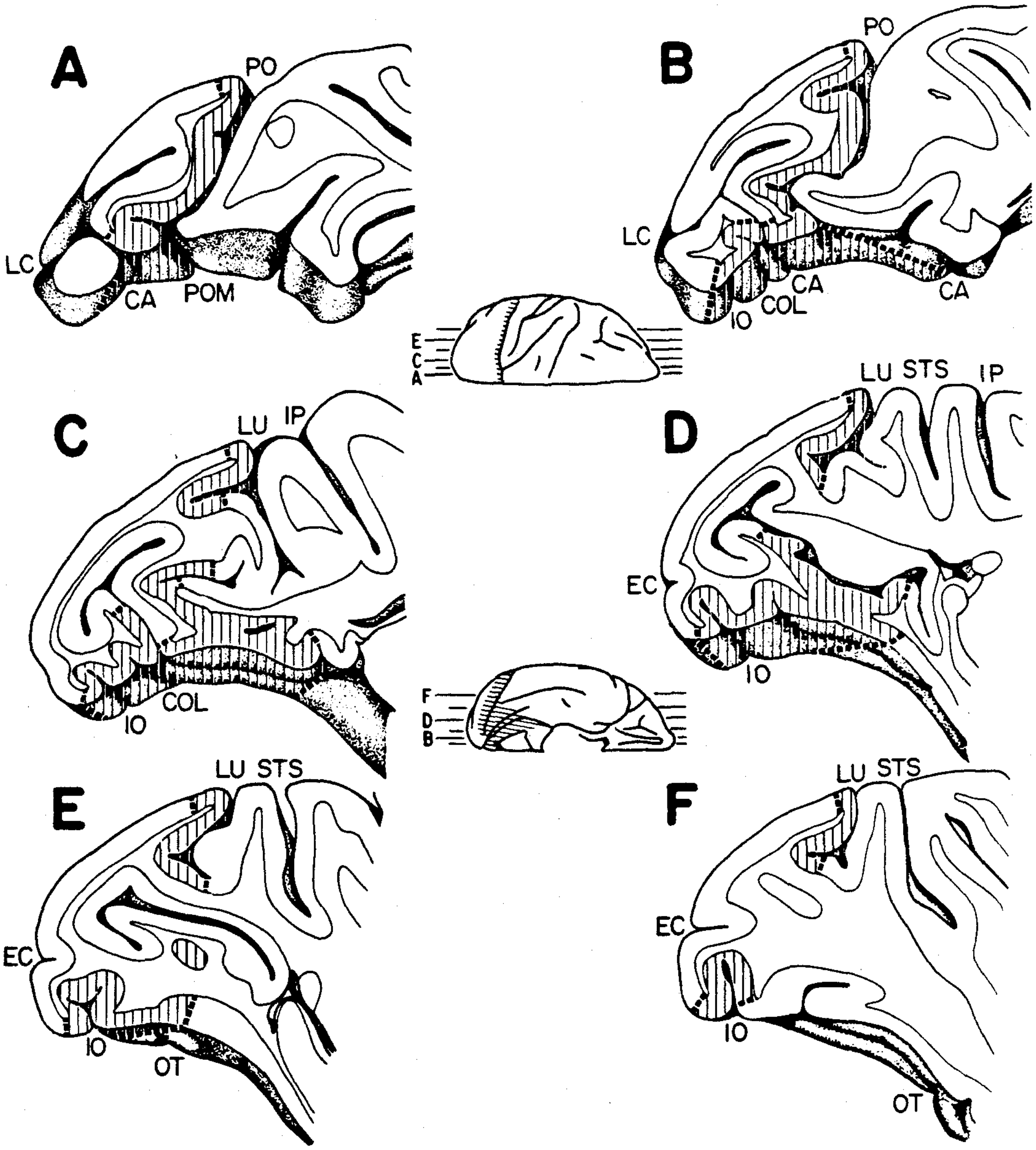


Figure 2. The extent of the second visual area (striped) represented on serial sections. From Gattass *et al.*, 1981.

cortex of primates. Typically, the cortex becomes torn during the unfolding, while the laminae remain significantly nonplanar due to residual intrinsic curvature. When the unfolded cortex is subsequently sectioned tangentially, the tears produce discontinuities in the unfolded surface, while the residual curvature precludes obtaining homogeneous sections of single laminae. In addition, the technique is difficult, being used relatively infrequently in comparison to the conventional parallel-plane sectioning of the intact brain. Finally, the physical unfolding of cortex can only be done postmortem, rendering it unsuitable for *in vivo* functional mapping in subhuman or human primates, made possible by the advent of noninvasive anatomical and physiological imaging techniques, such as nuclear magnetic resonance imaging and positron emission tomography (Mora, Carman, and Allman, 1989). For these and other reasons, one would like to be able to obtain such mappings without having to rely on a physical unfolding process.

1.2 Earlier Techniques

Despite these limitations, the physical unfolding of cortex demonstrates that such mappings can be obtained. This is possible because the global geometry of cortex is more like that of a plane than that of a sphere, in the sense that the cortex has low *intrinsic curvature* (Van Essen and Maunsell, 1980). Intuitively, surfaces of zero intrinsic curvature, such as the side of a cylinder, can be mapped onto a plane without distortion, while surfaces with nonzero intrinsic curvature, such as the sphere, must be distorted in such mapping (Gauss, 1827; Coxeter, 1961, Ch. 20; Hildebrandt and Tromba, 1984, pp. 69–70). To the extent that the sulci and gyri arise from *folding* transformations of the cortex, they do not contribute to the intrinsic curvature of the surface and can be removed by unfolding without introducing distortions.[2] However, once this is done, there

---

[2] Mathematicians prefer to use the term *bending* to describe the process by which a flat sheet would be deformed to resemble the surface of the cortex (see, for example, Hilbert and Cohn–Vossen, 1952). In this precise usage, bending refers to any transformation of a surface which leaves all arc lengths and angles on that surface invariant, and which therefore leaves the intrinsic geometry of the surface invariant. However, neuroscientists conventionally use the term *folding* in this context, and this convention will

will remain deviations of the cortex from planarity due to its small intrinsic curvature, and which will result in distortions in the course of any mapping onto the plane. Such intrinsically curved regions are known to occur in the macaque cortex, such as the regions where gyri or sulci merge, as evidenced by the fact that incisions need to be made in such regions in the course of manual unfolding (Olavarria and Van Sluyters, 1985). Since the intrinsic curvature is approximately zero elsewhere (Van Essen and Maunsell, 1980), we conclude that the curvature is not constant, but varies as a function of cortical location. Hence, we expect that distortions arising from such intrinsic curvature will also vary from one region of cortex to the next. Accordingly, any method for obtaining an unfolded mapping of the cortex should be able to measure such distortions and, when they must occur, attempt to minimize them in a controlled fashion (Gilbert, 1974; Bassett, 1972). This could be achieved either by distributing the distortion over large regions, or by concentrating the distortion in a few small regions, depending on the purpose for which the map is produced.

A number of previous investigators have attempted to obtain approximations to such unfolded maps with varying degrees of success. The simplest of these is the perspective projection of exposed, relatively smooth surfaces of the brain, as illustrated in Figure 3. This method is of limited use in that it only preserves the component of distances parallel to the plane of projection, resulting in a

---

be adopted here (see, for example, Van Essen and Maunsell, 1980). Nevertheless, it is useful to keep in mind that bending of a given surface may or may not lead to the development of folds in that surface.

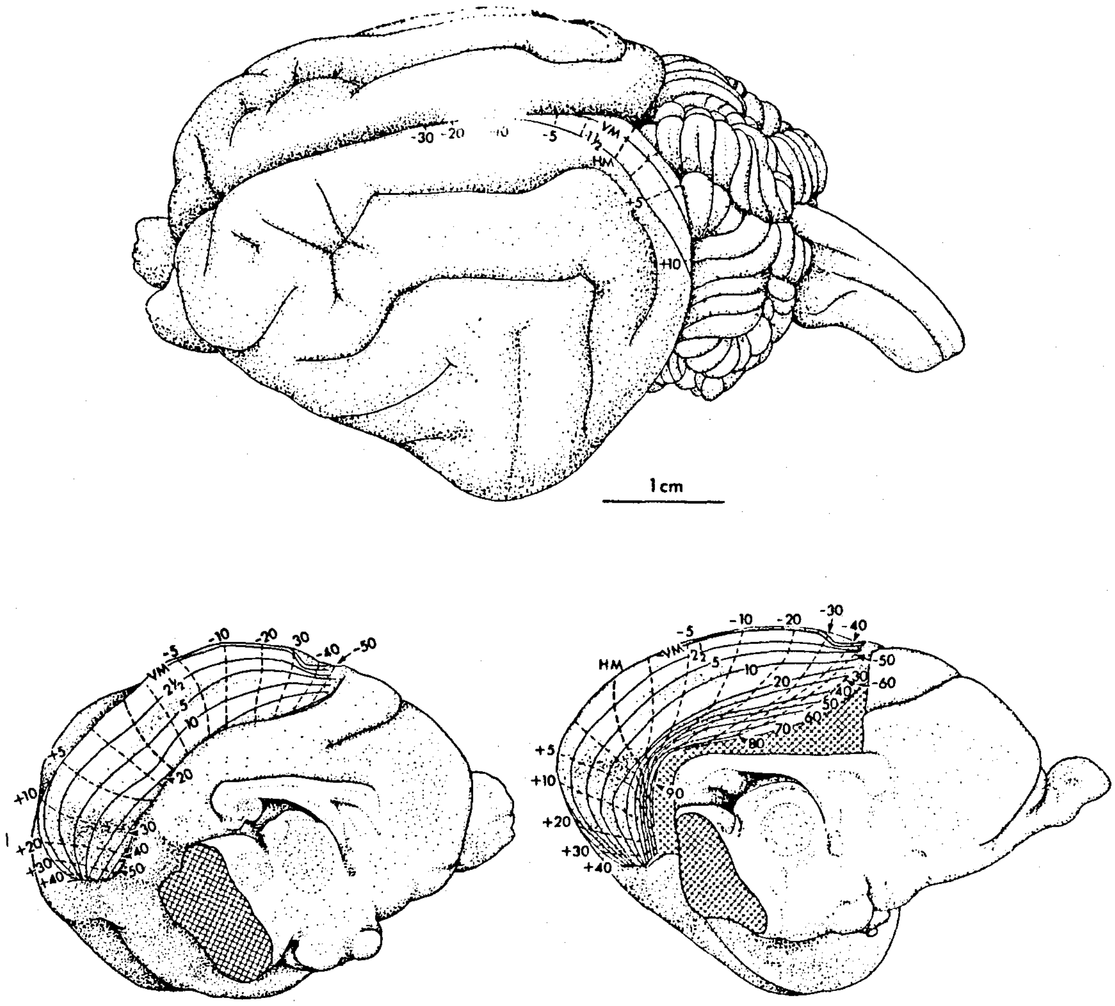


Figure 3. Perspective projection of exposed surfaces of cat visual cortex. Some portions of the cortex have been cut away in order to reveal underlying cortical surfaces. From Tusa *et al.*, 1978.

foreshortening of distances which will be a function of the particular "viewpoint" or axis of projection, as well as the shape of the surface being projected. In addition, information buried within sulci or beneath gyri can only be revealed by cutting way the obscuring portions of cortex. Since these regions may be part of the cortex which is of interest, some investigators have made use

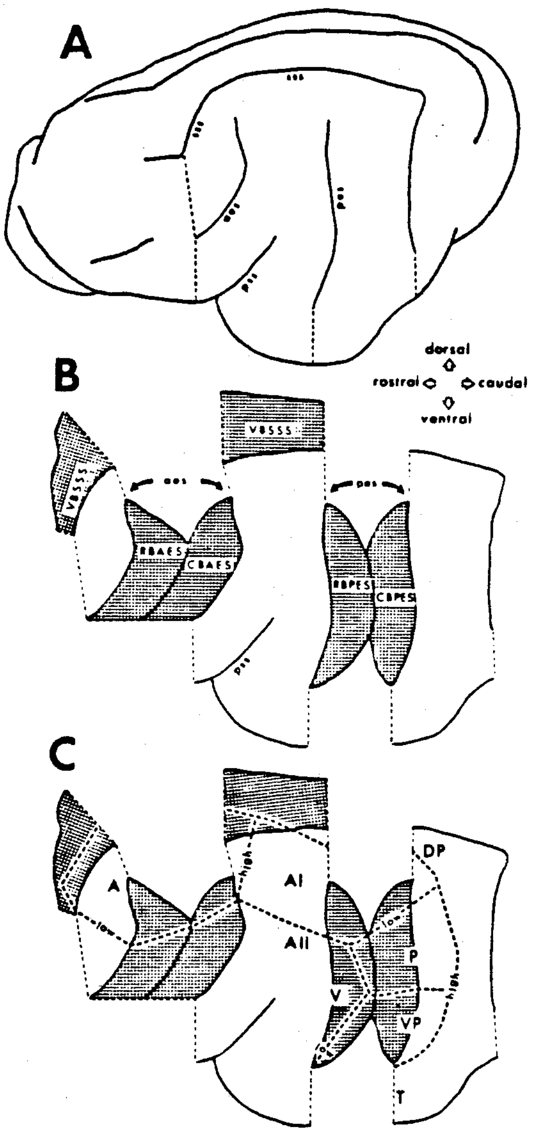


Figure 4. Schematic unfolding of surfaces of cat visual cortex. Regions which were hidden within sulci have been shaded. From Reale and Imig, 1980.

of schematic unfolded maps, as illustrated in Figure 4. Such schematic unfoldings typically preserve topology and geometry along some contours, but only at the expense of introducing discontinuities along others. Also, such schematic unfoldings approximate large regions of the cortical surface by flat polygons, thereby introducing distortions like those mentioned for the case of orthogonal projection above. Related to these are the "brainprints" illustrated

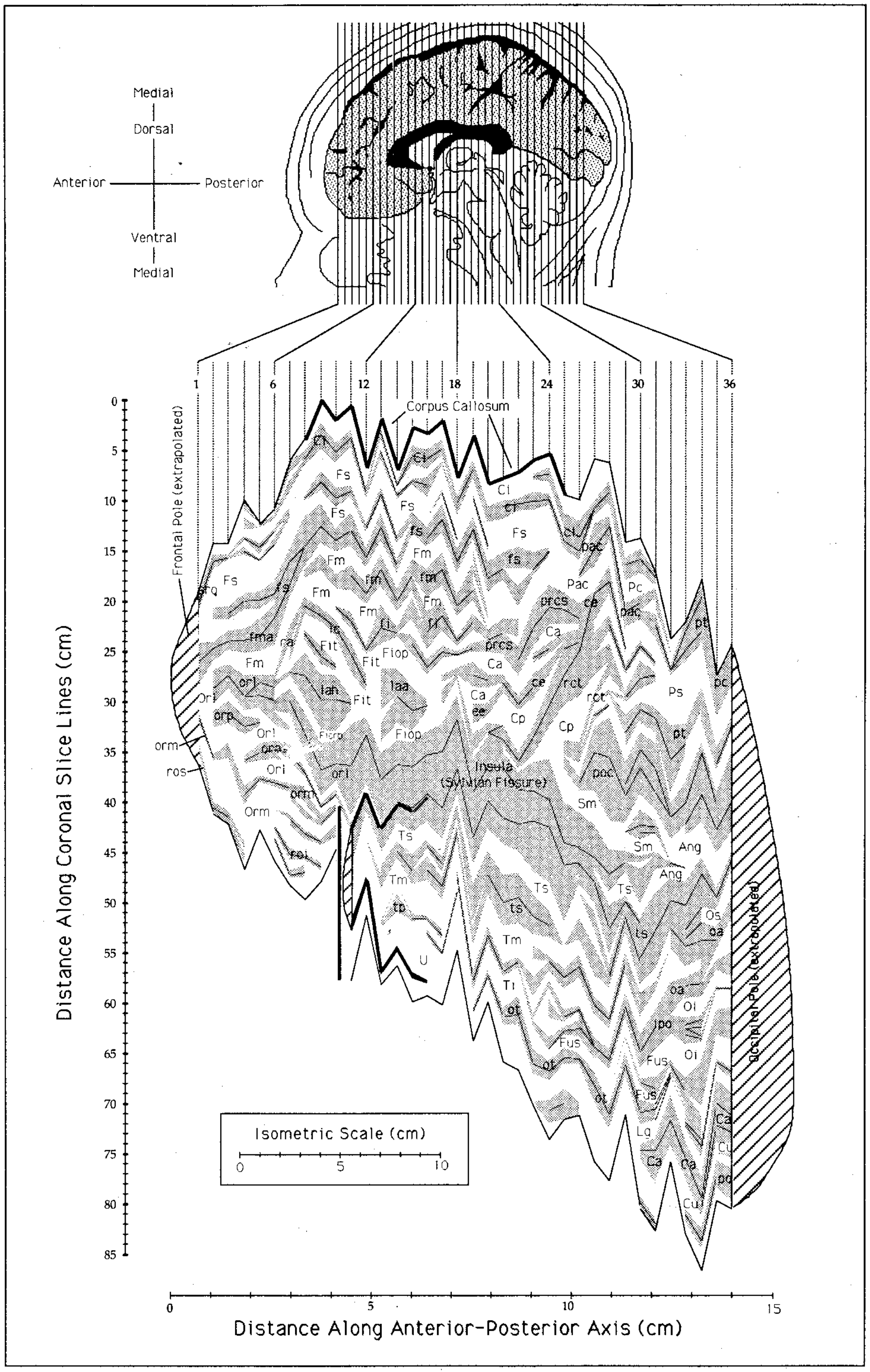


Figure 5. 'Brainprints' of human cerebral cortex. Regions which were hidden within sulci have been shaded. From Jouandet *et al.*, 1988.

in Figure 5, which preserve distances along, but not between, contours of section without introducing as many discontinuities in the mapped areas. However, the anisotropic preservation of length produces large distortions of geometry on the map as an unavoidable consequence when this technique is applied to the highly folded surface of the human brain. All of these methods share the major disadvantage that distortions are anisotropic and not quantified in most cases, making it difficult to determine the accuracy of representation of any given region on a particular type of map.

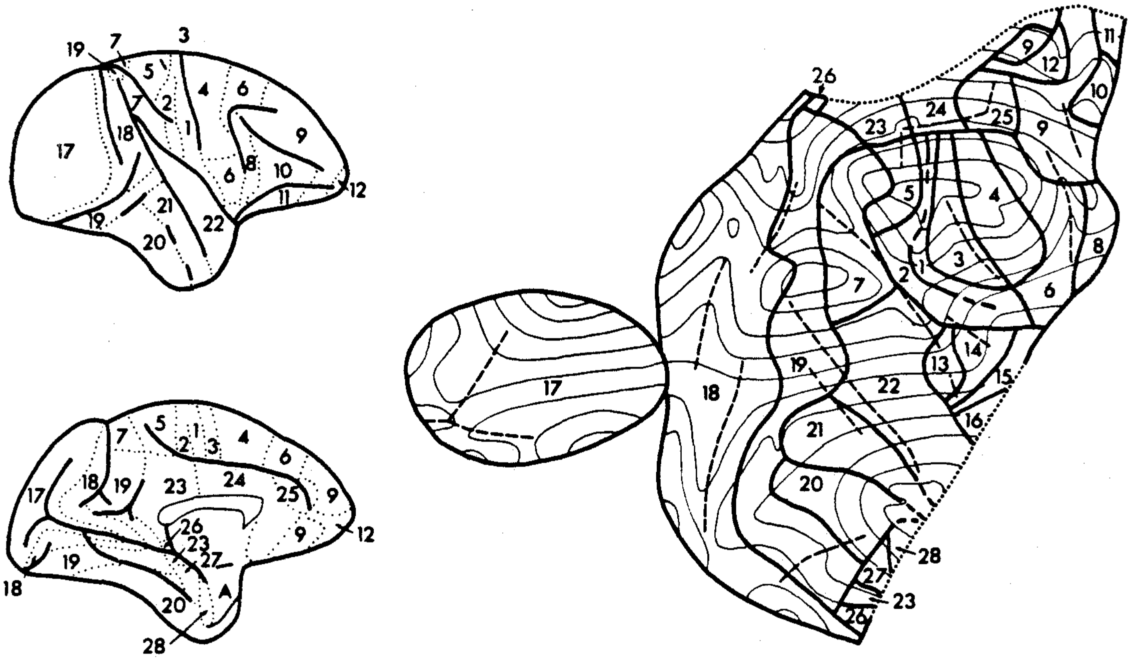


Figure 6. Two–dimensional maps of macaque cerebral cortex. On the left are lateral (top) and medial (bottom) views of surface of the hemisphere which is mapped on the right. Dark lines on the map represent boundaries of cortical areas, light lines the contours of section used in producing the map, and dashed lines the fundi of sulci. From Van Essen and Maunsell, 1980.

In an effort to overcome these limitations, Van Essen and Zeki (1978) and Van Essen and Maunsell (1980) developed a manual technique for the construction of two-dimensional, unfolded representations of the cerebral cortex (Figure 6). In this technique, the maps are produced by the iterative repositioning of the contours of section of a specific lamina of cortex obtained from histological sections (Van Essen and Maunsell, 1980, Figure 13). During this repositioning, the length of contours is kept constant while their shape is allowed to change, with some effort being made to keep the spacing of contours on the map comparable to their true spacing on the cortex. In addition, by tracking the distortions of small lengths, angles, or areas during the mapping, the technique enables one to quantify distortions as a function of position on the maps.

While this manual technique demonstrates the feasibility of producing such two-dimensional maps of the folded surface of the cortex from the contours of section, it has several shortcomings. The maps are tedious to construct, often taking several days to map a region the size of macaque primary visual cortex (1200 square mm) and weeks to map an entire cerebral hemisphere (8,000 square mm). They are also inherently limited in their acuracy, since they rely on manual techniques prone to cumulative errors, and since measures of geometry to be preserved are inherently anisotropic. Finally, since these maps are typically made without reference to a true three-dimensional reconstruction of the cortex, they rely on some subjective judgement to provide information about the spacing of contours of section from one region of cortex to the next, and therefore show variability between maps of the same tissue prepared by different cortical cartographers. Despite these shortcomings, the manual technique represented a substantial improvement over previous methods, and provided a strong foundation for the further development of the technique. That the manual technique involved little more than the iterative repositioning of contours in conjunction with preservation of local geometry suggested that computers could be programmed to perform the task by making use of variational or optimization methods. While the prospect of developing computational algorithms for the production of unfolded maps of the cortex appeared promising, it also raised questions as to what measures of local geometry should be preserved in such mappings. We now turn to a review of the properties which define the various mappings.

1.3 Preservation of Topology and Geometry

Perhaps the most familiar mappings are the various mappings of the Earth's surface (Richardus and Adler, 1972; Gardner, 1975). These maps exhibit several properties which are desirable for mappings of the cortical surface as well. These properties allow us to distinguish several different classes of mapping on the basis of what aspects of topology and geometry they do and do not preserve. In the present context, a mapping M can be formally described as a representation of one surface **S** on or by another surface **S'** for which every point **P** on **S** has a corresponding point **P'** on **S'**, and for which this correspondence is unique and reciprocal (Richardus and Adler, 1972). More precisely, a mapping M which transforms points **P** to points **P'** is given as:

$$\mathbf{P}' = M(\mathbf{P}),$$

while the inverse mapping is given as:

$$\mathbf{P} = M^{-1}(\mathbf{P}').$$

The simplest example of a mapping is the identity mapping I in which each point

P is mapped or projected onto itself:

$$P' = I(P) = P.$$

In order to avoid discussion of uninteresting cases, we consider that the mappings of interest are determined only to within an arbitray global translation, rotation, and scaling.

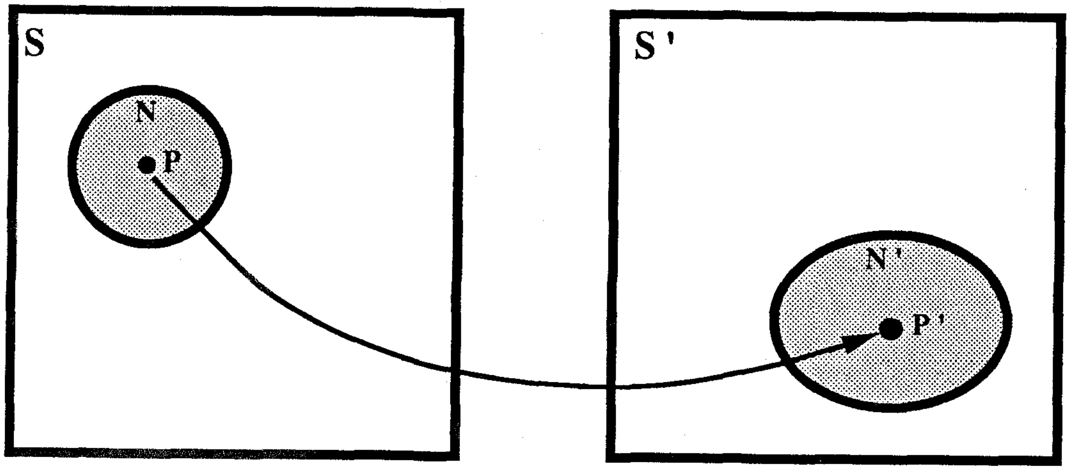


Figure 7. Continuous or topological mappings place points in the set S in one to one correspondence with points in the set S'. Thus each point P and its neighborhood N in set S are mapped to the image P' and its neighborhood N' in set S'.

Mappings of the cortex should be *continuous*, meaning that if P maps to P', then the neighbors of P should also map to the neighbors of P'. More precisely, a mapping *M* is *continuous* if for every neighborhood N of P all points in that neighborhood are mapped by *M* into a neighborhood N' of P' (Figure 7). Mappings which exhibit such continuity are referred to as being *topological*. Such mappings are possible between any two surfaces which are *topologically equivalent* by virtue of being of the same genus, that is to say, for surfaces which have the same number of bounding curves (Alexandroff, 1961). We note that to say that a mapping is topological amounts to a restriction on the ordering of points so as to preserve neighborhoods in the course of mapping. Thus a continuous mapping does not require the preservation of any geometry *per se*.

The property of continuity may also be shared by several more restrictive classes of maps, which preserve local measures of area, angle, or length.

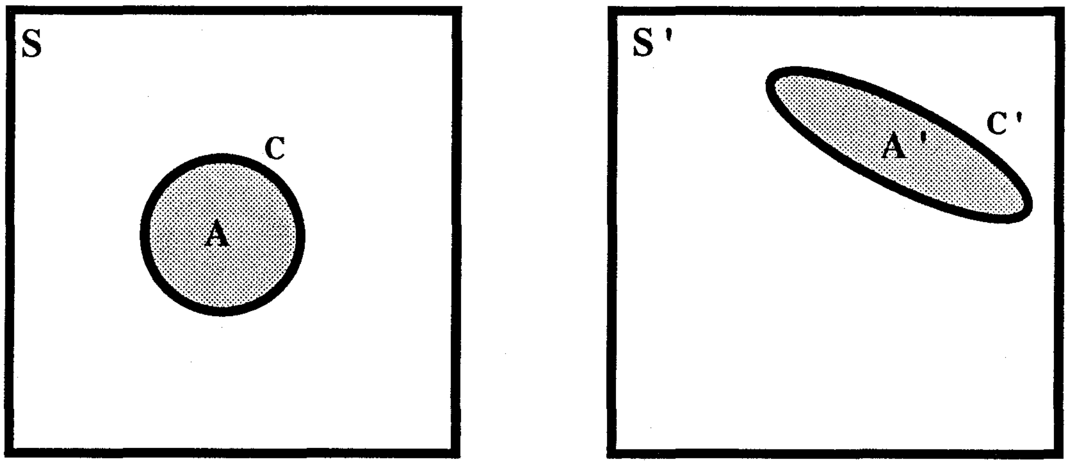


Figure 8. Area–preserving mappings project the region A bounded by contour C and having area a to the region A' bounded by contour C' and having the same area a. Note that lengths, angles, and shapes need not be preserved.

Of these, *area–preserving* mappings are the least restrictive, requiring only that the area bounded by any closed curve be preserved in the area bounded by the image of that curve (Figure 8). Clearly, however, one can stretch or squeeze local regions while leaving the area invariant, so that lengths and angles will not in general be preserved by such mappings. If we instead stipulate that the angle between any two curves be preserved in the angle of the image of those curves, then we obtain a *conformal* mapping (Figure 9). As the name suggests, the

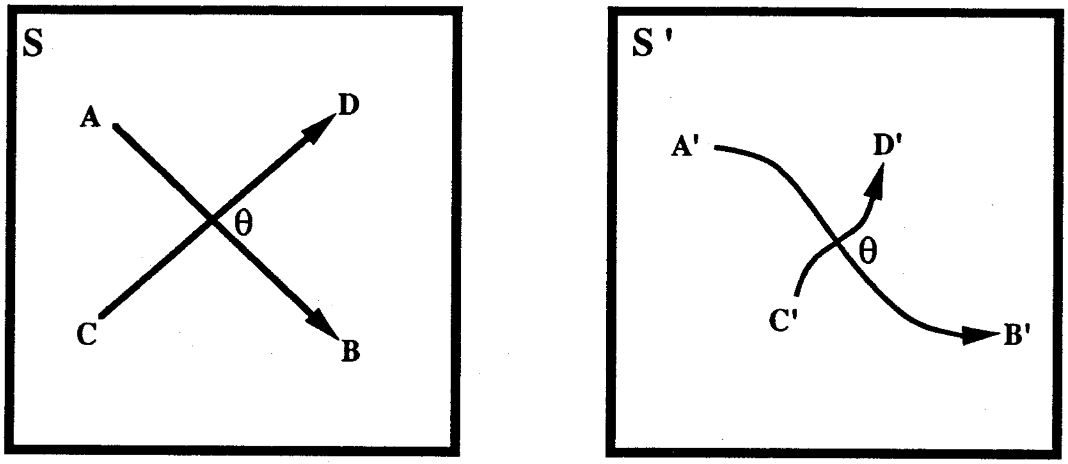


Figure 9. Conformal mappings project two curves AB and CD intersecting at an angle Θ to the curves A'B' and C'D' intersecting at the same angle Θ. Note that lengths and areas need not be preserved.

shapes or conformations of sufficiently small figures carry over to their images under such mappings. Alternatively, the shape or conformation of a figure will be preserved approximately and in a manner which improves as the size of the figure decreases. One can see this by considering a triangle: if all three angles are to be preserved under the mapping, then the triangle must be similar to its image, so that the ratios of lengths of the sides of the triangle will be preserved. Conformal maps are valued for this property, even though they do not necessarily preserve lengths or areas. Also, conformal mappings will by definition preserve the orthogonality of a coordinate system, such as the Cartesian grid in the plane, permitting conformal mappings to be described geometrically as transformations of a given coordinate system. Alternatively, if we stipulate that the length of any curve be preserved in the length of the image of that curve, we obtain an *isometric* mapping (Figure 10). This mapping is more restrictive than either of the earlier mappings, as can be seen by again

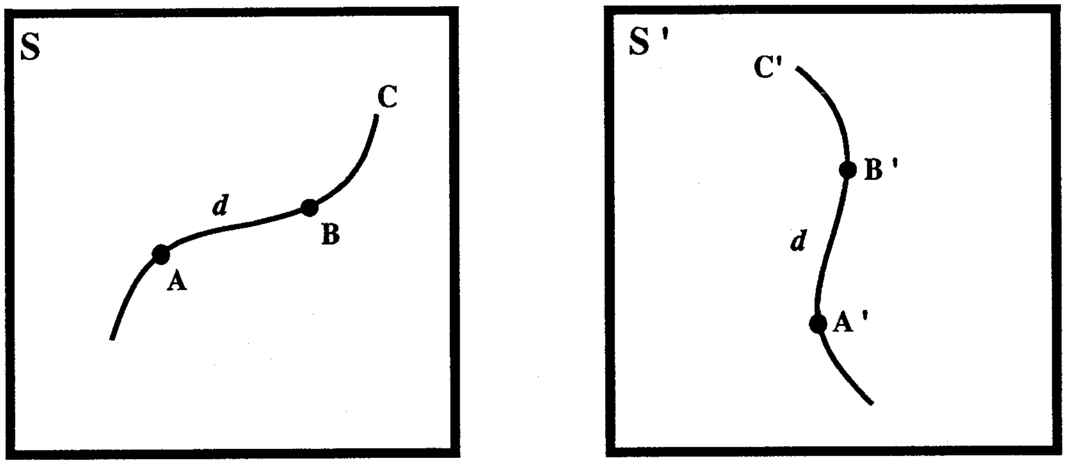


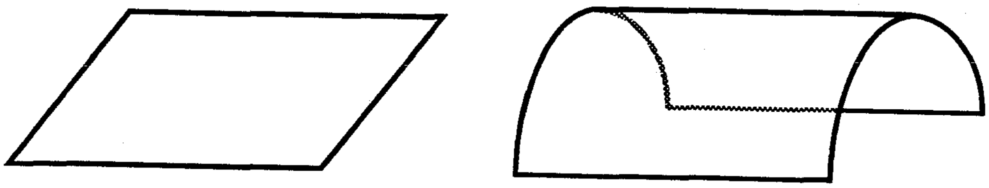
Figure 10. Isometric mappings project points A and B separated by a distance *d* along curve C to the points A' and B' separated by the same distance *d* along the curve C'.

considering a triangle: if the lengths of all three sides are to be preserved under the mapping, then the angles must all be preserved as well. In this sense, the isometric mapping can be regarded as the particular case of the conformal mapping having unity magnification. Finally, the *geodesic* mappings may also be distinguished. These map the geodesics, which are the curves of minimal length, of one surface onto the geodesics of another. Since they preserve a subset of distances, they can be regarded as a degenerate case of the isometric mapping, which we have already considered. Other properties of all these mappings are discussed in Hilbert and Cohn–Vossen (1952) and Richardus and Adler (1972).

Note that in all these cases we regard the areas, angles, or distances as being measured with respect to some some curve(s) on the surface being mapped. This enables us to measure the intrisic geometry of a surface regardless of how it may be embedded in a space of higher dimension. This intrinsic geometry will be invariant under the various possible foldings or *embeddings* of the surface in the larger space. For example, the distances along a curve on a two-dimensional surface embedded in a three-dimensional space will not depend on the way in which the surface is folded. Such intrinsic distances will therefore be preserved when the surface is unfolded in the course of mapping. In contrast, the Euclidean distance between any two points along the curve, as measured along a straight line connecting those points, will in general depend on the details of how the surface is embedded in space. Such extrinsic distances will change as the surface is unfolded, and are thus not suitable measures of geometry to preserve during mapping. The problem of finding a set of intrinsic measures of distance on a discrete polygonal surface is known as the *discrete geodesic problem*, and has been treated by Mitchell *et al.* (1987). In general, we would like to preserve as much of the intrinsic topology and geometry of a surface as possible in its mapping. We will now consider how continuity, area, angle, and length can each be preserved at least for subsets of a typical cortical surface when mapped onto the plane.

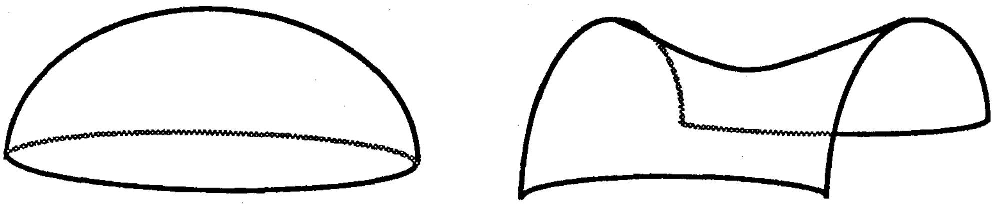
To show that there exists a continous mapping of such a cortical surface onto the plane, it is sufficient to note that the two are topologically equivalent. In particular, the surface of cortex and a region of the plane bounded by a single contour are both surfaces of genus zero having a single hole. As a consequence, one surface can be transformed into the other via *deformations* involving stretching or shrinking, but not cutting or gluing of portions of the surface. This can be accomplished by an imaginary "unfolding" of the cortex, so that its surface deforms into a sphere with a single hole, and then "flattening" this surface into a subset of the plane bounded by a single contour. Since such deformation can be accomplished without introducing discontinuities, it is by definition an example of a continuous mapping. Thus, as a consequence of their topological equivalence, it is possible to obtain a continous mapping of the cortex onto the plane. Such mappings can be obtained in practice by requiring that each point on the surface of cortex have a corresponding point on the map, and that points which are neighbors on the cortex are likewise neighbors on the map. Since continuity is one of the most fundamental properties sought for our maps, this operation will in fact become one of the steps of our mapping procedure.

Similarly, we can imagine producing an area–preserving mapping of the cortical surface on the plane as follows. Construct a model of the surface of interest using a uniform layer of perfectly deformable but incompressible material (something like plasticine comes to mind). Now carefully deform this model onto the plane so that it does not tear or break. Each unit area will be able to change shape as needed to allow the surface to deform onto the plane, although some angles and lengths may be changed in the process. However, since the material is incompressible, the total area and the area of each of the subregions will be preserved. More formally, we can show that an area–preserving mapping of the cortical surface onto the plane can be constructed as follows. We start by measuring the entire area a of a surface S enclosed within a boundary contour C. We create a map of this surface by drawing an arbitrary contour C' on the plane which encloses a surface S' having the same area a. Now we partition the surface into two disjoint regions and measure their respective areas a_1 and a_2. We proceed to partition the area of our planar map into two regions with areas a_1



Planar: $K_g = 0$, $K_m = 0$

Parabolic: $K_g = 0$, $K_m \neq 0$



Elliptic: $K_g > 0$

Hyperbolic: $K_g < 0$

Figure 11. The relationship between principle curvatures and local surface shape. The signs of the Gaussian and mean curvature are indicated.

and a_2, respectively. We now repeat this process of partitioning the surface and then the map so as to preserve the areas within each progressively smaller subdivision. In the limit of infintessimal areas a_i we obtain an area-preserving map of the original surface. Thus, by ensuring that the area of each successively smaller neighborhood is preserved, we have that the area within any closed curve whatsoever will be preserved. We can localize any point on such a map by finding the image of a small region in which the point lies, taking the region as small as we like through further subdivision until we obtain the point in the limit. Although this construction proves the existence of such an area-preserving map, we would not use it in practice since it preserves area without regard to any other aspects of geometry.

Other aspects of surface geometry, such as length or angle, can also be preserved. To help illustrate this, we develop the notion of the *curvature of a surface*, as follows. We can decompose any smooth, continuous surface bounded by a single contour into regions each characterized by the signs of the *principle curvatures* K_1 and K_2 (Hilbert and Cohn–Vossen, 1952). These principle curvatures, which correspond to the maximum and minimum curvatures along curves through a given point on a surface, determine whether the shape of a surface is locally flat, parabolic, elliptic, or hyperbolic (Figure 11). The average of these principle curvatures,

$$K_M = \frac{K_1 + K_2}{2}$$

known as the *mean curvature* K_M, provides a measure which depends both on intrinsic surface properties and on extrinsic folding of the surface. In contrast, the product of the principle curvatures,

$$K_G = K_1 * K_2,$$

known as the *Gaussian curvature* K_G, remains invariant under the folding transformations which are chiefly responsible for the formation of the sulci and gyri. Thus, the Gaussian curvature provides a measure of the intrinsic curvature of a surface. As noted earlier, the cortex has little or no intrinsic curvature over most of its surface, with the exception of certain regions where the intrinsic curvature is nonzero. Over those regions where the Gaussian curvature is zero, we can obtain an isometric mapping onto the plane (Hildebrandt and Tromba, 1984, p. 69). However, since the Gaussian curvature differs from zero in at least some regions, and since one cannot reduce the Gaussian curvature of a surface to zero by segmenting it, we we do not expect to obtain an isometric mapping in general. Thus, we do not expect to be able to map the cortex onto the plane without some distortion. This is true even if we use discrete polygonal approximations of a cortical surface, which although locally planar, are already distorted versions of the actual cortical surface. However, for those regions characterized by a constant Gaussian curvature, it is possible to obtain a conformal mapping of that region onto the plane (Hilbert and Cohn–Vossen, 1952, p. 269). While the nonuniformity of the Gaussian curvature of the cortex makes finding a single, continuous conformal mapping of this surface impossible, such a mapping might be obtained piecewise by segmenting the cortex into regions each having approximately constant Gaussian curvature. If this process of segmentation is carried out so as to give segments which are sufficiently small, we can obtain an approximation to a variable conformal mapping of the surface of the cortex.

Regardless of the mapping selected, we will need to make measurements of surface geometry to be preserved in the course of mapping. Our ability to make such measures will, however, be constrained by the form and availability of the data. In particular, the surfaces to be mapped will in general not be continuously sampled, so that measurements of topology and geometry of the surface cannot be made directly from the sample. Instead, we will need to reconstruct the surface from the sample before such measures can be taken. The reconstruction of such the highly folded surface of the cortex from the typically sparse sample will prove to be our most difficult undertaking. Since the possible approaches to the reconstruction of the cortex will depend on the nature of the data which are available, we next turn our consideration to the question of the selection of the surface to be mapped.

1.4 Selection and Sampling of the Surface to be Mapped

In the cortex, one can generally distinguish six or more layers which span the radial distance from its surface to the boundary of the grey and white matter. These layers can each be considered to be a continuous and smooth shell of gray matter of some thickness, laid one on top of the other to comprise the cortex. While it is possible to map each of the layers separately, or to generalize the mapping to a three-dimensional volume, for most current applications it is sufficient to select only one layer for mapping. Although in principle any of the layers could be selected, usually either the exterior surface of the cortex or layer IV are chosen, as they can be consistently distinguished in a wide variety of anatomical data. By neglecting the thickness of the chosen layer, we can approximate it by a two-dimensional surface embedded within the three-dimensional cortex, providing an ideal substrate for mapping.

In order to map the chosen cortical layer, we must make measurements of the local topology and geometry of this surface to be preserved in the course of mapping. To do so, we would ideally select a continuous and isotropic sample of sufficient density to determine the chosen surface at a desired resolution (Gabor, 1946; Shannon, 1948, 1949; Landau, 1967; Logan, 1977; Marr *et al.*, 1979; Requicha, 1980; Curtis and Oppenheim, 1987; Marvasti, 1989). If such sampling of the surface were possible, we could determine the topology and geometry as needed for mapping from direct measurement of sets of points in the sample. However, the chosen surface cannot be sampled in this manner, since in almost all cases the surface lies within an opaque volume of tissue. In order to expose the surface of interest, the volume of tissue must be "sliced" in some manner, and a subset of points sampled from the *contours* defined by the intersection of this surface with the planes of section.

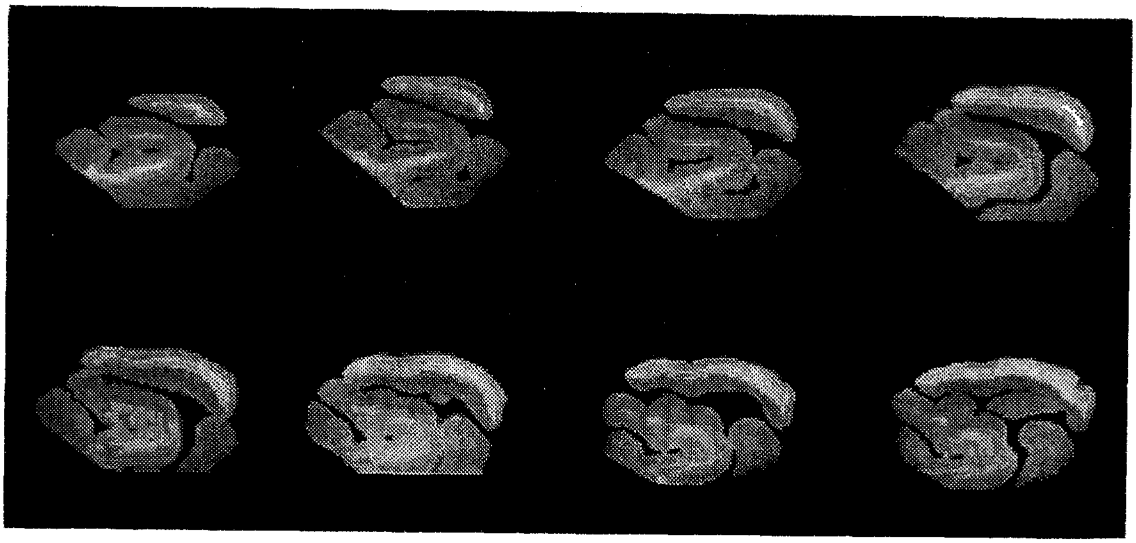


Figure 12. Histological sections of macaque cortex obtained by microtomy. Sections are separated by 500 microns on average.

Examples of such anatomical data include a series of histological sections obtained by postmortem microtomy of monkey cortex (Figure 12), or a series of noninvasive magnetic resonance images of human cortex obtained *in vivo*.

Since isotropic measurements of topology and geometry of the surface cannot be made directly from such contours, we must reconstruct a *reference surface* upon which such measurements can be made for the purpose of mapping. But in general, not all contours of section from a surface will be available for this purpose, due to technical limitations on the number and spacing of usable sections which can be obtained from a given brain. For example, histological processing of brain sections can result in tissue shrinkage and breakage, both within and across sections. Even when all sections are in principle available, histological processing can differentially distort adjacent sections, making the continuity of limited value for the reconstruction of the surface (Van Essen and Maunsell, 1980). But more typically, different subseries of sections undergo various histological procedures to obtain different kinds of data, effectively limiting the sample of the surface to be mapped to those sections subjected to comparable histological processing. Even if we distribute this sample uniformly across the brain so that it consists of, say, every sixth section, there still are a number of sections separating each pair of contours in the sample. For a typical section spacing of say 30 microns, this means that planes of section are 180 microns apart -- enough distance for surface contours to undergo significant changes in shape. Similarly, the various noninvasive procedures, such as magnetic resonance imaging of the human brain *in vivo*, have technical and economic limitations which effectively limit the number and spacing of contours of section which can be obtained (Brownell *et al.*, 1982; Norman and Brant-Zawadzki, 1985; Oldendorf, 1985; Sokoloff, 1985; Fox *et al.*, 1986; Raichle, 1986). Although the advent of confocal microscopy and advances in resolution of noninvasive techniques promise to bring us closer to an ideal continuous sampling of surfaces, at present these techniques are not in general use. For all these reasons, the contours of section will in general not be strictly serial or adjacent to one another, but rather will constitute a sparse, discontinuous, and anisotropic sample of the chosen surface. These characteristics of the sample can result in ambiguities in the subsequent reconstruction of the surface, as will be described in the next section.

An additional problem arises when the alignment of contours with respect to each other is lost in the process of acquiring the data. For example, once histological sections are cut from the brain, the original alignment of one section relative to the next may be difficult to determine. However, it is possible to obtain registration of the data by translation and rotation of the contours so as to bring them into alignment with one another or to a set of independent fiducial points (Weinstein and Castleman, 1971; Levinthal and Ware, 1972; Merickel, 1988; Herbin *et al.*, 1989). This alignment procedure is facilitated through the use of three-dimensional computer graphics, which permit the manipulation of contours so as to achieve registration (Figure 13). A series of images of the cut surface of the brain which record the absolute alignment of the sections can be used as a reference for such registration, as well as for rectification of other distortions arising from histological processing. Furthermore, the noninvasive procedures such as magnetic resonance imaging and positron emission tomography typically obtain data in register, at least within a given scan. Thus the registration problem does not arise in all cases, and where it does, solutions are available. We thus assume in the following that such registration of the contours has been achieved in one way or another, so that the distances and angles between points on the contours agree with the values that would be obtained if we could make the corresponding measures directly from the surface.

1.5 Reconstruction of the Sampled Surface

Formally, we describe the *reconstruction* of a surface to be the recovery of the intrinsic topology and geometry of the surface from measures taken on a sample or projection of the surface. In most cases, the intrinsic properties of a surface cannot be directly determined from the sample, due to the loss of topology from

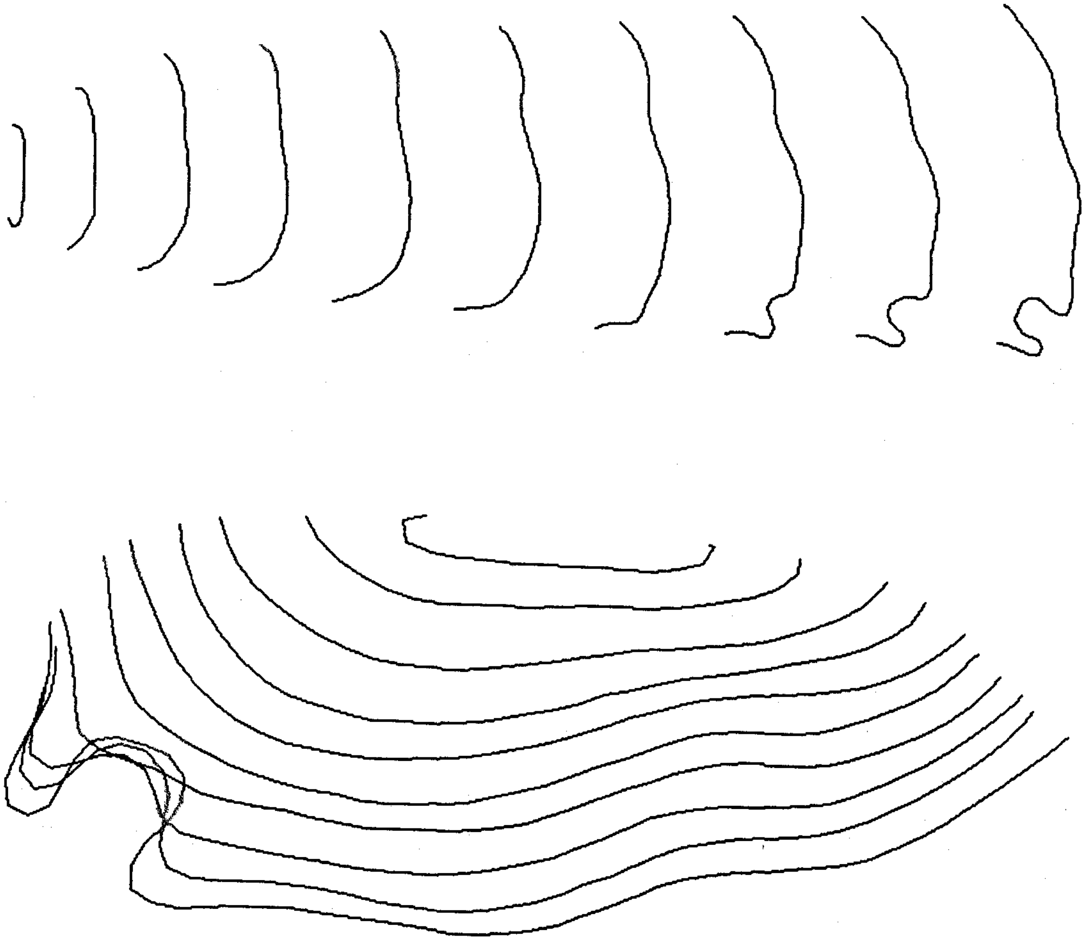


Figure 13. Contours obtained from sections of macaque striate cortex (above) are brought into registration with one another (below).

discontinuities in the sample, and due to differences between the geometry of the surface and that of the space in which the sample is embedded. We view it as the fundamental problem of vision, to which the other problems of vision can be reduced by an appropriate mapping. If this is the case, then it should be possible to solve these other problems from within a general unified theory of vision once a solution to the reconstruction problem is known (Carman 1987). A variety of approaches to the reconstruction problem have been proposed, mostly by mathematicians and computer scientists, and often with the intent of computing some visual surface representation.

We begin with a review of this literature to provide a taxonomy of surface representations for the neuroscientist interested in the general problem of surface reconstruction. This problem has been viewed alternatively as an exercise in tiling a two-dimensional surface, in fitting a bivariate function, in deforming a two-parameter model, or in constructing a mapping or vector field on a topologically equivalent domain. Regardless of the description, the representation chosen must be sufficient to capture the relevant measures of topology and geometry for surfaces of interest. Several representations of surfaces have been proposed in the context of various applications (for reviews, see: Marr and Nishihara, 1978; Srihari, 1981; Morganthaler and Rosenfeld, 1981; Marr, 1982; Foley and Van Dam, 1982; Barsky, 1983; Ritter and Tou, 1984; Samet, 1984; Brady *et al.*, 1985; DeFloriani *et al.*, 1984; Lancaster and Salkauskas, 1986; Horn, 1986; DeFloriani, 1989). Since surfaces are embedded in three dimensions, most representations make use of the familiar three Cartesian coordinates (x,y,z) to give the positions of points, with additional dimensions sometimes used to represent various attributes at each point or to permit the representation of time.

We can characterize the various representations in terms of the fields, or more generally, manifolds over which they are defined. Thus, we distinguish *pixels* defined on two-dimensional image planes from *voxels* defined on three-dimensional object volumes. In between these two extremes are found a variety of parametric surfaces, defined by way of implicit or explicit parametrizations of two-dimensional manifolds of various topology (Alexandroff, 1961; Munkres, 1984; Flanders, 1989). To distinguish the elements of such *p*arametric *s*urface representations from the alternatives, we refer to them as *psxels* (pronounced 'sik-sels). We will now consider these alternatives for surface or object reconstruction so as to provide an overview of existing approaches, as well as a running critique to motivate the development of our own representation for surfaces.

One common choice of representation is a scalar function of position on a two-dimensional field, often referred to as a *scalar field*. When this representation is used to represent surfaces in or on some computing device, the scalar field is often referred to as an *image* comprised of *pixels*, with the value of the scalar giving the *elevation* or *distance* of a surface. With this representation, the reconstruction problem can be formally described as finding a bivariate function $z = f(x,y)$ defined on a two-dimensional field (x,y) which interpolates the points sampled by our data contours (x_i, y_i, z_i), $i = 1,\ldots,n$. Examples of this approach have been developed by several authors (Grimson, 1982; Terzopoulos, 1983; Poggio *et al.*, 1985; Xu and Lu, 1988). Such scalar field models have an appealing simplicity, since the isoelevations can be interpreted as isopotentials of interpolating functions which are solutions to Laplace's equation, allowing them to be easily computed by relaxation on a digital computer or by analog resistive nets (Koch *et al.*, 1986). Surfaces obtained in this fashion are guaranteed to be of minimal area, although not necessarily unique (Hildebrandt and Tromba, 1984). Once this is done, the neighbors of points and their coordinates can be found either by travelling along the isoelevations or in the

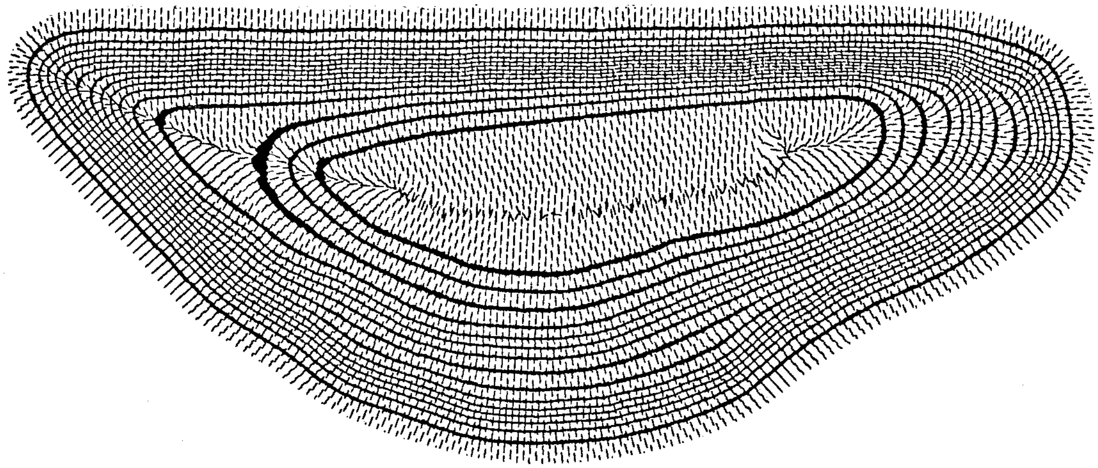


Figure 14. Determination of topology from the gradient of a potential representing the elevation of a surface. Contours show the lines of constant elevation of a surface, while needles show the direction of the local gradient. The isoelevation contours can be regarded as an interpolation of the four contours along which the surface was originally sampled.

orthogonal direction along the gradients (Figure 14). However, the properites of such minimal surfaces may not be well-suited to the reconstruction and interpolation of arbitrary surfaces. For instance, surfaces formed by folding of a planar sheet have zero Gaussian curvature everywhere, so that local regions can be either planar or parabolic, while the reconstructed minimal surfaces must have zero mean curvature, so that local regions are either planar or hyperbolic. More generally, such scalar functions of two-dimensional fields are profoundly insufficient as a general representation for surfaces of objects. The most obvious difficulty arises with the assumption that the surfaces to be represented are single-valued functions of the field. While the single surface of an opaque object can be represented as a scalar function of the image plane, the "near" and "far" surfaces of a transparent object cannot. This representation is also insufficient for representation of an integrated view of two or more images of various surfaces in the vicinity of an occlusion, such as might be obtained during stereoscopic or kineopic viewing.

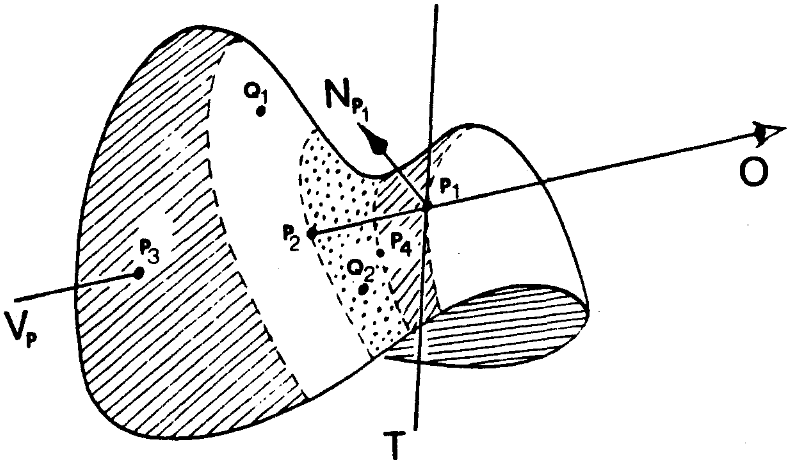


Figure 15. The projection of multiple and topologically discontinuous points P_1, P_2, and P_3 of a continuous surface along a single ray and onto a single point O of the retinal image (Koenderink, 1987).

These troublesome examples, which occur about as often as they are ignored, share the property that multiple and topologically discontinuous regions of a continuous surface may project to a given point on the image (Figure 15). The frequent occurrence of such nonfunctionality in our data precludes the use of scalar functions defined on two–dimensional fields in the case of surface reconstruction from planar contours.

One means of overcoming this difficulty also involves representing the surface as a scalar function, but makes use of a three–dimensional field. When this representation is used to represent objects, the scalar field is often referred to as a *scan* comprised of *voxels*, with the value of the scalar giving the (binary) existence, or (continuous) probability or density of the object at a given voxel.

Although this representation is usually used with solid objects, their surfaces alone can be represented, in which case the scalar gives some measure of the existence of the surface at a given voxel. With this representation, the reconstruction problem can be formally described as finding a function $s = f(x,y,z)$ defined on a three-dimensional field (x,y,z) which interpolates the points sampled by the data contours (x_i, y_i, z_i), $i = 1,\ldots,n$ where we set $s = 1$ on the surface and $s = 0$ otherwise. Such a representation can in principle be used to represent transparent or folded surfaces which cannot be represented as scalar functions of two-dimensional fields. Some authors have made use of such representations, in part because biomedical data are often represented in this fashion (Lorensen and Cline 1987; Sander and Zucker, 1988). However, there are few advantages to this approach. In particular, one cannot find the interpolant by solving Laplace's equation on the three-dimensional field, since adjacent or apposing surfaces tend to blend together. One might remedy this by proposing a *surface process* to segment the *volume* analogous to the use of line processes to segment an area (Koch *et al.*, 1986). However, since surfaces typically occupy only a small fraction of the volume in which they are embedded, such a scalar function defined on a three-dimensional field is a rather inefficient representation for the reconstruction of a surface. It may, however, be appropriate in other domains such as the reconstruction of volumes from a sparse series of scans.

A third possibility represents points on a surface as a mapping (discrete) or vector function (continuous) defined on a two-dimensional field or *manifold*. A manifold is a space which locally resembles either a metric space or a domain in such a space, but which is actually neither of these. Rather, it is a collection of local coordinate neighborhoods embedded in the space (Alexandroff, 1961; Flanders, 1989). Thus, the manifold is neither the plane of an image nor the volume of the scans of an object, but can be thought of as *a region topologically equivalent to the surface of the object.* Since topology varies among objects, reconstruction requires the estimation of topology for the construction of such a manifold, in parallel with estimation of the geometry defined upon the manifold which gives the shape of the surface. By parametrizing the surface by a two-dimensional coordinate net (u,v), we obtain a manifold upon which we can define a three-dimensional vector field whose points $P = (x,y,z) = f(u,v)$ represent the observed geometry of the surface. However, almost all previous work on reconstruction of three dimensional surfaces has involved either the estimation of topology, or the estimation of geometry, but not both in parallel, primarily because most procedures assume rather than construct the manifold which provides the domain for the reconstruction. The reason for this is that the parametrization (u,v) is in general difficult to obtain from typical samples (e.g., Sander and Zucker, 1988). Nevertheless, it is the most promising of the representations for reconstruction of topology and geometry of arbitrary surfaces. We will pursue this approach later in this section.

As is apparent from the above survey of surface representations, a general solution to the reconstruction problem will be relevant to problems of visual object representation and recognition. We will develop this parallel here and throughout this section for two reasons. First, the ease with which the human visual system solves this problem causes us to underestimate its difficulty, which becomes apparent as soon as one attempts to develop an algorithm to solve the problem in any degree of generality. Thus, despite over a decade of work by a number of investigators, the problem of surface reconstruction can still be solved in only a few limited cases. Second, the parallel suggests that surface reconstruction is isomorphic with certain other visual processes, which allows us to apply our knowledge of vision to obtain a solution to the reconstruction problem. Furthermore, by virtue of this isomorphism, this solution in turn provides us with a better understanding of these visual processes, and how they might be accomplished given the known structure of the visual pathway. We will elaborate on this point later in this section and in the **Discussion** at the end of this work.

What evidence is there that the visual system reconstructs both the topology and geometry of objects in the manner suggested above? When we are first presented with an unfamiliar object in the visual environment, we will not know anything about how the parts of the object are connected to one another (topology) or the shape they describe (geometry). This situation is exemplified by the random dot cinematograms and stereograms of Julesz (1971), which can present any of a number of different kinds of surfaces to the visual system through temporal or spatial correlations between otherwise random images. In such stimuli, the individual dots in each image constitute a sampling of the surfaces of an object which are presented in such a way as to remove explicit cues to the topology and geometry of the objects. The visual system can assemble or group these dots into sets corresponding to the *figure* and *ground*, demonstrating that the visual system is capable of discovering and representing the different parts of an object and their interconnections, which is tantamount to a specification of topology. Furthermore, objects having the same topology can be further distinguished on the basis of their shape. The ability of the visual system to determine and represent such shapes demonstrates that it is capable of determining the geometry of objects as well. Since neither topology or geometry are available directly from such random dot images, the visual system must determine these from correlations on the random dot image pairs. This requires solution of the well-known *correspondence problem* (Julesz, 1971; Marr, 1982) in order to obtain the mapping of one image onto the other of the pair which identifies the corresponding points. The correspondence problem can also be regarded as being solved by the reconstruction of the structure responsible for the correlations between the two images (Carman 1987). We shall pursue this point later as well.

The visual system can also distinguish differences in intrinsic and observed geometry. For example, objects having the same topology and intrinsic geometry, but differing in the way in which the objects are embedded in space by a bending of one surface into another, are seen as different conformations of the same object. Conversely, surfaces having the same topology, but differing in their intrinsic geometry by a deformation of one surface into another, are seen as having shapes which, if sufficiently different, will be regarded as different objects. Taken together, the discriminability of the aforementioned objects is evidence that the visual system is capable of distinguishing both the topology and geometry of objects embedded in space, and therefore of solving the reconstruction problem as we have defined it. Since objects differing by only a change in location or conformation should be recognized as the same object, the *reconstruction* of objects by the visual system should preceed the *recognition* of objects per se. Presumably the perceptual constancy of objects differing in this way arises from those aspects of topology and geometry which remain invariant under such transformations. Yet most proposals regarding object recognition fail to appreciate that both topological and geometric ambiguities can and do arise due to the effective undersampling of objects arising from the limited availability of cues (DeYoe and Van Essen, 1988).

Having described the parallel between some aspects of visual perception and the general problem of surface reconstruction, we return to the particular problem at hand. The reconstruction of the cortex for the purpose of mapping requires the determination of the intrinsic topology and geometry of a surface which is embedded within the familiar three-dimensional Euclidean space, based on a sample consisting of the contours of section. Since this sample is continuous within contours, there is no difficulty determining the neighbors of points, and therefore the distances between points, along the same contour. However, since the sample is discontinuous between contours, the neighbors of points on one contour which are located on other contours will in general be ambiguous. This topological ambiguity prevents the accurate determination of the curves of minimal length on the surface of the object, which are known as the *geodesics* of the surface (Hilbert and Cohn-Vossen, 1952; Hildebrandt and Tromba, 1984). Since the intrinsic geometry of a surface is completely determined by these geodesics, the topological ambiguity also produces an ambiguity of the intrinsic geometry of the surface. This is because the geodesics will in general change direction in regions where the surface changes shape, so that we cannot determine their direction on a sparse, discontinuous sample of a variable surface. Since the ambiguities in reconstruction of intrinsic properties of the surface arise from the presence of topological discontinuities in the sample of the surface, we refer to this as the the *topological problem* (Figure 16).

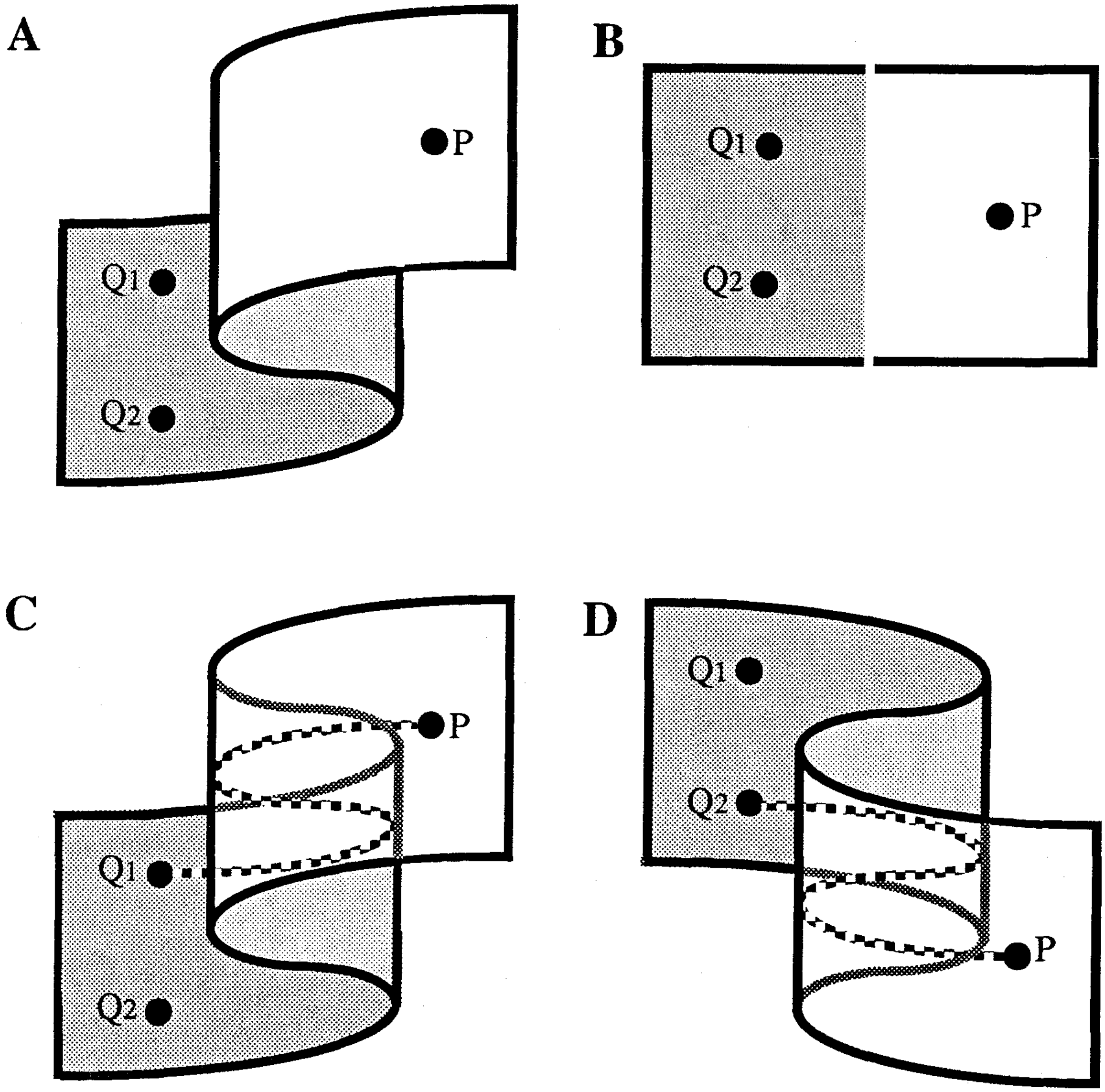


Figure 16. The topological problem. **A**: Perspective view of a folded surface. Surfaces are shaded in proportion to their distance, with the nearest surface being white and the farthest dark gray. **B**: Projection of surface shown in (**A**). Discontinuity in sampling (gap) arising from the fold produces an ambiguity as to whether and how point P is topologically connected to points **Q**1 or **Q**2. **C** and **D**: Two possible interpretations of the surface sampled in (B), showing the geodesics (checked lines). The topological ambiguity has metric consequences.

Although no notion of geometric measure is needed to define a continuous or topological space, geometry provides a convenient means of determining topology. Through the use of a suitable *metric*, which measures the distance *d* between any two points of a space, the topological notion of neighborhood can be expressed in terms of the geometric notion of distance. Thus, the neighborhood of a point **P** can be described as the set of all points **N** within a distance *d* of **P**. Futhermore, we can use this geometric description of neighborhoods to reconstruct unknown topology given the geometry of a sample, which in turn may allow us to determine the intrinsic geometry of the surface. Provided that the sample is sufficiently dense, points which are neighbors in the space of the sample will also be neighbors on the surface being reconstructed. If this is the case, then the ordering of distances from one point to its neighbors on a surface (intrinsic geometry) will be preserved in the ordering of distances among the corresponding points in the sample (observed geometry). Such a nonparametric ordering of points by distance requires only that relative distances along the surface be preserved in the sample space (*cf.* Todd and Reichel, 1989).

In the case of the reconstruction of the cortical surface for mapping, the contours taken from separated planes of section represent a sample of the surface which is continuous within, but not between, the planes of section. Thus, while neighboring points can easily be found for a given point within the same contour, the identity of neighboring points on adjacent contours can be ambiguous. If a surface is sectioned so that the contours are sufficiently close, the distances along any curve on the surface will be well–approximated by distances between samples along that curve, so that the topology recovered in this way will reflect the true topology of the surface. On the other hand, if the sections are sufficiently far apart, the relative distances between points may differ significantly from those on the surface, so that we are not able to unambiguously determine the topology of the sampled surface. Indeed, if the contours are sampled so sparsely that the surface is underdetermined, it may be impossible to reconstruct a unique surface from the contours alone (Gabor, 1946; Shannon, 1948, 1949; Landau, 1967; Logan, 1977; Marr *et al.*, 1979; Requicha, 1980; Curtis and Oppenheim, 1987; Marvasti, 1989).

There are, however, are a variety of definitions of neighborhoods. Many of these share the criterion of minimal distance in some space (Ahuja, 1982; Grunbaum and Shephard, 1989). For example, the neighbors of a set S of points on the plane or in a volume can be found by constructing the *Voronoi tesselation*, which subdivides the plane or volume into polygons or polyhedra each of which contains all points closer to a given member of S than to any other member (Voronoi, 1908; Rogers, 1964; Watson, 1981; Riedinger *et al.*, 1988; Grunbaum and Shephard, 1989). Alternatively, if we are only interested in the neighbors within S , we may construct the *Delaunay triangulation*, which connects members of S that are closer to each other than to any other member (Delaunay, 1906; Rogers, 1964; Lee and Schachter, 1980; Watson, 1981; Riedinger *et al.*, 1988; Grunbaum and Shephard, 1989). The Delaunay triangulation can thus be regarded as the dual[3] of the Voronoi tesselation (Figure 17). While the Delaunay triangulation can be constructed from straight line segments on the plane or in a volume, these should be generalized to curves when constructing the tesselation on arbitrary surfaces. More typically, however, piecewise–linear approximations to these curves are employed.

---

3. Two figures or axioms which have the formal property of remaining unchanged by exchanging "straight lines" and "points" are referred to as *duals*. It is a remarkable fact that all the axioms of plane projective geometry have duals (e.g. "two points determine a line" and "two lines detetmine a point" are duals). The principle of duality is discussed at length by Hilbert and Cohn–Vossen (1952).

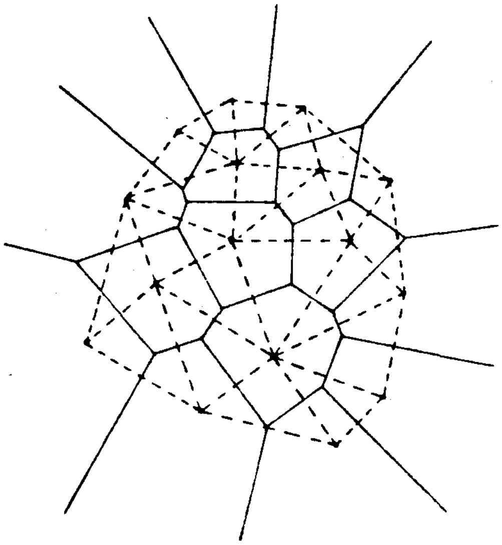


Figure 17. An example of a Delaunay triangulation (dashed lines) of a set of points (filled circles) and its dual, the Voronoi tesselation (solid lines) (Lee and Schachter, 1980).

In some cases, the topological problem is "solved" by specifying neighbors via interactive grouping of points on the surface so as to define and assemble contours to approximate the surface of interest. Exploiting human vision to assist in this process, Levinthal and Ware (1972) presented pairs of contours stroboscopically so as to produce apparent motion between neighboring points on adjacent contours, while Marko *et al.* (1988) and Leith *et al.* (1988) presented stereoscopic views of sections so as to produce binocular fusion of neighboring points. In other cases, the topology was explicitly sought by algorithms which grouped points on the surface into discrete polygonal tiles, and in turn assembled the tiles to approximate the surface of interest. Such a representation is referred to as a tiling or *tesselation* of the surface. Many early developers of such computer imaging or object modelling algorithms made use of heuristics to select the grouping of points defining the tiles. Keppel (1975) proposed an algorthm which sought to subdivide the surface into nested convex hulls, and introduced the use of planar *graphs* to represent neighborhood relationships between points (Harary, 1969; Berge, 1985). Alternative heuristics based on minimal area, equalaterality, similarity in orientation, and other measures of triangular tiles were also proposed (Fuchs *et al.*, 1977; Christiansen and Sederberg, 1978; O'Rourke, 1981; Batnitzky *et al.*, 1981; for early reviews and more heuristics, see: Cook, 1981; Ganapathy and Dennehy, 1982; Funnell, 1984).

Many of these early heuristic algorithms did not obtain appropriate or unique solutions, since objects are not generally shaped according to such heuristics. They also typically could not reconstruct branched objects for which a single cross section splits into two. A recent proposed extension to one of these early algorithms offers no demonstration that the extension is sufficient (Livadas, 1989). In an attempt to obtain a uniquely defined tesselation, several investigators defined tesselations according to the Delaunay triangulation of a sparse sample projected into two dimensions (Lee and Schachter, 1980; Watson, 1981; Ahuja, 1982; Sloan and Houlsby, 1984; Watson and Philip, 1984). Unfortunately, the Delaunay triangulation does not always produce triangles whose edges approximate the contours of section, and produces triangles on the interior and between exterior surfaces, requiring extensive editting to obtain just those triangles on the surface. Recently, Boissonnat (1988) proposed an extension designed to overcome these difficulties, but even this extended algorithm does not properly reconstruct sparse samples. This is a consequence of grouping on the basis of straight−line distances between points in the process of constructing the Delaunay triangulation. Due to the topological problem, such distances do not follow the geodesics in general, and thus do not reflect the true minimal distances between points. Thus points which are actually quite distant on the surface can be grouped together to form tiles which are not part of any real surface. However, the construction of a Delaunay triangulation would provide a unique solution to the problem of surface reconstruction if measurements of geodesic distances between points were available, or if the sample is dense enough so that the difference between geodesic and sampled distances were negligible.

The discrete, polygonal tesselation can be extended to include blending deformations of the surface tiles, allowing us to obtain a piecewise continuous covering of the surface using *patches* or two-dimensional *splines* or, in our terminology, *psxels*. Here a manifold is specified by an orthogonal parametrization (u,v) of the surface, and the spline as a vector field $(f(u), g(v), h(u,v))$ defined over local regions of the manifold. Such a parametrization can always be obtained for a continuously sampled surface, for example, by using length along an orthogonal net of curves on the surface to give $f(u)$ and $g(v)$. However, obtaining this parametrization for a sparse, discontinuous sample of a surface requires solving the topological problem. In most cases in the literature on psxels, this problem is not even addressed, since they have been used primarily for modelling and imaging known surfaces rather than reconstructing unknown ones. Thus, we will not review this literature at length here (for reviews, see: Foley and Van Dam, 1982; Barsky, 1983; Lancaster and Salkauskas, 1986). Some exceptions worthy of note can be found. Dierckx *et al.* (1988) describe an algorithm for generating a tensor product psxellation of surfaces starting from contours of section. They require that the parametrization either be given *a priori*, or be obtainable by a trivial subdivision of the contours of section into an equal number of segments. Neither of these conditions is acceptable in our case. A geometrically sophisticated approach was taken by Sander and Zucker (1988), who used a series of biomedical scans represented on voxels to obtain a local parametric surface patch covering, or psxellation, of relatively simple surfaces contained therein. These authors recognized the topological problem and made use of several results from differential geometry of surfaces in order to reconstruct the toplogy. They were able to accomplish this using measures of surface geometry, namely the principle curvatures and the surface normal at each point. These were locally available in the scans used by Sander and Zucker, but are not in our contour data, so that their methods cannot be applied directly.

The local surface patch or spline representation can be further developed into *models* representing an entire simple object or part of a more complicated one. Here the manifold and model are defined as for the psxels above, except that a single global model has replaced the collection of locally defined psxels. The various models include *generalized cones* or *cylinders* which either represent the surfaces themselves, or are used as a manifold upon which a scalar function can be defined to represent the surface (Binford, 1971; Marr and Nishihara, 1978; Marr, 1982). Both cases were proposed as a means to obtain an *object–centered representation* for visual images and memory. Since the shape of an object is not known *a priori*, both the topology of the model and the mapping of visual images onto the model must be determined. Usually, the class of model to be found in an image is specified *a priori*, which essentially ignores the topological problem. The mapping of visual information is also simplified through restricting the problem to objects and models which have a one-dimensional axis of symmetry or "skeleton" which is found using symmetry-seeking processes operating on the image. As a result, the surfaces which can be represented are thus restricted to be functions of either spherical or cylindrical coordinates (van Oosterom, 1977; Brevdo *et al.*, 1987; Terzopoulos, 1987; Terzopoulos *et al.*, 1987; Dierckx *et al.*, 1988; Horaud and Brady, 1988; Terzopoulos *et al.*, 1988; Bresler *et al.*, 1989). The more general superquadric surfaces were described by Barr (1981, 1984) in the context of solid modelling and deformation, and Solina and Bajcsy (1990) have described an algorithm to fit them to rangefinder data. Unfortunately, all these models are insufficient for representing the various surfaces of the brain to be reconstructed. Neither the surface nor the cross sections of cortex posess the assumed symmetries and therefore cannot be directly represented by one-parameter models. While this might be remedied by developing the capability to construct shapes on more complicated "skeletons" of two or more parameters, and perhaps a hierarchy of shape, such a generalization does not yet exist. Furthermore, the folds of the cortical sulci and gyri cannot be represented by radial scalar functions defined on the models or generated by radial deformation of the models such as have been demonstrated to date.

A more general model-based approach to surface reconstruction can be based on the notion of shape inference and reconstruction. One possible approach relies on the simple partitioning of smooth surfaces into regions of positive or negative Gaussian curvature by the parabolic lines, for which the Gaussian curvature is zero (Hilbert and Cohn-Vossen, 1952). Thus the zeros of curvature on the occluding contour of an object can be used to infer the possible segmentations of the object into parts which are locally convex, concave, parabolic, or hyperbolic (Beusmans *et al.*, 1987). The generic pairings of such zeros of curvature are limited to those for which the parabolic lines do not intersect. We applied this approach to the reconstruction problem in vision, and found that the same inferences could be made for the contours of section as well (Carman, 1987). In this case, the parabolic line segments reconstructed by pairing the zeros of curvature on different contours subdivide the surface spanned by the contours into local regions having different shape. Interestingly, generic pairings of the zeros of curvature also obey the restriction that the resulting parabolic line segments do not cross. Milios (1989) proposed segmenting contours into convex and concave primitives and then finding a minimum cost graph which matches the primitives on neighboring sections. A similar approach was taken by Kehtarnavaz *et al.* (1988), with the exception that their primitives were not based on curvature but rather a semantic description of local contour shape subject to syntactic comparison, very similar in spirit to the codon primitives proposed by Hoffman and Richards (1984). Kehtarnavaz and de Figueiredo (1988) also proposed using the zeros of the total curvature along three−dimensional contours as a means of segmenting and matching contours. This total curvature is the length of the Darboux vector, which includes the normal curvature on a surface plus the *torsion* about a contour. In the case of contours contained within the planes of section, the torsion is zero, and hence the segmentation would be the same as proposed above. Thus, the main contribution of this proposal is the obvious generalization to three−dimensional contours which may be needed in other reconstruction problems.

A completely different approach to shape matching was proposed by Witkin *et al.*

(1987), who described an elegant algorithm for tracking the structure of one-dimensional signals across scales to obtain a deformation which maps such signals onto each other. Their algorithm incorporates many useful components found in earlier approaches, making this an important work for anyone interested in this general class of problem. Although they do not consider the question of reconstruction *per se*, their approach can be readily adapted to the matching of pairs of contours of section for which local orientation has been computed as a function of length along the contour. We implemented a version of this procedure based on an earlier report by Witkin (1983), and found it to work satisfactorily, except that ambiguities sometimes arose which could not be resolved by the algorithm. An example of this approach has recently been published (Algazi *et al.*, 1989). These approaches based on shape or signal matching are all limited in that they are intended for intrinsically one-dimensional data, and no true generalization to two dimensions has been proposed. This becomes a problem for structures that split or merge. In the case of a surface sampled by contours of section, this results in the splitting of a single contour into two or more contours, or the merging of two contours into a single contour. Such splitting and merging of surfaces occurs rather frequently within the cortex, in the form of surface folds which develop to form the sulci and gyri, so that some means of accomodating them must be provided. While it is possible to handle these cases using some form of interactive proceedure, the most appropriate solution would be to find a matching or correspondence proceedure which is genuinely two-dimensional. A final proposal by Lin *et al.* (1989) attempts to reconstruct surfaces from contours of section by a combination of interpolation within contours, interpolation between contours, and spline fitting. Although most of their reconstructions are rather simple, they provide one solution to the problem of branched structures by solving the interpolation and reconstruction separately for all possible combinations of pairs of contours from two adjacent sections, and then taking the solid union of those reconstructions as the desired surface. While this approach will succeed for visualization, it will not yield a topological surface useful as a reference for subsequent mapping without additional processing. Nevertheless, the notion of superposition of surface structure may be useful for other possible approaches to splitting and merging of contours in the course of solving the reconstruction problem.

None of the techniques which we have reviewed above are capable of solving the problem of reconstruction of surfaces from sparse, discontinuous samples. Each lacks some crucial feature, such as a sufficiently general representation of surfaces, or the use of sound geometric principles to obtain the reconstruction. Nor does any proposal appear to proceed in a natural, biological style of computation that can be understood within the context of the structure of the visual system, which we know must also solve this problem. There are several key features of the organization of the visual system which suggest how such computations might be accomplished. First, all stages of the visual pathway have a laminar organization, suggesting that the processing of visual information is intrinsically two-dimensional. This restriction to two-dimensional, laminar organization as opposed to three-dimensional, nuclear organization permits a marked reduction of complexity of the neural circuitry, but also restricts us to computations which can be implemented in such architectures. Second, most of these stages have both feedforward and feedback connections, suggesting that processing at each stage is influenced by that at the next. Thus we contrast the conventional view of a primarily feedforward flow of visual information, which projects from one stage to the next, with an alternative view in which information is iteratively transformed via the feedforward and feedback projections. Third, the topographic transformations from one stage to the next can be substantial, and presumably reflect the processing of visual information at each. Even small differences in topography and magnification can profoundly affect the flow of information between one stage and the next, since the transformations are cumulative under such iteration. Fourth, the known connectivities and selectivities of neurons within the visual pathway suggest that the representation of information at each stage is not static but rather dynamic and dissipative. By this, we mean that the iterative mapping of information under these topographic transformations will either contribute to a stable flow or dissipate otherwise. We have developed a solution to the reconstruction problem which is based on computing such flows, which was driven by these fundamental observations of the visual pathway and the fundamental properties of dynamic systems (for reviews, see Abraham and Shaw 1982; Crutchfield *et al.*, 1986; Stewart, 1989; more detailed treatments of relevant issues are given by Crutchfield, 1984; Hogg and Huberman, 1985). We now turn to an overview of this proposal. Although we discuss the proposal in terms of the reconstruction problem, it should be clear that many if not all problems in vision can be solved by computing such flows.

We begin by noting that the problem of reconstructing an entire surface can be subdivided into a reconstruction of the portion of the surface which spans the space between the contours on two neighboring planes of section. Since the sample is sparse, we must solve the topological problem in order to find the

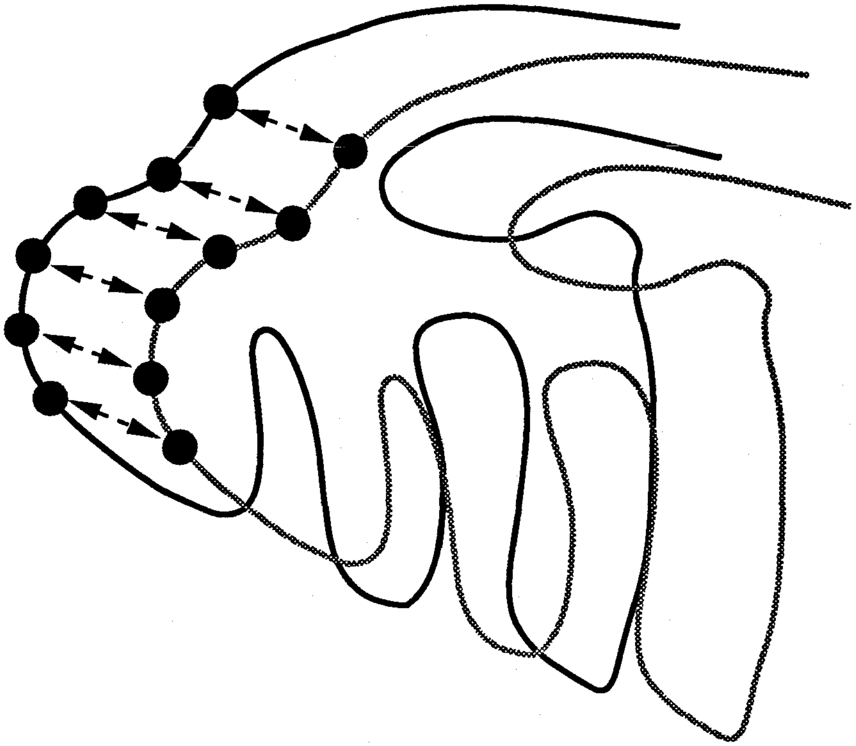


Figure 18. The representation of topology by a discrete graph. Points sampled from the contours of section (solid and shaded curves) are represented by nodes (filled circles). Nodes considered to be neighbors are joined by edges (arrowhead pairs).

points on one contour which are neighbors of points on another contour. One possible representation for this problem would be a graph (Harary, 1969; Berge, 1985) consisting of a set of *nodes* representing the discrete sample along a contour, which are linked by *edges* representing associations or matches between the nodes which are considered to be neighboring (Figure 18). The graph can be considered to be *undirected*, in which case the topology is symmetric, or *directed*, in which case asymmetric neighborhoods can be represented. However, this representation is discrete and has no intrinsic topological structure for incorporation of information about ordering constraints (Yuille and Poggio, 1984) or partial solutions obtained at various scales (Carman, 1987) in any simple manner. An alternative representation

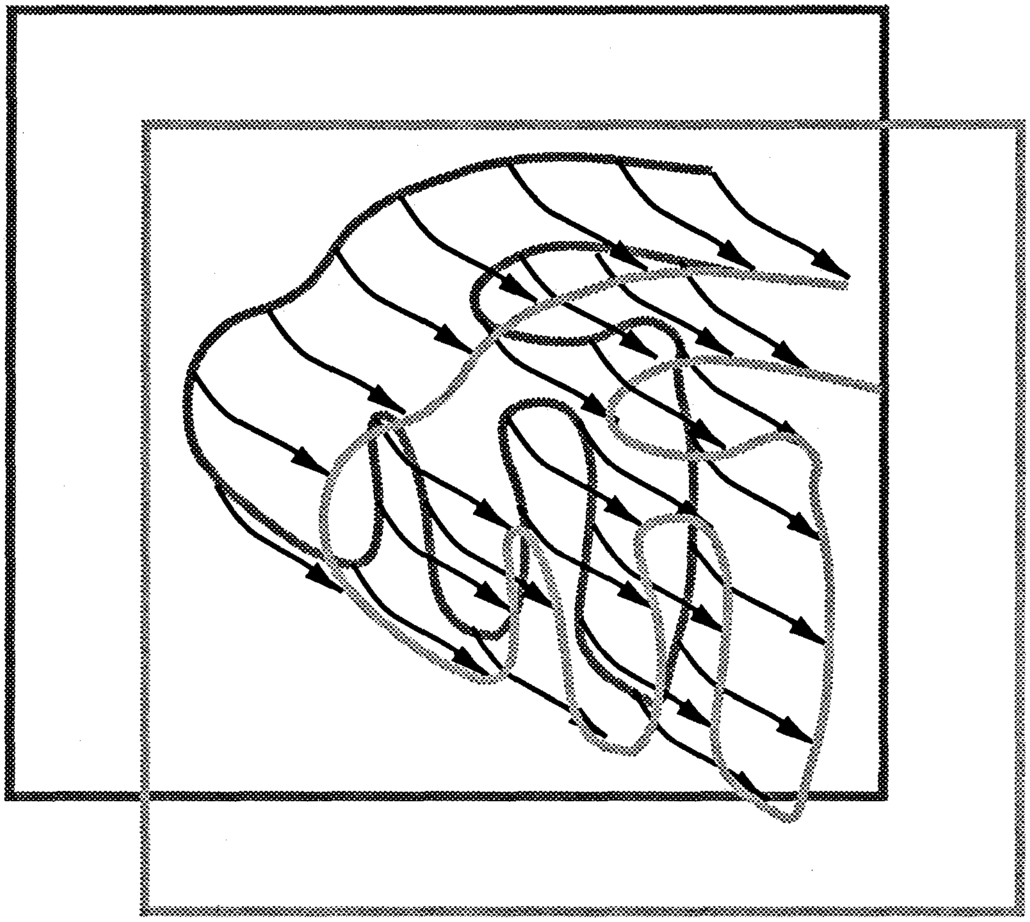


Figure 19. The representation of topology by a continuous flow. All points on the contours of one section are mapped to the contours of the other section. The arrows show the direction, if any, of the flow.

would be a continuous *flow* (Figure 19). This flow can be regarded alternatively as a mapping of the plane onto itself, or a mapping of one plane onto another (Whitney, 1955). On the latter interpretation, the flow can be regarded as the continuous generalization of the discrete graph. Surprisingly, the relationship between the discrete graph and the continuous flow does not appear to have been previously noted. The flow can be regarded as being either undirected or directed, in analogy with the graph.

Unlike the graph, the flow is defined on a continuous two dimensional domain which is intrinsically topological. Thus, flows can be defined which are themselves mappings of the plane which preserve neighborhoods at each point. Such topological flows between one plane and another thus provide a very general representation of topology, correspondences, or associations subject to ordering constraints. Such constraints apply in the topological problem, since valid reconstructions will preserve the order of points between contours. More explicitly, the order of points on one contour must be the same for the order of the neighbors of those points on another contour. Such constraints appear in many of the early visual processes, suggesting a common representation (Carman, 1987). For example, the ordering constraint is applicable in the correspondence problems of stereopsis (Yuille and Poggio, 1984) and kineopsis (Hildreth, 1984), as well as to the problem of determining the shape of an object from the correspondences of the parabolic points on its occluding contour (Beusmans *et al.*, 1987). We note that since the flow is defined on the plane, the ordering of points under the action of the flow will be preserved regardless of what direction is chosen. Thus the common difficulty of generalizing a one–dimensional constraint to two–dimensional domains does not arise.

Topological flows are represented by single–valued, continuous complex functions, which by definition preserve point neighborhoods (Curtiss, 1948; Sokolnikoff and Redheffer, 1958; Fuchs and Shabat, 1964). These functions are defined on the complex plane whose points (x,y) are represented by a complex number z such that

$$z = x + iy$$

where i is the imaginary squareroot of minus one. We consider complex functions f of the form

$$w = f(z) = u(z) + iv(z).$$

where f is single-valued and continuous for some domain D of the plane. We recognize these two condtions as equivalent to the conditions of one-to-one correspondence and preservation of neighborhoods defined for continuous or topological mappings. The derivative of f with respect to z is given by:

$$f'(z) = \frac{\partial u(z)}{\partial x} + i\frac{\partial v(z)}{\partial x}$$

The differentiability of f implies its continuity. For f to be differentiable for each point of D, its partial derivatives must exist on D and satisfy the Cauchy-Reimann equations:

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}$$

Applying these equations to the previous result, we obtain an alternative equation for the derivative of f:

$$
\begin{aligned}
f'(z) &= \frac{\partial u(z)}{\partial x} + i\frac{\partial v(z)}{\partial x} \\
&= \frac{\partial u(z)}{\partial x} - i\frac{\partial u(z)}{\partial y} \\
&= \overline{\mathbf{grad}\,u(z)}
\end{aligned}
$$

where the bar denotes the complex conjugate. Thus the derivative of $f(z)$ is given by the conjugate of the gradient of $u(z)$. Functions f which are single-valued and differentiable are termed *regular* or *analytic*. Such functions are of particular importance because it can be shown that their real and imaginary parts each satisfy Laplace's equation, for which reason they are termed *harmonic* and *conjugate harmonic* functions. A function f which is analytic and whose derivative f' is nonzero for all points of D is said to be a *conformal* mapping of D. We have briefly discussed these before in our survey of the various kinds of mapping.

Such conformal mappings have the property that for some sufficiently small neighborhood, they approximate an affine or similarity transformation composed of a translation from z_0 to $f(z_0)$, a rotation of angle $\arg(f'(z_0))$, and a scaling of magnitude $|f'(z_0)|$ (Fuchs and Shabat, 1964, pp. 59–60). As such, two contours of section which differ locally from each other by similarity tranformations can be mapped onto each other by determining a complex analytic function f which takes the values of translation, rotation, and scaling observed in the neighborhood of each sample. These are related to the mapping f by:

Translation: $f(z_0) = u(z_0) + i\,v(z_0)$

Rotation: $$arg\,(f\,'(z_0)) = tan^{-1}\left(\frac{-\,\partial u\,(z_0)}{\partial y} \Big/ \frac{\partial u\,(z_0)}{\partial x}\right)$$

Scaling: $$|f\,'(z_0)| = \sqrt{\left(\frac{\partial u\,(z_0)}{\partial x}\right)^2 + \left(\frac{\partial u\,(z_0)}{\partial y}\right)^2}$$

These same properties of the conformal mapping also permit the use of local linear interpolation on sparse samples:

$$w = az + b,$$

where a and b are complex numbers given by:

$$a = f'(z_0)$$
$$b = f(z_0) - f'(z_0)\,z_0$$

Such interpolation will be valid over a sufficiently small neighborhood in the vicinity of a sample. Thus, if we make local estimates of $f(z_0)$ and $f'(z_0)$, they can be used to approximate the mapping f in the neighborhood of z_0. A collection of such estimates of f can be considered a distributed representation of the reconstruction. We emphasize that the values of translation, rotation, and scaling will depend upon the actual sampled data, and so will differ in general from point to point in the planes of section.

We refer to surfaces generated by conformal mappings from one contour of the surface to the next as the *conformal surfaces*. Such surfaces are quite general, as they include all the surfaces which can be generated by analytic functions.

Furthermore, conformal mappings can develop multiple contours into multiple surfaces above a point, so that surfaces which are folded or laminated so as to have multiple elevations above a given point of the plane can be easily represented. We therefore overcome the restrictions of earlier methods, such as those based on the minimal surfaces or thin plate splines. While here we consider the reconstruction of a single fixed "slice" of the surface spanning two sets of contours of section, the representation is readily extended to include extended, deforming surfaces, which can be represented by variation of the conformal mapping from "slice to slice" or over time. Such mappings can be expressed as functions of the elevation e and the time t:

$$f(z,e,t) = u(z,e,t) + iv(z,e,t).$$

Alternatively, iteration of a single mapping over time will suffice to represent those surfaces which have the required self-similarity, so that the mapping becomes a function of an iteration parameter n:

$$f(z,n+1) = u(f(z,n)) + iv(f(z,n)).$$

Such surfaces include the well-known fractals (Mandelbrot, 1982) which have recently been applied to image compression of biological structure (Barnsley and Elton, 1988; Barnsley and Hardin, 1989; Barnsley *et al.*, 1989). Finally, we note that although we have considered the case of a single conformal mapping of an entire domain D, our approach can be generalized to include a mapping of D which is piecewise continuous, so as to accomodate surfaces of different objects as well.

The reconstruction of such surfaces from their contours of section can be accomplished by determining the local values of $f(z_j)$ and $f'(z_j)$ for a series of samples of the contours. Since surfaces in general exhibit structure across a continuum of scales (Koenderink, 1984), we must also make these estimates of f and f' across scales. In other contexts, discrete pyramids have been proposed as one means of obtaining such a sample (Adelson *et al*, 1984). However, this approach suffers from the difficulty that pyramids represent structure at only a few discrete scales (e.g. every octave), rather than as a continuum. Thus, structure at one scale is ambiguously connected to form at the next scale, producing topological ambiguities within the scale space itself. We propose an alternative means of reconstruction which involves estimating f and f' across such a continuum of scales, thereby obtaining a single conformal mapping which represents the superposition of the structure found at each scale. The method is based largely on the complex variable theory and the known macroscopic structure of the first stages of the visual pathway. In the following, we provide an overview of this method, followed by consideration of its parts in some detail.

We begin by taking our samples across a continuum of scales σ which vary from the periphery (coarse) to the center (fine), as seen in the primate retinal sampling (Figure 20) (Koenderink, 1986). These samples are resolution-limited by convolution with a Gaussian of width determined by the scale of the sample. Thus each isoeccentric annulus of samples represents structure at a given scale and resolution. For each sample, the local component of $f(z_j)$ is obtained by computing a flow which finds the local translations on the samples. This flow is given by the conjugate of the gradient of a potential $\varphi 1$, producing a local radially symmetric flow which displaces points of the two samples in opposite directions so as to remove translational differences between them. Given the

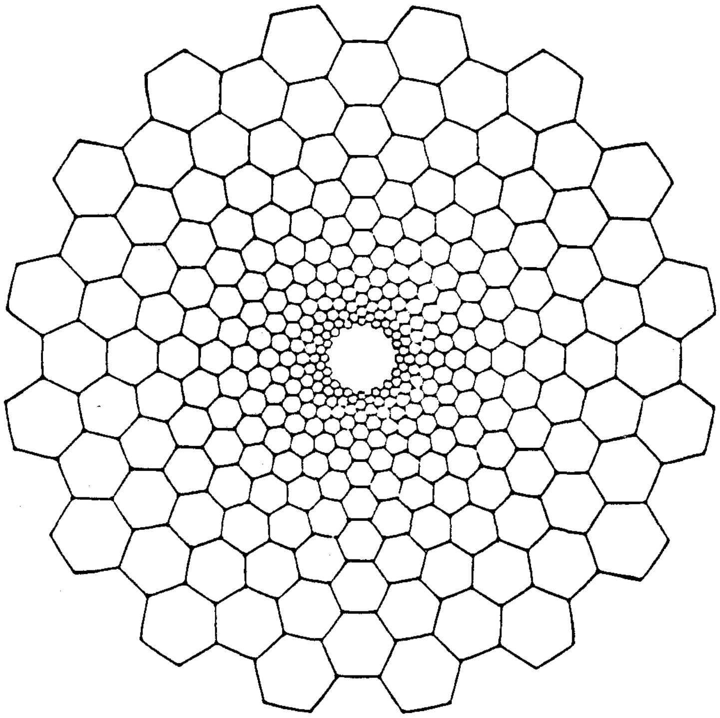


Figure 20. Continuous scale sampling by the primate retina. The size of each hexagonal tile is proportional to the scale of the corresponding sample (Koenderink, 1986).

values of this potential at samples on an annulus, we can determine the values of the potential on the interior of this annulus by solving the *Dirichlet problem* (Sokolnikoff and Redheffer, 1958). This is done for all annuli, and the potentials determined on the interior of each superposed across scales, so that at the center of the sampling we obtain a potential which incorporates information from all scales to determine the translational component $f(z)$ of the mapping. In the course of solving the Dirichlet problem, we will generate a log transformation of the original samples. Since rotations and scalings on the input correspond to translations on this log representation, we can determine the local component of $f'(z_j)$ by repeating the above process. Thus we obtain a second potential $\varphi 2$ whose conjugate gradient gives the local component of the rotation and scaling $f'(z)$ of the mapping. These two potentials $\varphi 1$ and $\varphi 2$ will thus provide our distributed representation of the reconstructed surface. We note that these potentials do not represent the surface geometry directly, but rather describe the values of $f(z)$ and $f'(z)$ of a conformal mapping f of one set of contours of section onto the other. We now turn to a more detailed exposition of the calculation of the local flows used to compute translations on the sample or its log transform.

For each local sample of the two planes of section, we find the local components of $f(z_j)$ or $f'(z_j)$ by constructing a potential φ whose conjugate gradient gives a flow field which will advect or transport the samples onto each other, so as to remove any translational difference between the two images. This unconventional choice of relation between potential and flow yields a potential which is a harmonic function, and yet which also produces a flow field which is radially symmetric and proportional to the distance between corresponding points in the two samples. An example of such a potential evaluated at z_0 due to a point in the sample at z is given by the real part φ of the complex function F:

$$\begin{aligned} F(z,z_0) &= \varphi(z,z_0) + i\psi(z,z_0). \\ &= q(z - z_0)^2 \\ &= [q(x - x_0)^2 - (y - y_0)^2] + i[2q(x - x_0)(y - y_0)] \end{aligned}$$

where q is plus or minus one according to whether the point at z is in the first or second of the two planes of section. That φ is harmonic can be easily verified.

The gradient of the real part φ is given by:

$$\mathbf{grad}\varphi = 2q(x - x_0) - i\,2q(y - y_0)$$

while the conjugate gradient is given by:

$$\overline{\mathbf{grad}\varphi} = 2q(x - x_0) + i\,2q(y - y_0).$$

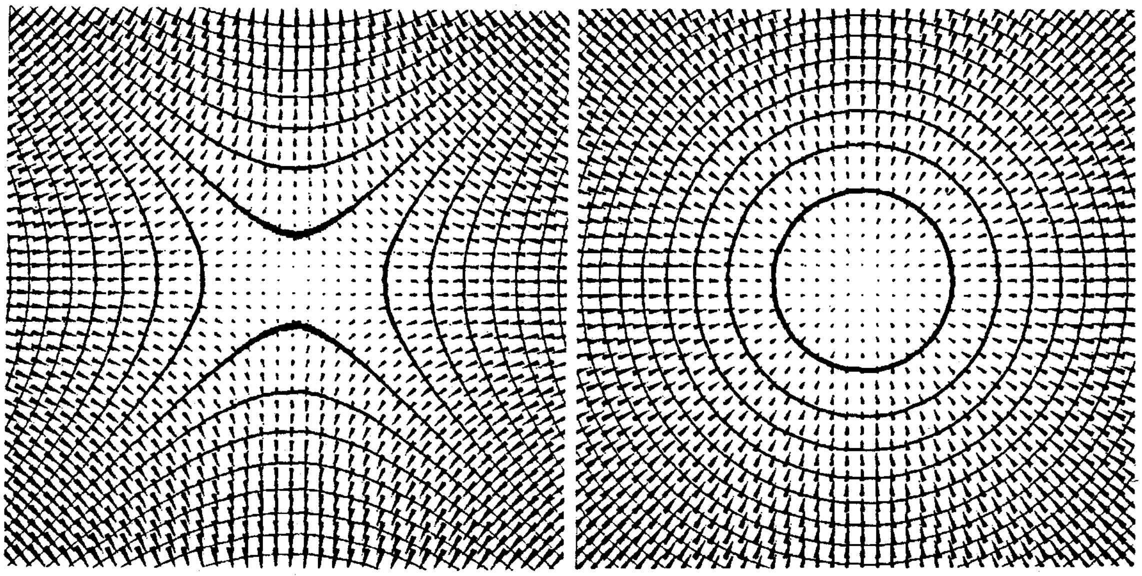


Figure 21. **A**: The isopotentials of φ (lines) and its gradient (arrows) in the vicinity of a point z_0. **B**: The isomagnitudes (lines) an directions (arrows) of the conjugate gradient for the same region.

The general form of such a potential and the conjugate gradient flow field is illustrated in Figure 21.

To obtain the potential or flow over an entire sample due to a number of points in both planes of section, we simply take the sum of the potentials or flow components over all points:

$$\varphi_{sum} = \sum_j (q_j (x_j - x_0)^2 - (y_j - y_0)^2)$$

$$\overline{\mathbf{grad}\varphi}_{sum} = \sum_j 2q_j (x_j - x_0) + i\, 2q_j (y_j - y_0).$$

This flow field for the entire sample is recognized as the first moment over the area of the sample. It will therefore be zero when two samples which differ by a translation are superimposed or fused. To achieve this, each point j is subjected to a flow $\mathbf{v}_j$ according to its 'sign' q_j :

$$\mathbf{v}_j = -q_j \overline{\mathbf{grad}\varphi}_{sum} .$$

Convergence is guaranteed for two local samples which differ by a translation. Furthermore, differences in rotation and scaling do not affect the final result due to the symmetry of the flow field. A simulation of the computation of translation by conjugate gradient flow is shown in Figure 22. Note that here we do not include the contribution from regions outside the local sample to either the potential φ or the flow $\mathbf{v}$. In practice, global components of the potential would be summed with the local potential before the computation of the conjugate gradient and the determination of the net flow. While this is a rather elaborate scheme for the computation of local moments, it has the advantage of allowing precomputation of the spatial component of the potential by mapping given by $F(z, z_0)$, so that linear sums over such a map could compute the potential φ and determine the flow $\mathbf{v}$ appropriate for a given potential and the value of q for each point. We note that the potentials thereby obtained will be spatially oriented with a period of π, but that this orientation will vanish at the center of the sample z_0. These and other aspects of the spatial organization of

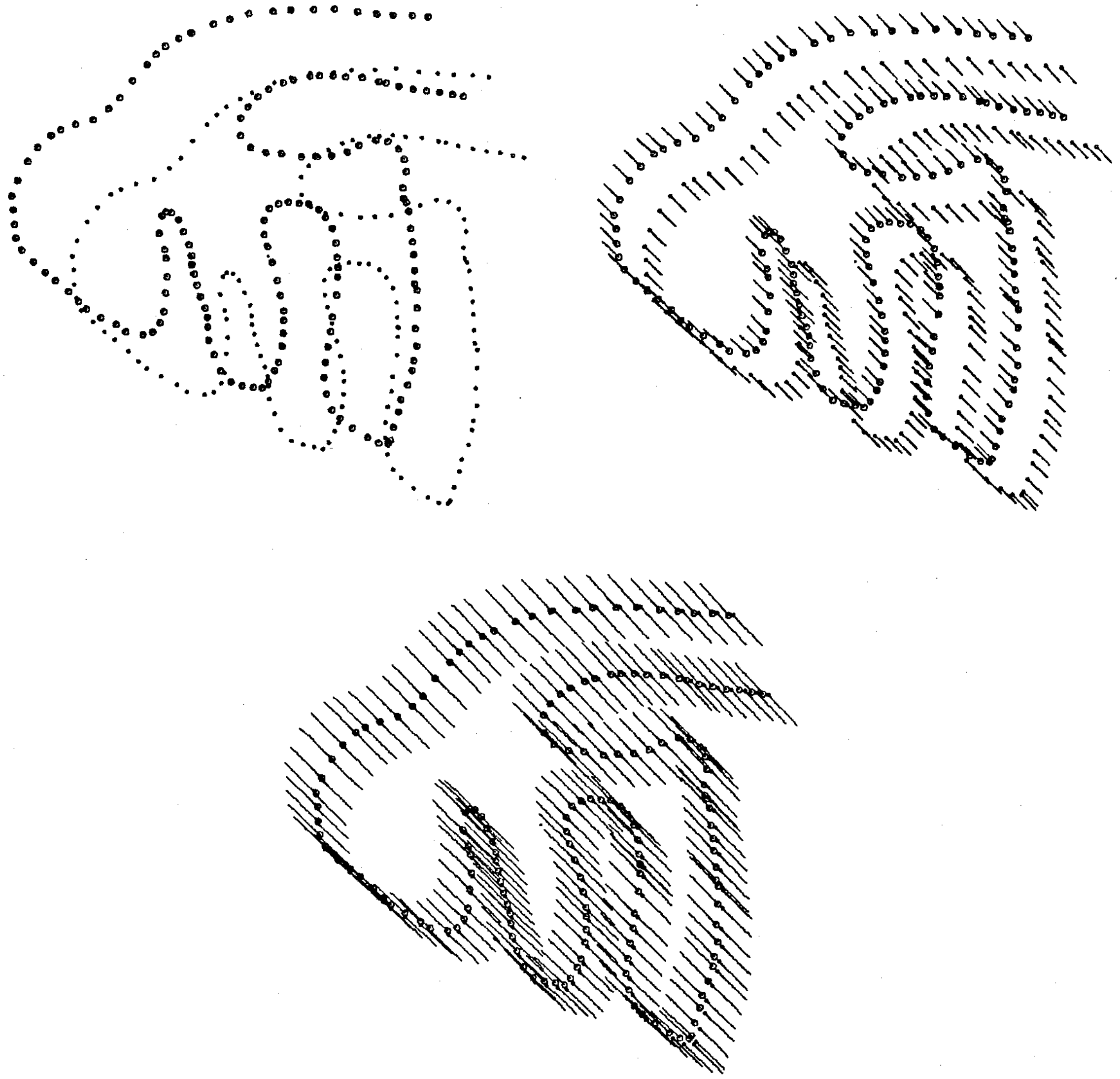


Figure 22. Computation of translation by conjugate gradient flow. Two local samples differing by a translation are shown superimposed, with the large open circles and the small filled circles representing points of each of the two samples. The fine dots show the path taken by points under the action of the flow. The three frames show an initial, intermediate, and final positions of the points during the simulation of the flow. Two samples differing by the same translation will experience the same flow, regardless of the distribution of points.

such a local computational map are similar to those observed in the vicinity of each cytochrome oxidase 'blob' of primate visual cortex.

Given a series of such local estimates $\varphi(\zeta)$ of a harmonic function on an annulus parametrized by ζ, we can find the potential on the interior of the annulus by solving the Dirichlet problem through use of a Green's function (Sokolnikoff and Redheffer, 1958; Fuchs and Shabat, 1964). The Green's function $g(z,z_0)$ is given by:

$$g(z,z_0) = -\ln |f(z,z_0)|$$

where $f(z,z0)$ is a complex analytic function having a sole zero at z_0 and chosen so as to produce a conformal mapping of the domain D of interest onto the unit circle centered on the origin. If our samples lie within a disc of radius R centered on z_0, then we obtain the desired mapping using the bilinear transformation:

$$f(z,z_0) = R\,(z - z_0) \,/\, (R^2 - z^*\overline{z_0}).$$

Substitution of this into the previous formula gives the Green's function:

$$g(z,z_0) = -\ln |R\,(z - z_0) \,/\, (R^2 - z^*\overline{z_0})|\,.$$

With this Green's function we proceed to solve the Dirichlet problem through use of Green's formula:

$$\varphi(z) = \frac{1}{2\pi} \int_C \varphi(\zeta) \frac{\partial g(\zeta, z_0)}{\partial n}\, ds$$

where ζ parametrizes the boundary of the unit disc onto which the samples have been mapped, and n is the inward normal on that boundary. Alternatively, since we are seeking the values of a harmonic function, we can make use of a resistive net to find the solution (Koch *et al.*, 1986).

This can be done by mapping the values of the potential $\varphi(\zeta)$ onto a resistive disc using a conformal mapping whose real part gives the above Green's function:

$$w(z,z_0) = -\,\mathrm{Ln}\,[R\,(z - z_0)\,/\,(R^2 - z^*\overline{z_0})]\,.$$

The values taken by the nodes of such a resistive disc will be the desired solution. The flow field at each point can be computed by finite difference approximation of the conjugate of the gradient for each sample node of the grid. We note that the solution obtained will be valid even for $\varphi(\zeta)$ which has a finite number of discontinuities, provided it is piecewise–continuous. Thus multiple structures imaged in the same region can be reconstructed with the aid of line processes (Koch *et al.*, 1986).

This complex logarithmic mapping of the input domain provides a representation in which log length and angle of input vectors are explicitly represented. If we repeat the identical computation taking w as our domain to be sampled, we will obtain a flow which determines translations along u and v, which correspond to scalings and rotations on z. Thus, the local similarity $f(z)$ and $f'(z)$ at each point z can be determined, and a composition of flows which map one image conformally onto another can be determined (Figure 23). Although we have described this for the case of images consisting of two contours of section, it should be clear that the method can be applied to problems in early vision in general, and can be considered a model for the first stages of the visual pathway.

Finally, we note that computational mappings such as described above can be applied to the more general problem of the reconstruction of entire surfaces of objects from generic views of their surfaces. We note that any smooth surface

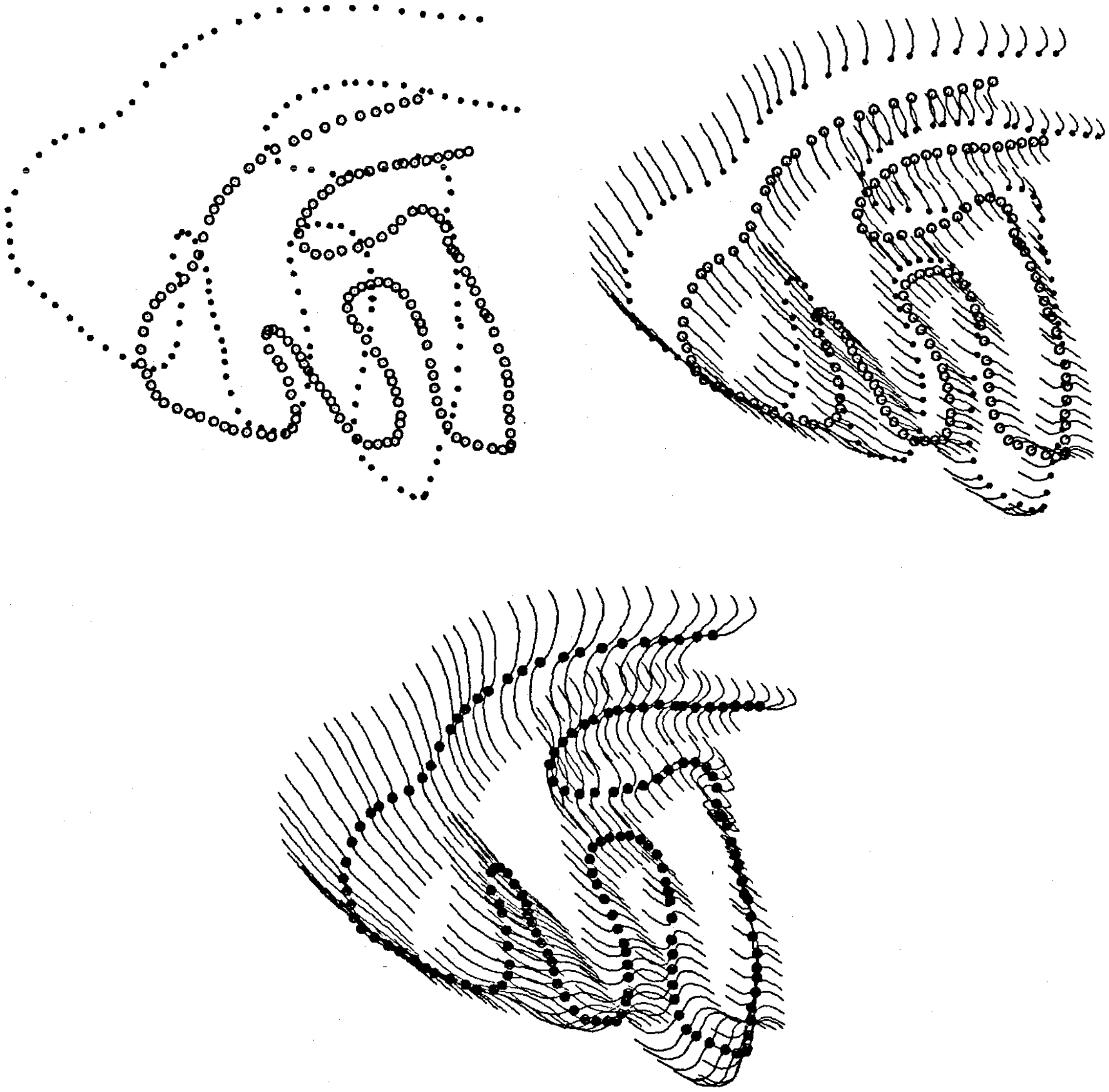


Figure 23. Computation of general similarity by conjugate gradient flow. Two local samples differing by a translation, a rotation, and a scaling are shown superimposed, with the large open circles and the small filled circles representing points of each of the two samples. The fine dots show the path taken by points under the action of the flow. The three frames show an initial, intermediate, and final positions of the points during the simulation of the flow. Two samples differing by a translation, rotation, and scaling experience a flow having a spatial dependence appropriate to produce synchronous fusion of all points at the same time.

can be parametrized along its *lines of curvature*, which are developed by following the directions of the principle curvatures[4] from one point to the next

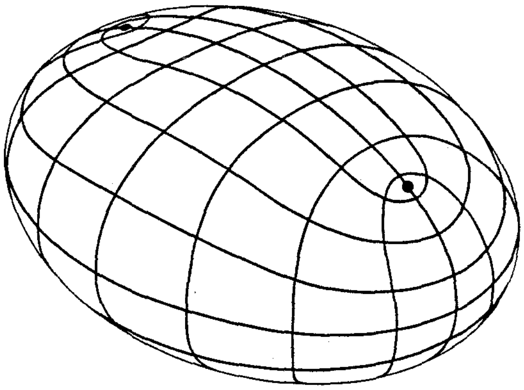


Figure 24. The lines of curvature on a surface form an orthogonal net which parametrizes the surface (Hilbert and Cohn–Vossen, 1952).

on the surface (Hilbert and Cohn–Vossen, 1952). The lines of curvature can be regarded as being generated by pairs of geodesics anchored at foci on the surface; these foci are the *umbilic points* for which the curvature is the same in all directions on the surface. For example, each line of curvature on the ellipsoid is generated by geodesics of constant total length (Figure 24). Since the principle curvatures are orthogonal everywhere they are defined, the lines of curvature are guaranteed to be locally orthogonal at their intersections. Due to their orthogonality, the lines of curvature can be regarded as the *lines of level* and *lines of flow* of a complex harmonic function defined on a topologically equivalent manifold (Curtiss, 1948). On this interpretation, the umbilic points are the *sources* and *sinks* of a geodesic *flow* which is described by the lines of curvature of a surface. In the case of the ellipsoid shown above, this

---

4. Recall that the priciple curvatures are the maximum and minimum normal curvatures defined at each point of a surface.

flow could have sources at the two foci shown atop the object, and sinks at the two foci on the bottom of the object. Although we picture the ellipsoid as it might appear in visual projection, a more 'cortical' representation could be obtained by a conformal mapping which preserves the orthogonality of the lines of curvature. Given such a representation, various surfaces could be generated by placing the appropriate sources or sinks of geodesic flow upon it. The theory we have described above gives a general technique for obtaining a distributed representation of any conformal mapping, and as such provides a foundation for the reduction of visual processes to a common algorithm.

1.6 Optimal Mapping by Elastic Deformation

As described above, the intrinsic curvature of the surface of cortex permits only a subset of its topology and geometry to be preserved when mapped onto the plane. Since the intrinsic curvature is not constant, but varies with location on the cortex, no single class of mapping based on the preservation of a single geometric measure will produce a minimally distorted map across all regions. However, we can attempt to preserve several such measures simultaneously, so that as much of the surface topology and geometry will be preserved as is possible within each mapped region. This can be achieved by quantifying the distortion associated with each measure, and reconfiguring the map so as to minimize a numerically weighted sum of these component distortions across all the mapped regions. The ability to combine several possible mappings by selecting such a weighting of distortions offers a significant advantage over the *ad hoc* choice of any one of them, and is arguably the best that can be done under the circumstances. For example, the cortical cartographer may choose to allow anisotropic shearing or bending distortions, in those areas where it is not possible to preserve all aspects of geometry in the course of mapping. This can be done by selecting a weighting which places a greater emphasis on preservation of linear rather than angular measures, with the result that the measured distances will be preferentially preserved as compared to angles or areas. Once such a weighting has been selected, the mapping is obtained in effect by deforming the surface of cortex so as to minimize the weighted sum of distortion, thereby producing a map which preserves as much of the geometry as possible for a chosen weighting. We refer to such a mapping as *optimal* in the sense that no other map can be obtained which has a smaller weighted sum of distortion for the chosen weighting. We also refer to the mapping as *adaptive* in that the same procedure will result in an optimal map regardless of the details, such as the distribution of intrinsic curvature, of the surface being mapped.

We could attempt to preserve each of the geometric measures of topology, length, angle, and area by including measures of each in the optimization, and allowing the extent of preservation of each to be determined during the course of the optimization. However, we will argue that it is preferrable to treat the preservation of topology as a fixed constraint or boundary condition, and to eliminate the explicit preservation of area *per se*. The reasons for these choices are as follows. Although we could introduce discontinuities on the basis of surface geometry by modelling the tearing of a surface subjected to forces exceeding some elastic limit (Terzopoulous and Fleischer, 1988), the resulting "torn" maps will not be very useful if they include discontinuities through regions of interest. Hence, we require that discontinuities be introduced not in the specific locations which if torn would minimize distortion, but in nearby locations which correspond to the boundaries of cortical areas. For example, the boundary of striate cortex is usually circumcised in the course of manual unfolding. We therefore will require that the mapping maintain topology except at specified boundaries of cortical areas. These locations are in general chosen to coincide with those selected by Van Essen and Maunsell (1980). Second, we choose not to preserve the measure of area since this measure is completely determined by the measures of length and angle.

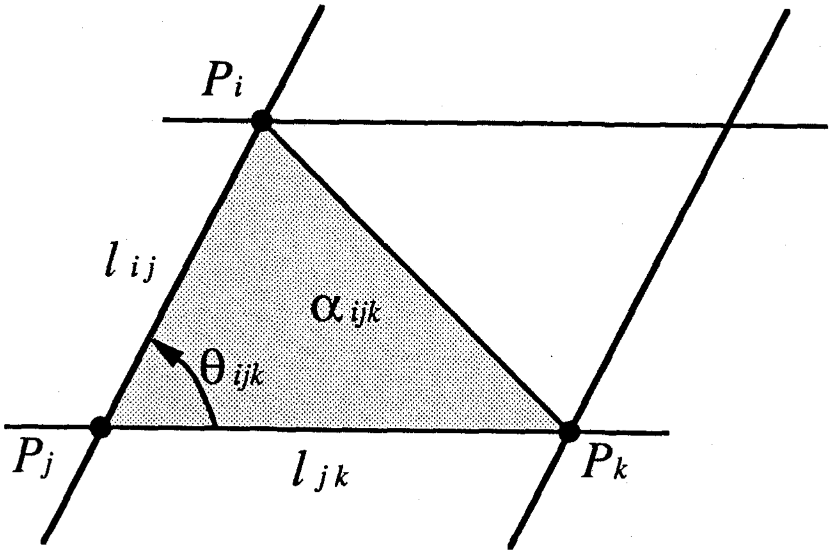


Figure 25. Relationship between length, angle, and area for a triangular surface tile. See text for details.

This can be seen by considering the relationship between area, length, and angle for a given region on the surface of cortex and in the plane of the map (Figure 25). Each triangular tile spanned by three points P_i, P_j, P_k has area

$$\alpha_{ijk} = \frac{1}{2} l_{ij}\, l_{jk} \sin \Theta_{ijk} ,$$

so that these values of length l and angle Θ determine the area of the tile. If these triangular tiles are sufficiently small, the area on the surface spanned by geodesic curves through the three points is approximated by α_{ijk}. Thus, if the lengths and angles for each such tile are preserved, then so will be the area of each tile. Since the total area of either the surface or the map is obtained by summing over all such tiles, the preservation of each tile's area guarantees the preservation of the total area on the surface on the map. However, preservation of each of these lengths and angles may not be possible due to frustration in the mapping, whereas preservation of area is always possible. Should applications arise in which exact preservation of areas are deemed more important than the alternative, we can use the measures of length and angle with the above equation to obtain a measure of area, allowing us to perform an area−preserving mapping instead. In practice this change requires editting only a single line of computer code.

By attempting to preserve a weighted measure of length and angles, we achieve an optimal mapping which will minimize distortions differentially for different regions of a surface. For example, a region with zero intrinsic curvature can be mapped isometrically onto the plane, while a region with constant intrinsic curvature can only be mapped conformally onto the plane. By minimizing both the linear and angular distortion, we obtain an *adaptive* mapping which attempts to preserve as much of the surface geometry as possible. Under such an adaptive mapping, regions will be mapped isometrically where possible (zero intrinsic curvature) by preserving measures of length. Alternatively, regions will be mapped conformally where possible (constant intrinsic curvature) by preserving measures of angle. Provided our surface is reconstructed using sufficiently small surface patches, so that intrinsic curvature is approximately constant over each, one of these two conditions should always hold. The result of the adaptive mapping will thus in general be a discrete approximation to a continuously variable conformal mapping of the surface of cortex onto the plane, but for those particular cases where the cortex has zero intrinsic curvature, the mapping will become isometric. The residual distortions which may be present in any given map may be quantified in terms of the linear and angular components of distortion, so that the deviation of the map from isometricity or conformality can be displayed as a function of location, and taken into account in subsequent analyses based on these maps. This adaptive feature of the mapping will be demonstrated later by application of the same mapping process to a variety of surfaces having different values of intrinsic curvature.

Thus the measures of local connectivity, length and angle are sufficient to characterize the local geometry of the cortical surface for a wide range of possible mappings.

A mapping of the cortical surface which preserves topology, length, and angle can be obtained by deformation of an elastic model (Sokolnikoff, 1946; Terzopoulos *et al.*, 1987). In such models, topology is maintained by coupling a sampling of points or *nodes* by elastic elements which tend to keep the nodes together. In most elastic models, these elements are usually longitudinal springs having a resting length l^o_{ij} corresponding to the length between the nodes i and j they join, as measured on the surface to be mapped. When the actual length l_{ij} of the spring differs from this resting length, the spring provides a restoring force $\mathbf{f}_{ij}$ along its longitudinal axis only, of magnitude:

$$|\mathbf{f}_{ij}| = - \kappa_{ij} (l_{ij} - l^o_{ij}) ,$$

where κ_{ij} is the longitudinal elastic coefficient (units of force per unit elongation), and is given by:

$$\kappa_{ij} = \frac{\kappa^o}{l^o_{ij}}$$

where κ^o is the longitudinal elasticity (units of force per unit elongation per unit resting length) of the surface (Figure 26).

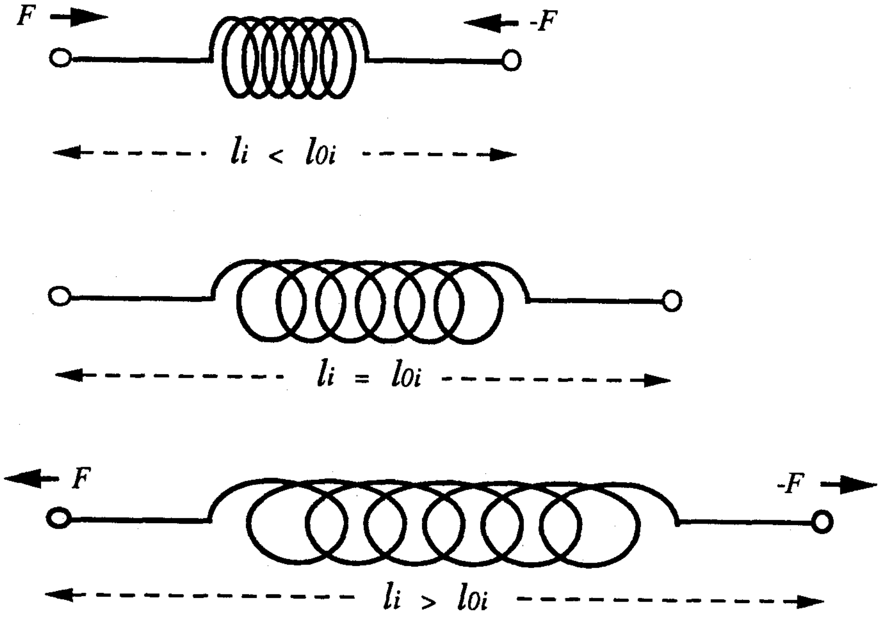


Figure 26. Longitudinal spring under compression, at rest, and elongation. Arrows show the direction of the force $\mathbf{f}_{ij}$ exerted by the spring in each condition.

This relationship permits us to obtain the value of κ_{ij} for a particular spring connecting nodes i,j from the value of the elasticity κ^o, which is a constant for a homogeneous and isotropic medium. This provides a normalization of the restoring force per unit resting length so that the elastic response of the surface to length deformations does not depend on the spacing of neighboring nodes. Two longitudinal springs of different resting length will thus exert the same force in response to proportional deformations, say a doubling of the resting length.

While models constructed from, say, a Cartesian grid of such longitudinal springs respond accurately to homogeneous deformations, such as stretching uniformly in all directions, they fail to respond properly to inhomogeneous deformations, which result in *shearing* and *bending* along the longitudinal axis of the spring. Thus the above model of an elastic surface provides no restoring force for the preservation of angles during deformation, and is insufficient for our purposes. A more realistic model would join neighboring nodes with deformable "elastic beams" offering restoring forces and torques when they are stretched, sheared, bent, or twisted. The response of such beams and surfaces constructed from them to deformation is complex, and the full description requires the introduction of tensor notation and stress–strain relationships (see Sokolnikoff, 1946, Chapter IV). These will not be pursued here, since the construction of an elastic model suitable for mapping by deformation of the cortical surface does not require the inclusion of all such responses to deformation. For example, twisting about the longitudinal axes of the beams should be allowed to occur freely in the course of unfolding the elastic surface, as such twisting does not represent a distortion of geometry on the surface *per se*.[5] Hence we do not include restoring torques against such twisting in our elastic model. In addition, if the beams are sufficiently short, then bending and shearing can be approximated by the rotation of a rigid beam about an elastic base, which offers a restoring torque proportional to the angle of rotation (see Appendix I). Thus, for our present purposes, it will suffice to approximate such elastic beams by a superpostion of springs which provide a restoring force $\mathbf{f}_{ij}$ described above to oppose stretching along their longitudinal axis, and a restoring torque to oppose rotation of this axis about the nodes.[6] Accordingly, we add to the above model a torsional spring producing a restoring torque $\mathbf{t}_{ijk}$ when the angle Θ_{ijk} between neighboring beams ij and ik differs from the corresponding resting angle Θ^o_{ijk} as measured on the surface to be mapped.[7] This restoring torque has magnitude:

$$|\mathbf{t}_{ijk}| = -\lambda_{ijk}(\Theta_{ijk} - \Theta^o_{ijk})$$

where λ_{ijk} is the torsional elastic coefficient (units of torque per unit rotation), and is given by:

$$\lambda_{ijk} = \frac{\lambda^o}{\Theta^o_{ijk}}$$

and λ^o is the torsional elasticity (units of torque per unit rotation per unit resting angle) of the torsional spring (Figure 27). This relationship permits us to

---

5. While twisting deformations will occur when surfaces embedded in three dimensions are deformed to lie in the plane, such deformations cannot be compensated by motions in the plane alone. Thus, including restoring torques against such twisting deformations would offer no improvement in the ability of an elastic model to serve as a substrate for cortical mapping.
6. Note that the addition of a longitudinal spring joining nodes j and k does not achieve the same result. Use of a longitudinal spring in this manner would introduce an additional dependence on the length of the beams ij and ik as well as on the angle between them Θ_{ijk}. In comparison, use of a torsional spring introduces only the dependence on the angle Θ_{ijk}, so that linear and angular distortions will produce separable restoring forces and torques.
7. Whereas the longitudinal restoring force can be defined in terms of the difference in the length of the beams, requiring only two subscripts to specify each longitudinal component, the restoring torque is defined in terms of the difference in the angle between two beams, requiring three subscripts to specify each torsional component.

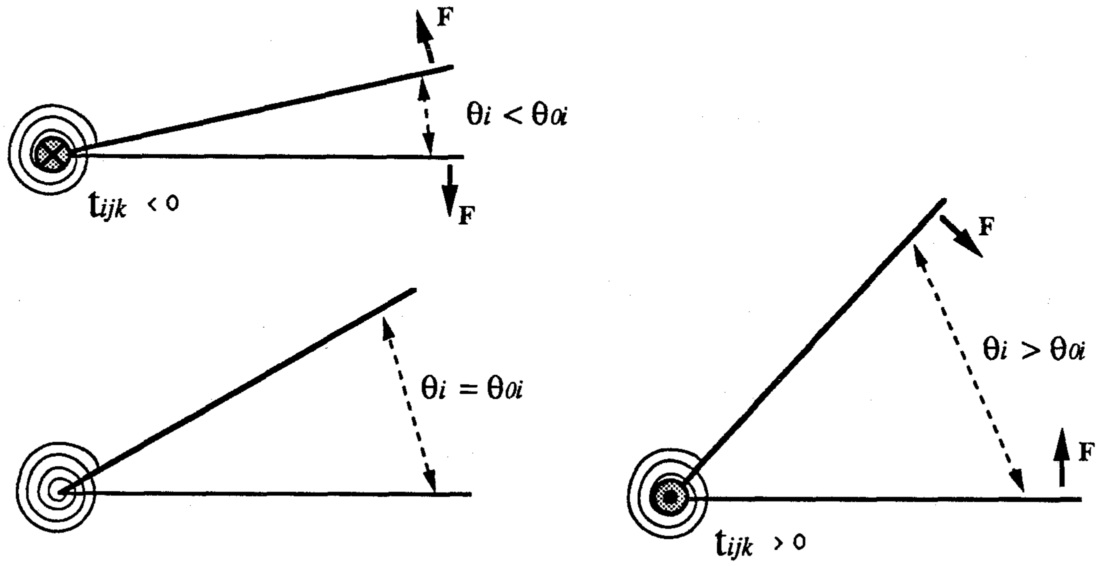


Figure 27. Torsional spring under compression, at rest, and elongation. Arrows show the direction of the torque $\mathbf{t}_{ijk}$ exerted by the spring in each condition, and the direction of the forces **f** which result at the end of the radial beams.

obtain the value of λ_{ijk} for a particular torsional spring acting between beams i,j and i,k from the value of the torsional elasticity λ^o, which in the present model can be shown to be a scaled version of the longitudinal elasticity κ^o (see Appendix I), and therefore also a constant for a homogeneous and isotropic medium. This provides a normalization of the restoring torque per unit resting angle so that the elastic response of the surface to angular deformations does not depend on the angular spacing of neighboring nodes. Two torsional springs of different resting angle will thus exert the same torque in response to proportional deformations, say a doubling of resting angle.

If we wished to simulate the dynamics of such an elastic model, it would be necessary to resolve the restoring forces acting on each of the n nodes in order to obtain a set of n coupled equations of motion, whose solutions could be obtained numerically. However, we are not interested in the actual motion of the nodes of such an elastic model, but rather in the shape or *conformation* of the model which has least distortion. This distortion is measured by the differences in local geometry on the deformed model as compared to those on the resting surface conformation. For our choice of mappings which preserve length and angle, these distortions are proportional to the magnitudes of the forces $\mathbf{f}_{ij}$ and torques $\mathbf{t}_{ijk}$ described above. These forces and torques can be integrated to obtain the associated scalar energies, with the energy due to linear distortion of a single beam being given by:

$$e^{l}_{ij} = \frac{1}{2}\kappa_{ij}\,(l_{ij} - l^{o}_{ij})^2 \,,$$

and the energy due to angular distortion by a pair of beams being given by:

$$e^{\Theta}_{ijk} = \frac{1}{2}\lambda_{ijk}\,(\Theta_{ijk} - \Theta^{o}_{ijk})^2 \,.$$

Thus we see that the beams which connect adjacent nodes on our elastic model will have zero energy when their lengths l_{ij} and angles Θ_{ijk} are at their resting values l^{o}_{ij} and Θ^{o}_{ijk}, and will have positive increasing energy when they are either stretched or compressed by increasing amounts or bent by increasing amounts in either direction. This energy is determinined in the neighborhood of a given node i by summing over the beams i,j connecting that node to its neighbors j, to give the energy for each node due to linear distortion:

$$e^{l}_{i} = \frac{1}{2}\sum_{j}\kappa_{ij}\,(l_{ij} - l^{o}_{ij})^2 \,,$$

and by summing over the angles spanned by adjacent pairs of beams i,j and i,k in cyclic fashion to give the energy for each node due to angular distortion:

$$e_i^{\Theta} = \frac{1}{2} \sum_{j,k} \lambda_{ijk} (\Theta_{ijk} - \Theta_{ijk}^{o})^2 .$$

Since the energy stored in these elastic beams is proportional to the linear and angular distortion, the conformation of least distortion will have minimum total energy, where this minimum will in general be greater than zero. This total energy is given by summing over the linear and angular energies contributed by each node:

$$\begin{aligned} E^{total} &= E^{linear} + E^{angular} \\ &= \frac{1}{2} \sum_{i} \sum_{j} e_{ij}^{l} + \sum_{i} \sum_{j,k} e_{ijk}^{\Theta} \\ &= \frac{1}{2} \sum_{i} \sum_{j} \kappa_{ij} (l_{ij} - l_{ij}^{o})^2 + \sum_{i} \sum_{j} \lambda_{ijk} (\Theta_{ijk} - \Theta_{ijk}^{o})^2 \end{aligned}$$

where i ranges over all nodes so that all lengths and all angles on the surface contribute to this sum. We note that each term E^{linear} and $E^{angular}$ is similar to the root-mean-square error for n measures x, given by:

$$\text{RMS Error} = \left(\frac{1}{n} \sum_{i=1}^{n} (x_i - x_i^{o})^2\right)^{\frac{1}{2}}$$

which is a common measure of distortion of maps (Richardus and Adler, 1972; Gilbert, 1974; Bassett, 1972). In the present case we have normalized not by the number of measurements n, but rather by their resting values, so as to give a root sum of the error or distortion per unit measure. Thus, by altering the conformation of the elastic model of the cortical surface so as to minimize the sum of the linear and angular energies associated with each node across all nodes on the map, we obtain a mapping of that surface which has minimal distortion.

1.7 Optimal Conformations and Global Energy Minima

An elastic surface will spontanously adopt this minimal energy configuration, provided that there exist no dynamic barriers to doing so. Such barriers are well known to occur in a variety of physical and chemical processes, and occur for elastic surfaces when the energy cannot simply be decreased, but must first be increased in order to go from one equilibrium conformation to another of lower energy. The amount of energy increase needed is termed the *transition energy* of the conformational change. Such barriers to conformational change exist for elastic surfaces in three dimensions, as can be seen by considering an elastic hemispherical shell such as half of a hollow rubber ball. If such a hemispherical shell is turned inside–out, it will remain that way even though its energy is higher than that of its inside–in, resting conformation. In order to get the shell to return from the inside–out to the inside–in conformation, one must overcome the transition energy via deformation forces sufficient to allow the shell to pass through an unstable transition conformation of higher energy.

Such barriers also arise when surfaces having a three–dimensional resting conformation are restricted to two dimensions, which reduces the number of degrees of freedom available for conformational changes. For intrinsically curved regions of the surface, this can give rise to *frustration* (Kirkpatrick, 1977), in which no single conformation of the surface minimizes the distortion of all geometric measures. Instead, different subsets of these measures are satisfied by different conformations having local minimum energy, separated by intermediate conformations of higher energy. As a result, the energy landscape upon which this optimization occurs can become pocked with relative energy minima which correspond to conformational *folds* of two kinds (Figure 28). The first kind involves a folding of one continuous contour on the surface about another due to shearing within the plane, and produces a non–local distortion

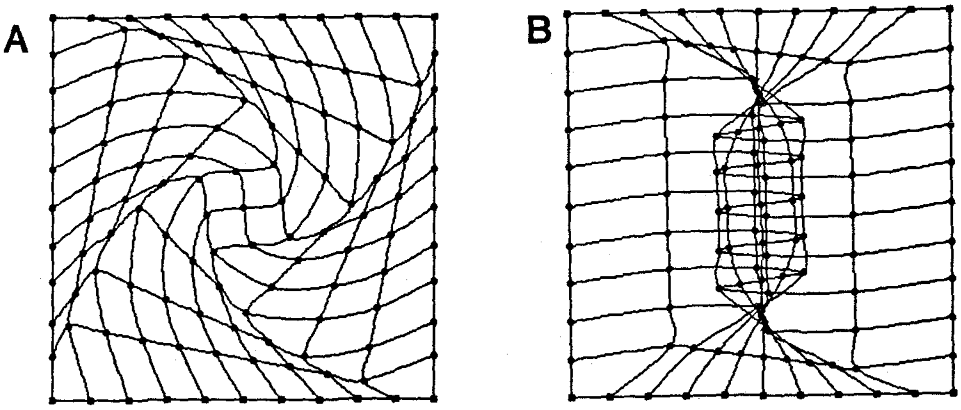


Figure 28. Two kinds of folds which can arise from relative energy minima. A. The first kind results from shearing within the plane. B. The second kind results from shearing of the plane "over itself" so as to produce a degeneracy in the mapping. Some folding of the first kind has been added to permit visualization of the contours in (B).

of shapes of the contours on the elastic map. The second kind involves a folding of one region over an adjacent region due to shearing of the plane "over itself" so as to produce a degeneracy in the mapping, wherein two or more points on the cortical surface project to a single point on the elastic map. Simple optimization algorithms, which attempt to proceed along trajectories of strictly decreasing energy, become trapped in such local energy minima, and will then be unable to find configurations at or near the absolute minimum energy and distortion (Bounds, 1987). In order to find configurations at or near the global energy minimum, we applied the stochastic optimization technique known as *simulated annealing* (Kirkpatrick *et al.*, 1983). In our application of this technique, we assume the map to be in some initial suboptimal configuration, so that the sum of energy E^{total} is greater than its minimum value. Nodes on the map are then visited randomly, and local random displacements $\mathbf{dr}_i = (dx, dy)$ in the plane of the map attempted, so as to simulate the thermal motion of a collection of particles (Metropolis *et al.*, 1953). Such displacements are evaluated to determine if they are to be accepted or rejected by calculating the change in energy that would result:

$$\Delta E\,(\mathbf{r}_i, \mathbf{dr}_i) = E^{total}(\mathbf{r}_i + \mathbf{dr}_i) \;-\; E^{total}(\mathbf{r}_i)\,.$$

We note that only those local beams defining lengths l_{ij} and angles Θ_{ijk} which would be deformed by the displacement $\mathbf{dr}_i$ need be considered in this calculation, since only those will contribute to a change in the energy. A displacment $\mathbf{dr}_i$ resulting in a negative value of $\Delta E\,(\mathbf{r}_i, \mathbf{dr}_i)$, corrersponding to a decrease in local energy and an improvement in local geometry on the map as compared to the cortical surface, is always accepted. However, displacements which result in a positive value of $\Delta E\,(\mathbf{r}_i, \mathbf{dr}_i)$, corresponding to an increase in local energy and a less favorable local geometry, are also accepted with a probability given by the Boltzmann distribution:

$$P(\Delta E\,(\mathbf{r}_i, \mathbf{dr}_i)) = e^{-\Delta E\,(\mathbf{r}_i, \mathbf{dr}_i)\,/\,k_B T(t)}$$

where k_B is Boltzmann's constant and $T(t)$ is a control parameter termed the *temperature* which is a function of time t. The initial value of the temperature $T(0)$ is set high, so that effectively all displacements are accepted, allowing for a randomization of the positions of the nodes from their initial positions. The temperature is then decreased over time, according to:

$$T(t) = T(o)\,\alpha^{t}$$

where α is chosen slightly less than unity to produce a slow exponential decrease in temperature over time, hence the term "annealing." This results in a progressive increase in the probability that a displacement will improve the local geometry over the course of the annealing. Nevertheless, there is always some probability that a displacement which actually increases energy and degrades local geometry will be accepted. This provision makes it possible for the algorithm to escape from local minima so that it can continue on to obtain a map of absolute minimum global energy, provided that the annealing is done slowly enough, so that all the nodes can be considered to be at approximate thermodynamic equilibrium (Geman and Geman, 1984).

Having outlined the theory behind the mapping of the cerebral cortex, we now turn to a description of the practical methods employed in the computation of such maps.

2. Methods

Having developed the theoretical background for the mapping, we now consider the methods employed in practice for mapping the cortex. The process of cortical mapping generally involves four steps:

1. Selection of the lamina to be mapped, and of a subset of points to obtain a *sample* approximating the laminar surface;
2. Reconstruction of the surface to be mapped, so that geometry on this surface can be measured to provide a *reference* for lengths and angles on the map;
3. Creation of the *map* and optimization of its local geometry by comparison to the reconstructed reference ("unfolding"); and
4. Projection of data from the cortical lamina to *color* the surface of the unfolded map.

Each of these steps will be discussed in detail below. All algorithms were implemented in either Fortran or C in a Unix environment. Computers used for this work included a Masscomp 5400 computer with three 68020 CPUs and Aurora graphics displays, a Silicon Graphics Iris 2400 with 68010 CPU, and a Silicon Graphics Personal Iris 4D20G. Most algorithms are portable with the exception of those for the reconstruction of surfaces, which made extensive use of hardware–dependent features of the Silicon Graphics Iris family.

2.1 Sampling of the Surface to be Mapped

In order to sample the surface of interest, the tissue containing the surface is invasively or noninvasively sectioned to generate the contours C defined by the

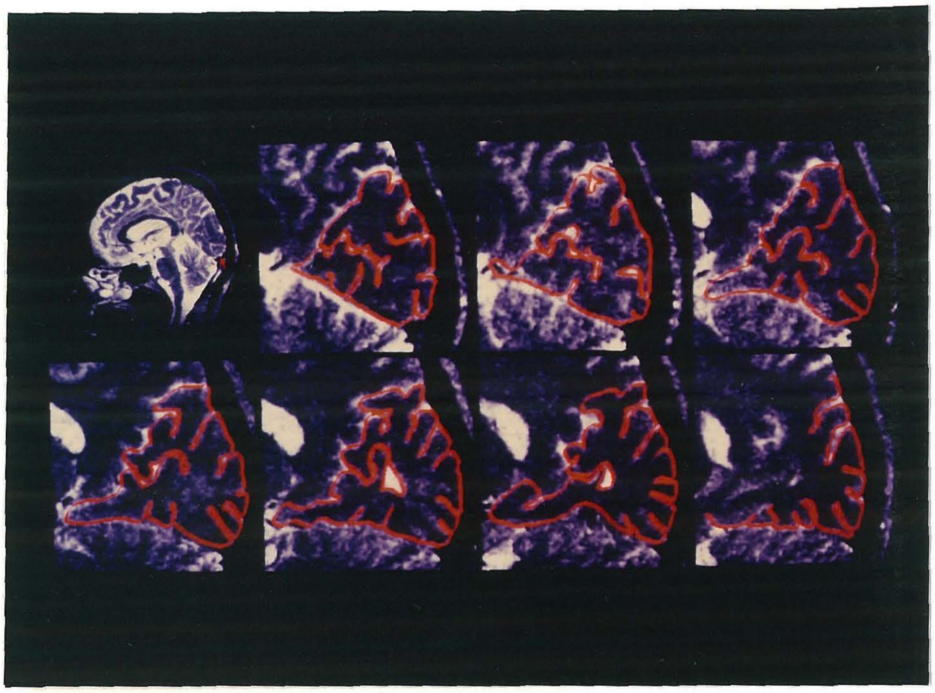


Figure 29. A series of magnetic resonance images of human occipetal cortex and the contours of section obtained from them. Images are in the parasaggital plane, spaced 2.7 mm apart.

intersection of with the parallel planes of section. Examples of such contours can be seen in histological sections obtained by postmortem microtomy of monkey cortex, or a series of magnetic resonance images of human cortex obtained *in vivo* (Figure 29). Such contours can be digitized directly from histological sections by use of a rear-projection digitization tablet, or from images by use of a grapics cursor manipulated via a mouse (Forbes and Petry, 1979). These digitized data consist of an ordered list or series of points, together with any other information which may be associated with the points, such as anatomical or physiological data. The digitized data thus consist of a series of contours $\mathbf{C}_D$, each consisting of a total of N_D data vectors $\mathbf{D}(i)$ indexed by the integer i, where $0 <= i < N_D$. The data vectors thus have the form:

$$\mathbf{D}(i) = [x(i), y(i), z(i), \alpha(i), \beta(i), \ldots],$$

where x, y, z give the three-dimensional coordinates of each point, while α, β, ... code a series of alphanumeric labels specifying what data if any are found at that point. The data α, β, ... might be sampled directly from the images being digitized, sampled from other images which are in register with the image being digitized, entered via a microscope interfaced to a computer, or entered manually via keyboard. The *order* or *sense* in which the data are digitized provides an implicit orientation of the surface useful for subsequent reconstruction and mapping operations, and is therefore preserved. We arbitrarily select a counterclockwise sense for digitization of the contours when viewed from above, *i.e.*, when looking down the z axis. The data are taken at the full resolution of the digitization device so that they may be subsequently smoothed, resampled, and represented either as a series of discrete points taken along the contour, or as a piecewise-continuous spline of the contour. Regardless of which contour representation is used, the points which comprise the discrete sample or which define the splines will be used as the *nodes* which are mapped in the course of computational unfolding.

The discrete representation consists of a series of contours $\mathbf{C}$ which are smoothed and subsampled versions of the digitized contours $\mathbf{C}_D$. Some smoothing is almost always required to remove noise introduced by manual tracing or other digitization. For the discrete representation it is achieved by low-pass filtering the contour so as to remove or reduce these sources of noise. For uniformly digizited contours, we compute the following approximation to a convolution with a Gaussian *along the length of the curve*:

$$\mathbf{P}(i) = \frac{\sum_{j=-s,s} w(j,s)\mathbf{D}(i+j)}{\sum_{j=-s,s} w(j,s)},$$

where the coefficients $w(j,s)$ can be taken from the binomial distribution:

$$w(j,s) = \binom{2s}{s+j} = \frac{(2s)!}{(s+j)!(s-j)!},$$

with $2s$ giving the width of the smoothing kernel. Typical values of s range from 2 to 3 pixels or digitization units. Once smoothed in this fashion, the digitized contours can then be decimated by sampling points every S units. We choose S to be uniform over the length of the contours for reasons which will be explained below. The data components α, β, ... are attributed to the nearest sample **P**. This results in an ordered set or list of n nodes indexed by i, where $0 <= i < N$. Each such node can thus be characterized by position $\mathbf{P}(i)$, tangent $\mathbf{T}(i)$, and curvature $\mathbf{K}(i)$:

$$P(i) = [x(i), y(i), z(i)]$$

$$T(i) = \frac{dP(i)}{ds}$$

$$= \frac{P(i+\Delta i) - P(i)}{|P(i+\Delta i) - P(i)|}$$

$$K(i) = \frac{dT(i)}{ds}$$

$$= \frac{T(i+\Delta i) - T(i)}{|P(i+\Delta i) - P(i)|}$$

where ds is the infintesimal displacement along the contour, here approximated by finite differences computed at the resolution Δi of the digitized data, and $|A|$ denotes the norm or magnitude of the vector **A**. This representation of the contours has the advantage that the geometry of the curves is simply computed and explicity represented. This geometry was used in the course of performing reconstructions of some the sampled surfaces, as will be described later.

Alternatively, the contours **C** can be represented by piecewise–continuous cubic splines specified by position and/or tangent values at nodes (also known as *knots* in the spline literature) along the contours. We select the piecewise cubic spline as it is the lowest order spline for which the measures of position and tangent can be continuous where the endpoints of two splines meet. In addition, once coefficients determining the spline are calculated, the values of the position, tangent, and curvature along the spline can be readily computed. In three dimensions, each piece of a parametric cubic spline is given by P(u), where the parameter u can be considered the normalized length along the spline, so that $0 <= u <= 1$:

$$P(u) = [x(u), y(u), z(u)],$$

where

$$x(u) = a_x u^3 + b_x u^2 + c_x u + d_x,$$
$$y(u) = a_y u^3 + b_y u^2 + c_y u + d_y,$$
$$z(u) = a_z u^3 + b_z u^2 + c_z u + d_z.$$

Differentiation of these parametric equations with respect to the parameter u gives the tangent vector T(u):

$$T(u) = \frac{dP(u)}{ds}$$

$$= \frac{[t_x, t_y, t_z]}{(t_x^2 + t_y^2 + t_z^2)^{1/2}},$$

where

$$t_x = dx(u)/du = 3a_x u^2 + 2b_x u + c_x,$$
$$t_y = dy(u)/du = 3a_y u^2 + 2b_y u + c_y,$$
$$t_z = dz(u)/du = 3a_z u^2 + 2b_z u + c_z.$$

Differentiating again gives the curvature vector **K**(u):

$$
\begin{aligned}
K(u) &= \frac{dT(u)}{ds} \\
&= \frac{[k_x, k_y, k_z]}{(k_x^2 + k_y^2 + k_z^2)^{1/2}},
\end{aligned}
$$

where

$$
\begin{aligned}
k_x &= d^2x(u)/du^2 = 6a_xu + 2b, \\
k_y &= d^2y(u)/du^2 = 6a_yu + 2b, \\
k_z &= d^2z(u)/du^2 = 6a_zu + 2b.
\end{aligned}
$$

Depending upon data available and the desired properties of the interpolating spline, for each coordinate one can obtain a series of four equations in four unknowns which can be solved given four data values at the nodes (Foley and Van Dam, 1982; Lancaster and Salkauskas, 1986). Thus, given two endpoints $P(0)$, $P(1)$ and the tangents at the endpoints $T(0)$, $T(1)$ we can obtain the *Hermite spline* which passes through and takes the specified values of the tangent at those endpoints. Alternatively, given four points and the desire for continuity of position, tangent, and curvature of the endpoints, while sacrificing the requirement that the curve pass exactly through the data points, we can obtain the *B-spline*. The latter has the desirable property that it smooths over the data points and so can be used to effectively filter noise from either manual digitization or digital quantization. The B-spline is computed by matrix multiplication:

$$P(i,u) = \mathbf{U}\ \mathbf{M}\ \mathbf{N}$$

where

$$U = [\,u^3 \quad u^2 \quad u \quad 1\,]$$

$$M = \frac{1}{6}\begin{bmatrix} -1 & 3 & -3 & 1 \\ 3 & -6 & 3 & 0 \\ -3 & 0 & 3 & 0 \\ 1 & 4 & 1 & 0 \end{bmatrix}$$

$$N = \begin{bmatrix} x(i-1) & y(i-1) & z(i-1) \\ x(i\ \ ) & y(i\ \ ) & z(i\ \ ) \\ x(i+1) & y(i+1) & z(i+1) \\ x(i+2) & y(i+2) & z(i+2) \end{bmatrix}$$

The B−spline thus obtained is defined on the interval between node i and $i+1$ as u is varied from 0 to 1. When the end of this interval is reached, the nodes defining N are all incremented by one to obtain the spline on the next interval. Again we select the spacing of the nodes which determine N to be uniform for reasons to be given below. Note that the product **MN** provides us with a matrix of coefficients for the components of $P(i,u)$ which permit the determination of the components of the tangent and curvature as well. Finally, once an unfolded map is obtained, the coordinates of the mapped nodes can be used in N to give images of the splines on the unfolded map. The cubic splines thus provide a more efficient representation of contour data, and are to be preferred over the discrete representation when the number of contours and desired resolution become high.

In order to accurately map the chosen surface of cortex at a given resolution, the nodes which comprise the discrete sample or which are used to define the splines of the contours **C** should satisfy two criteria. First, the sampling should be homogeneous, having equal density in all regions of the surface, so as to provide equal resolution of all parts of the surface during the mapping. Second, the sampling should be isotropic, having a comparable density in all directions along the surface. Optimal sampling schemes which attempt to sparsely sample curves where their shapes change most rapidly do not satisfy these criteria (Pavlidis, 1974; Pavlidis, 1978; Dunham, 1986; Dorst and Smeulders, 1987). Achieving such a homogeneous, isotropic sample is difficult in practice due to the variation in distances between contours generated by sectioning of a nonuniform shape by parallel planes of section. For a surface inclined by an angle ϑ to planes of section Δz apart, the distance d between points on adjacent contours will be:

$$d = \frac{\Delta z}{cos(\vartheta)} \quad .$$

Thus the spacing Δz between planes of section effectively limits resolution of the mapping, since from the above equation d must be at least Δz. Furthermore, since the angle ϑ varies from one region of the surface to another, the value of d will vary along the contour, so that a sample taken along the contours cannot in general be both homogeneous and isotropic.

One solution to this problem would be to provide for an adaptive sampling which would iteratively reconstruct the surface based on the existing sample and then adjust the sampling to obtain the desired properties. Since the final sample would not necessarily lie along the original contours, any data sampled along those contours would require interpolation. Such adaptive sampling and interpolation are beyond the scope of our current methods and have thus far not proven necessary. Instead, we find it sufficient to seek a sample which is homogenous along the contours and isotropic on on average. The latter implies a sampling Δs along contours which is greater than the minimum spacing Δz but less than the maximum spacing which may occur. Empirically, we have found that a spacing $\Delta s = 1.41*\Delta z$ provides an approximately uniform, isotropic sampling of surfaces. One can show that the expectation of d will have approximately this value for surfaces intersecting the planes of section at angles ranging from $-2\pi/5 < \vartheta < 2\pi/5$, with all surface angles equally likely. Once reconstruction has been completed, statistics on the internode distances in various directions on the surface are computed so as to determine the density and isotropy of the sampling of the surface. If these statistics suggest large anistropy, the spacing Δs along the contour can be changed to correct for the average anisotropy. Alternatively, one can obtain additional contours of section either directly from the data set, or by interpolation of existing contours, so as to better sample the surface. This interactive proceedure is often necessary where the surface is approximately tangential to the planes of section, as distances between contours can be quite large in such areas.

2.2 Reconstruction of the Sampled Surface

As described in Section 1.5 above, the reconstruction of surfaces from sparse samples is a challenging problem. Accordingly, we initially performed the reconstruction interactively. This was accomplished through creation of a graphic interface between the user and the data comprising the sample of the surface. This interface, implemented on Silicon Graphics Iris computers, provided a dynamic perspective projection of the three-dimensional contours and nodes of the sample, enabling the user to view the data from various directions and to manipulate the data so as to provide kineoptic cues. Both perspective and dynamic cues were thus available to provide information about the shape of the surface being reconstructed. Information about the topology of the surface was stored within the program as a list of nodes $\mathbf{n}_i$, each of which had a list of edges giving the indices of other nodes $\mathbf{n}_j$ which are neighbors of $\mathbf{n}_i$. We refer to this abstract set of nodes and edges as the *neighborhood graph* $\mathbf{G}$.

Figure 30. User interface for graphical viewing and reconstruction of three-dimensional reference surfaces. The menu on the right permits selection of one of several working environments. The menu at the top permits selection of options within a given environment. A cursor (arrow) manipulated via a mouse is used to select one or more nodes, to be joined by an edge as neighbors, here represented by line segments between the nodes.

The display provided the user with a visualization of this neighborhood graph (Figure 30). Nodes were represented by small points of different colors for different planes of section. Selection of menu entries enabled a variety of options for construction of the neighborhood graph **G**. Most of these options involved selecting nodes through use of a cursor tied to a mouse, and creating edges between nodes in a manner which depended upon the menu option selected. The simplest option created a single edge between two nodes. Such edges were displayed as line segments between selected nodes. Other menu entries enabled interactive and automatic interpolation between existing edges, editting to undo errors or remove unwanted edges, options to move the data in three dimensions so as to permit better viewing or selection of nodes, printing of views for documentation, and saving the current neighborhood graph **G** to disk. Neighborhoods of nodes within a given contour were automatically determined by their order along the contour, with sequential nodes taken to be symmetric neighbors.

Once this graphical user interface was developed, we attempted to automate the reconstruction process using a variety of approaches which we developed or which had appeared in the literature reviewed earlier. These efforts included (1) computing Delaunay triangulations of the surface, (2) reconstruction of minimal surfaces defined on the domain of the planes of section and represented by a scalar elevation satisfying Laplace's equation, (3) matching of shapes of contours based on zeros of curvature across a continuum of scales, as well as (4) methods based on the differential geometry of surfaces. When used on sparse samples of a surface, all these methods succumbed to one or both of the following problems: (a) the representation was incapable of representing folds in the surface, or (b) the method was incapable of reconstructing surfaces spanning different numbers of contours on the two sections. Both of these difficulties can be traced to underlying assumptions regarding the local topology of the domain upon which the reconstruction takes place. The first problem (a) arises because topology of the surface is assumed to be identical to that of the projective plane (parallel to the plane of section), which is only true if there are no folds in the surface. The second problem (b) arises because reconstructions between pairs of contours assume that the surface will have the same number of boundary curves in both sections, which will also be true only if there are no folds in the surface. In order to surmount these problems, two approaches have been taken. First, an algorithm was developed which makes use of the differential geometry of surfaces in order to reconstruct the unambiguous subset of neighborhoods, so as to accomplish at least part of the reconstruction automatically. This has become our standard method of reconstruction and will be described below. Second, we have obtained the solution of the general reconstruction problem, as described in Section 1.5 above, which has been partially implemented. We anticipate that this solution will supplant existing methods once implementation is complete.

Certain topologcal and geometric properties of surfaces can be used to determine a valid process for reconstruction of a discrete neighborhood graph **G**. First, we note that points along a contour have a definite order, so that three points **A**, **B**, **C** on one contour which have respective neighbors **A'**, **B'**, **C'** on another contour will have the same order along the two contours. That is, if **B** is a neighbor of **B'**, then **A** cannot be a neighbor of **C'** and **C** cannot be a neighbor of **A'**. In effect, this means that the edges of two neighboring nodes must not cross, although they may converge upon a single node. If nodes along a contour are sequentially indexed by an integer i for one contour and j for the other contour while traversing the contours in the same direction, the index of nodes along one contour will be a monotonic function of the index of nodes along the other contour. This is true even for closed contours, provided we take the periodicity of the indexing into account. That the neighborhood graph must be everywhere monotonic reduces the size of the required representation from N^2 to $N\log_2 N$. Second, if all contours are traversed in the same direction (e.g. counterclockwise when viewed from above), we can compute local approximations to the tangent of the curve at each node. These tangents must be approximately in either the same direction or in opposite directions for neighboring nodes, depending upon whether the nodes are on the same section or on different sections. Their relative direction can be determined by the sign of the dot product of the tangents, $\mathbf{T}_i \cdot \mathbf{T}_j$, which will be positive if the tangents are in the same direction and negative if they are in the opposite direction. Thus, given two nodes $\mathbf{n}_i$ and $\mathbf{n}_j$ on contours C_1 or C_2, the tangents of neighboring nodes must obey the following constraints:

Contour of $\mathbf{n}_i$	Contour of $\mathbf{n}_j$	$\mathbf{T}_i \cdot \mathbf{T}_j$
C_1	C_1	–
C_1	C_2	+
C_2	C_1	+
C_2	C_2	–

Application of this direction constraint rules out numerous incorrect neighborhoods, such as those which might leave the surface and extend through an object. Yet it still permits the determination of correct neighborhoods across local convex or concave portions of the surface, where nodes have their neighbors on opposite sides of the same contour. Third, and finally, we note that two points which are on geodesics are by definition symmetric neighbors separated by a minimal distance. Thus, nodes which are each other's symmetric nearest neighbors are likely to be neighbors on the surface. Such symmetric nearest neighbors can be found by examining the matrix of all internode distances, and selecting those which satisfy the aforementioned constraints. In practice, when the nodes are sparsely sampled from a highly folded surface, it is not always possible to find such symmetric nearest neighbors on the surface. In such cases, either interpolation or interaction must be applied to reconstruct the remaining neighborhoods. Several views of such a reconstruction is shown in Figure 31. Once such a reconstruction of the surface topology is obtained, local measurements of lengths l_{ij} and angle Θ_{ijk} in the vicinity of each node are obtained from simple geometry.

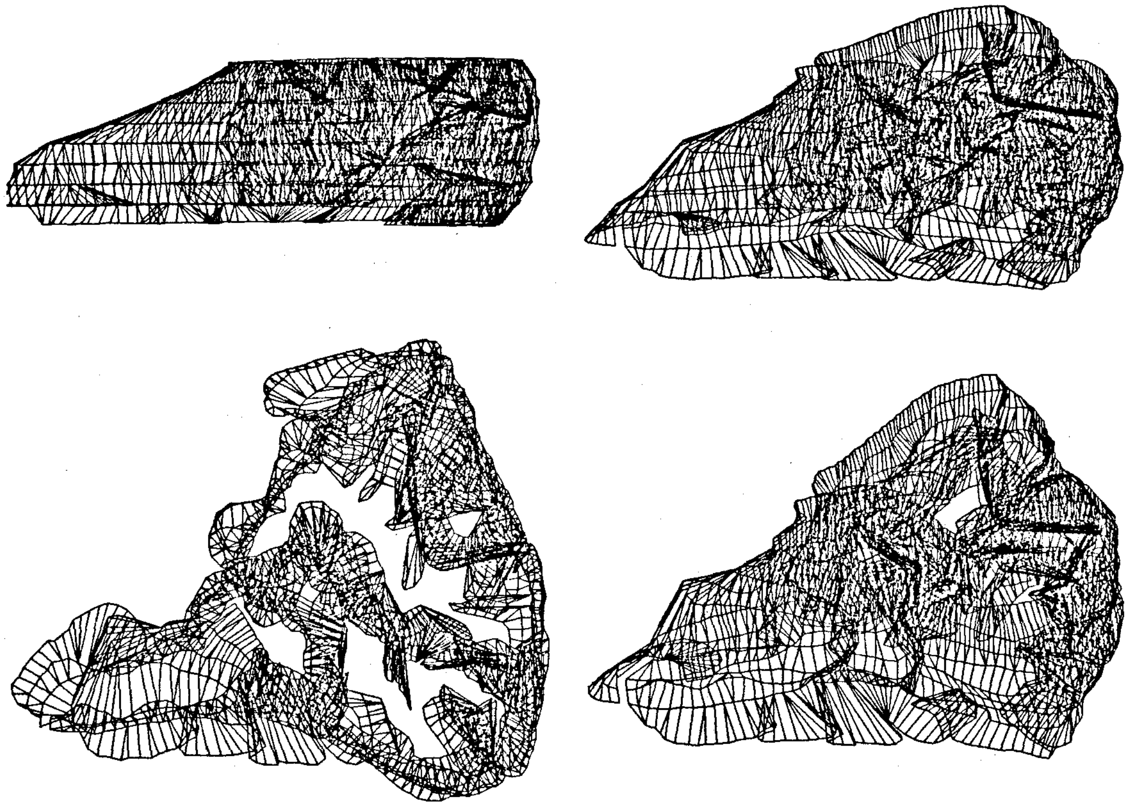


Figure 31. Four views of a surface reconstructed from the contours of section shown in Figure 29 above. The view proceeds from a ventral aspect to a medial aspect going clockwise from the top left.

2.3 Mapping of the Reconstructed Surface

The process of mapping begins with the creation of a node *image* on the map for each node on the reconstructed reference surface. Whereas the positions $\mathbf{r}_i$ of nodes of the reference surface have three coordinates,

$$\mathbf{r}_i^{ref} = (x_i, y_i, z_i),$$

the nodes on the map are given only two coordinates,

$$\mathbf{r}_i^{map} = (x_i, y_i),$$

making the map planar by definition. In the following, we will refer to the nodes on the map simply as nodes for brevity, unless the distinction is necessary.

In some cases these nodes are all initially placed at the origin (0, 0) of the plane, and topography allowed to emerge in the course of mapping. While this provides a powerful demonstration that the algorithm can generate topographic organization even when initially absent, it is not the most efficient means of beginning the mapping process. In order to save the computation required to initially establish topography, we can begin with an initial estimate for the map with correct topography but not necessarily correct geometry. Such a topographic ordering of the nodes can be obtained through use of their neighborhood relationships to sort them into order on the plane. Although there are several approaches to obtain such an ordering (Acton, 1970), we obtain it by converting the previously obtained neighborhood graph **G** into a *regular graph* for which there are an equal number of edges originating from and terminating upon each vertex.[8] Once this is done, this "regularized" graph can be imaged on the plane by identifying each vertex with one of the points of a cartesian grid. For those cases where the neighborhood graph is already regular,

---

[8] Normally the term *regular* signifies that the sum of edges originating and terminating from each vertex are equal for all vertices. Here we relax the definition so as to also apply to subgraphs of a regular graph, whose vertices have equal numbers of originating and terminating edges, the total number of which may differ from one vertex to another. Thus a subgraph **G'** formed by "trimming" vertices and edges originating or terminating from those vertices from a regular graph **G** is also referred to as regular.

the initial estimate for the map can be produced directly from it.

This sorting is accomplished as follows. The number of edges which originate from or terminate upon each vertex $\mathbf{v}_i$ of the neighborhood graph **G** are counted. The former quantity, known as the degree of divergence $d^+(\mathbf{v}_i)$, while the latter is known as the degree of convergence $d^-(\mathbf{v}_i)$ of each vertex. By construction of **G**, the divergence $d^+(\mathbf{v}_i)$ will have values 2 or 3 at boundaries of the graph, and 4 otherwise. However, the convergence $d^-(\mathbf{v}_i)$ will vary between 2 and considerably larger integers. The process of regularizing these graphs is accomplished by reducing the convergence to equal the divergence for each vertex, so that $d^-(\mathbf{v}_i) = d^+(\mathbf{v}_i)$. This can always be done by searching the graph for vertices where this does not hold, and replacing each such vertex with a pair of vertices each having $d^-(\mathbf{v}_i)$ reduced by one. In the process of doing so, the value of $d^-(\mathbf{v}_i)$ of a neighboring vertex can be increased by one, which will generate a chain of such replacement which continues until it either closes onto itself or reaches a boundary. This process is continued until all vertices have $d^-(\mathbf{v}_i) = d^+(\mathbf{v}_i)$, at which point the graph is termed regular. The vertices of this graph are then put into correspondence with points in a planar grid consisting of squares whose sides have the average length between nodes on the reference surface. The positions of these points are then used as the initial positions for the images of nodes of the reference surface, each of which has a vertex in the original neighborhood graph **G**.

In order to find the two-dimensional map configurations at or near global minimum distortion, we applied the stochastic optimization technique known as *simulated annealing* (Kirkpatrick *et al.*, 1983). We begin with an initial map configuration found as described above, and randomize it by raising the temperature exponentially until effectively all attempted displacements are accepted. This effectively destroys the original configuration and precludes any bias of the final map configuration. Thus, the map attains a randomly distorted configuration whose energy E^{total} is greater than its minimum value. Nodes on the map are then visited randomly, and local random displacements $\mathbf{dr}_i = (dx, dy)$ in the plane of the map attempted, so as to simulate the thermal motion of a collection of particles (Metropolis *et al.*, 1953). Such displacements are evaluated to determine if they are to be accepted or rejected by calculating the change in energy that would result:

$$\Delta E\,(\mathbf{r}_i, \mathbf{dr}_i) = E^{total}(\mathbf{r}_i + \mathbf{dr}_i) \;-\; E^{total}(\mathbf{r}_i)\,.$$

We note that only those local lengths l_{ij} and angles Θ_{ijk} which would be deformed by the displacement $\mathbf{dr}_i$ need be considered in this calculation, since only those will contribute to a change in the energy. A displacment $\mathbf{dr}_i$ resulting in a negative value of $\Delta E\,(\mathbf{r}_i, \mathbf{dr}_i)$, correrspondin g to a decrease in local energy and an improvement in local geometry on the map as compared to the cortical surface, is always accepted. However, displacements which result in a positive value of $\Delta E\,(\mathbf{r}_i, \mathbf{dr}_i)$, corresponding to an increase in local energy and a less favorable local geometry, are also accepted with a probability given by the Boltzmann distribution:

$$P(\Delta E\,(\mathbf{r}_i, \mathbf{dr}_i)) = e^{-\Delta E\,(\mathbf{r}_i, \mathbf{dr}_i)\,/\,k_B T(t)}$$

where k_B is Boltzmann's constant and $T(t)$ is a control parameter termed the *temperature* which is a function of time t. The value of the temperature $T(0)$ at the start of annealing is set high, so that effectively all displacements are accepted, allowing for a randomization of the positions of the nodes from their initial positions. The temperature is then decreased over time, according to:

$$T(t) = T(0)\,\alpha^{t}$$

where α is chosen slightly less than unity to produce a slow exponential decrease in temperature over time, hence the term "annealing." This results in a progressive increase in the probability that a displacement will improve the local geometry over the course of the annealing. Nevertheless, there is always some probability that a displacement which actually increases energy and degrades local geometry will be accepted. This provision makes it possible for the algorithm to escape from local minima so that it can continue on to obtain a map of absolute minimum global energy, provided that the annealing is done slowly enough, so that all the nodes can be considered to be at approximate thermodynamic equilibrium. In practice, the annealing schedule is usually determined by a compromise between maintaining this equilibrium and obtaining a reasonable rate of convergence to the desired result. Finally, when the temperature has been reduced to the point where essentially none of the displacements are accepted, the map is considered "frozen" and the annealing halted, giving the result as the final conformation of the map.

We can obtain an evaluation of the distribution of *distortion* over the map at any time during the mapping process by computing the RMS error of lengths, angles, or a combination of the two in the neighborhood of each node. If errors in both lengths and angles are combined, the contribution from each kind of distortion can be treated equally, or given a weighting according to the relative values of the longitudinal elasticity κ^{o} and torsional elasticity λ^{o} (see Appendix I). These measure of distortion in the vicinity of each node can be plotted as a function of position of the node on the map at any stage of its computation, to show which regions are undergoing the greatest distortion. In general, the amount of distortion for a given region of the map will depend upon the intrinsic curvature of the corresponding surface, with higher values of intrinsic curvature resulting in greater distortion on the map. Thus, examination of the distribution of distortion will help determine whether discontinuities should be introduced in the course of mapping, and where they should be made in order to relieve the distortion. Unlike physical unfolding processes, in which deliberate cutting or inadvertent tearing is irreversible, it is possible to map the same surface several different ways, thereby obtaining several different mappings which can preserve the geometry of different regions of the cortical surface. From such distributed measures of distortion, the mean RMS distortion can be computed for the map as a whole, so as to provide a single measure of the accuracy of the map in representing linear and angular measures of surface geometry. We provide examples of both of these approaches in the **Applications** which follow.

2.4 Coloring the Map with Data

During the course of mapping, the image of the current map conformation is displayed either as a series of contours of section by drawing line segments joining the mapped nodes on the same section, or as a polygonal mesh by drawing the line segments between each mapped node and its neighbors. This permits inspection of the map during the computation. Later, once the final map conformation is obtained, any data which were been encoded along with the map nodes can be projected so as to "color" the map with data, with the color drawn chosen to identify different categories of data, or to represent gradations in one or a two categories of data through use of an appropriate gray scale or color map. These colorings can be used to display information about the quality of the mapping, such as the regional linear or angular distortion, or actual anatomical or physiological data or landmarks.

Different kinds of data are plotted differently according to their spatial distribution on the surface being mapped. If the data are obtained for discrete points, such as the locations of labelled cells, a single point can be plotted at that location, with bilinear interpolation being used to estimate the mapped position of data which does not coincide with the grid. Curvilinear distributions can be drawn as a series of line segments whose endpoints either coincide with or can be obtained by interpolation from the mapped nodes. Areal distributions are typically drawn by construction of a suitable color map for the intended display device, and then plotting a series of Gouraud–shaded polygons across areas of the map for which data exists (Foley and Van Dam, 1982). This technique has the advantage that dense data, such as may be obtained through video or other imaging techniques, can be judiciously reduced to a sampling of data values at each mapped node.

If more than one data set exists for a given map, multiple copies of the map are created and colored separately in order to facilitate comparison of the various data sets. Since the same computed map is used in all the colorings, all differences observed between maps will be due to differences in the data, and not in the mapping itself.

3. Applications

To provide demonstrations of the accuracy and utility of this technique, we applied it to obtain mappings of geometric surfaces as well as the macaque and human cortices. The geometric surfaces were selected to provide illustrations of how the algorithms performed in cases where the shape of the surface to be mapped was well defined and for which reasonable outcomes could be predicted. The cases for the macaque and human cortices were selected so as to provide a comparison of the results obtained to the maps constructed by hand for the macaque, as well as a demonstration that the technique could successfully unfold the highly convoluted human cortex.

3.1 Mapping of Geometric Surfaces

Model geometric surfaces, such as portions of a plane, a cone, and a hemisphere, were mapped during the initial development of these algorithms in order to evaluate their performance. These models were chosen since maps of their surfaces could readily be predicted in each case from knowledge of their symmetry and curvature. Intuitively, we expect to be able to map a region of the plane onto another plane without distortion. This is possible since both surfaces have zero mean and Gaussian curvature, which permits one to be mapped onto the other with zero distortion. The cone, which can be constructed from cutting, bending, and glueing a planar surface, also has zero Gaussian curvature, but nonzero mean curvature arising from bending of the surface during its construction. This mean curvature can be reduced to zero, however, by introducing a single discontinuity so as to allow an unbending of the surface, allowing for the surface to be mapped with zero distortion. In contrast, the hemisphere has a constant and nonzero mean and Gaussian curvature at every point, so that it cannot be constructed by a finite cutting, bending, and glueing of a planar surface. Thus, we cannot reduce this curvature to zero by the introduction of any finite discontinuity, and we expect some distortion in mapping the hemisphere onto the plane. The distribution of this distortion should reflect the symmetry of the hemispherical surface.

In the following, we present the details of the mapping of the cone and the hemisphere. These mappings were performed with approximately the same number of nodes, the same values of parameters, and the same coding of the mapping algorithm to facilitate comparison. After some initial exploration of the parameter space, the following values were chosen. The value of the longitudinal elasticity κ^o was chosen to be 1.0 and the torsional elasticity λ^o was chosen to be equal to 0.0769, so as to provide a greater emphasis on the preservation of lengths than on the preservation of angles. The value of the exponential base α was chosen to be 0.95 during annealing. All other parameters were chosen adaptively by the algorithm. In both cases, the map was initialized with all nodes at the origin of the plane, so that information regarding the *topography* of these surfaces was completely absent from the map initially. We expected the algorithm to reconstruct a topographic map of the reference surface based on local measures alone.

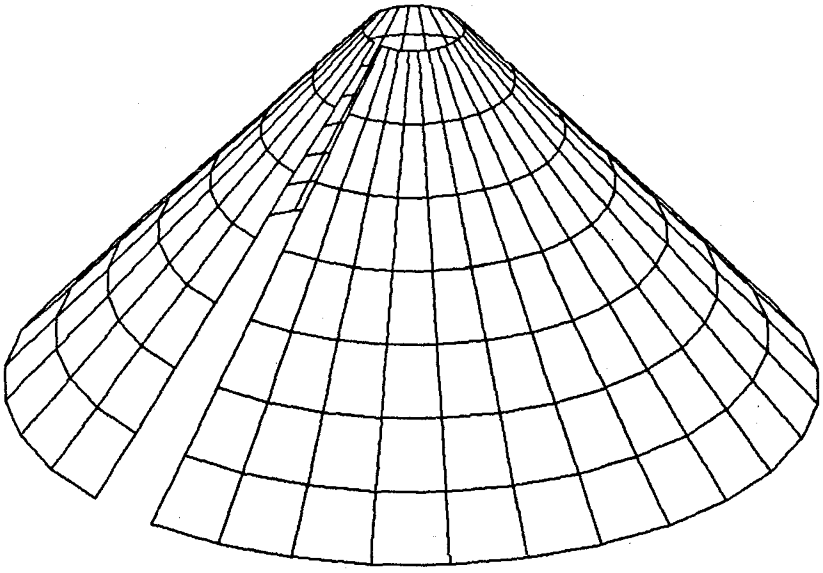


Figure 32. The surface of a cone.

The surface of a cone (Figure 32) has principle curvatures which either lie along straight lines or along circles at each point, so that the surface has nonzero mean curvature but zero Gaussian curvature (excepting the apex). Thus in principle an isometric mapping of the cone can be obtained, provided that the mean curvature can be reduced to zero in some way. This is most readily accomplished by introducing a single discontinuity in the surface, allowing the cone to "unfold" during the mapping so that its entire surface (excepting the apex) can come to lie in the plane. Such an unfolding can be demonstrated by using a paper model of a conical surface into which a single cut is made. Thus, an isometric mapping is possible from the surface of the cone onto the plane, and all measures of geometry can in principle be preserved. Thus, we expected that the map would be identical to the reference surface, up to an arbitrary uniform translation, rotation, or reflection. The latter can occur since measures of local geometry are based on the relative position of a given node with respect to its neighboring nodes, and will therefore be invariant under a uniform translation, rotation, or reflection.

We mapped a conical surface using a sample of nodes consisting of a polar grid having 8 nodes per radius and 33 nodes per cycle (Figure 33), from which the values of the reference lengths and angles were determined. The results of this mapping are shown in Figure 33. The map was initialized with all nodes at the origin, and the temperature increased so as to permit the nodes to "evaporate" from this singularity (Figure 33A). The temperature was exponentially increased so as to allow at least 95 percent of all displacements to occur, resulting in a rapid expansion of the map (Figure 33B,C). This continued until the map had attained a distorted but nevertheless topographic organization (Figure 33D). As the temperature was reduced, these distortions progressively decreased, resulting in a contraction of the map to an approximately correct size (Figure 33E). Finally, when the temperature had been reduced to a value such that fewer than 0.1 percent of displacements are accepted, the map was considered finished and the annealing halted, giving the final result (Figure 33F).

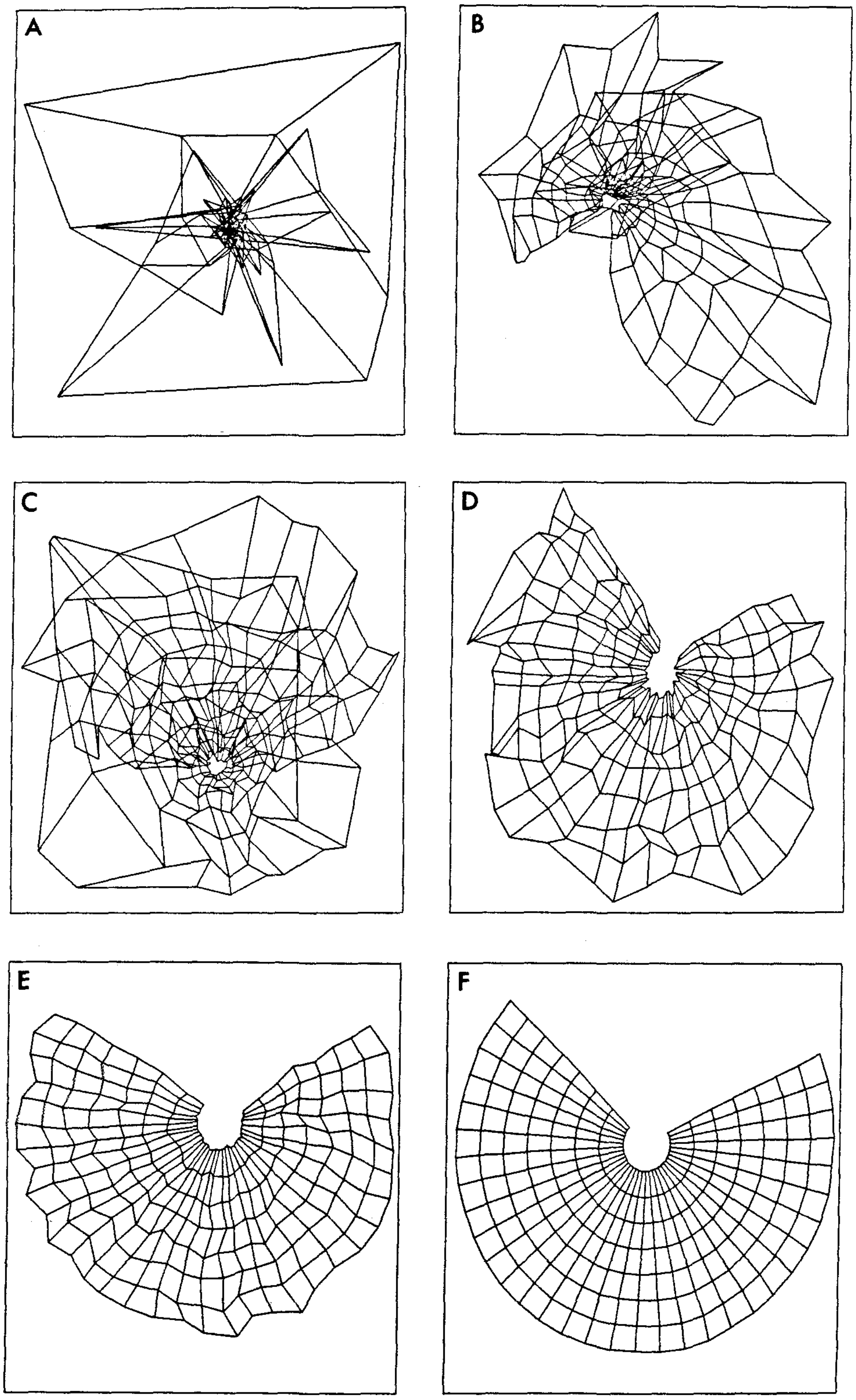


Figure 33. Map of the cone. **A–F**: Conformations 1, 100, 200, 300, 400, and 665 are shown.

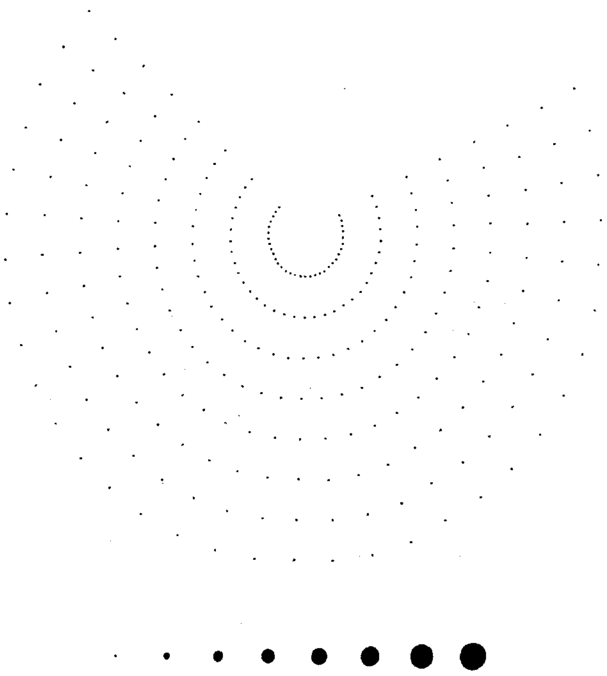


Figure 34. Distribution of linear and angular distortion for the map of the cone. The distribution of distortion over the map is shown by plotting a filled circle whose radius is proportional to the combined local linear and angular distortion at the position of each node. The scale (below) shows RMS distortions of 0.0 to 0.70, in increments of 0.10, with larger dots signifying greater linear or angular distortion. Here the dots are barely visible, representing the almost zero distortion of this map.

Since conical surfaces have zero intrinsic curvature, they can in principle be mapped onto the plane with zero distortion. We computed the sum of RMS linear and angular distortions in the neighborhood of each node of the map, and plotted this distribution of this combined distortion in Figure 34. In addition, the mean linear and angular distorion for the entire map was computed. We found an average of 0.00014 linear RMS distortion and 0.0011 angular RMS distortion per node for this map. These distortions are negligibly small, and could be made arbitrarily smaller by choosing α closer to unity, so as to produce a more gradual annealing. Note also that the linear distortion is an order of magnitude smaller than the angular distortion, in agreement with our initial choice of κ^{o} and λ^{o} to emphasize the preservation of linear measures.

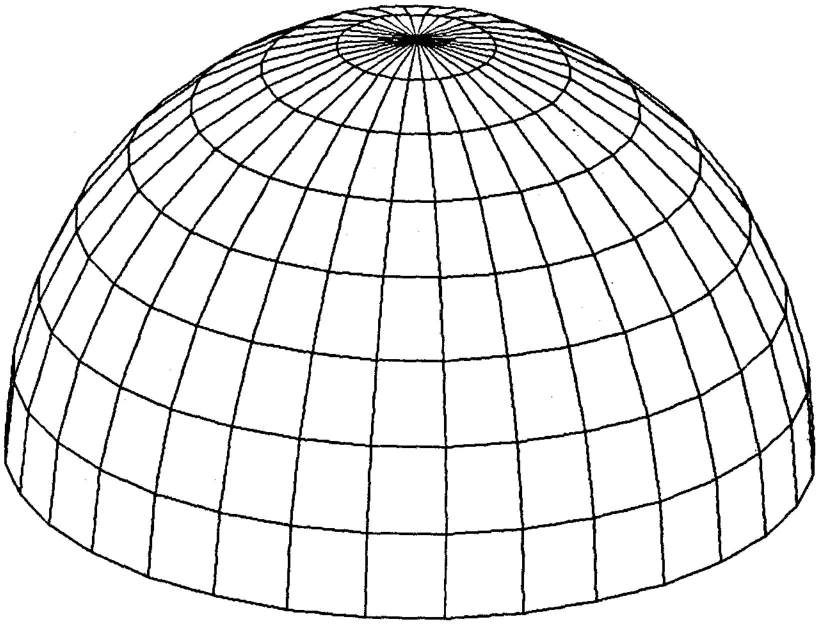


Figure 35. The surface of a hemisphere.

The surface of a hemisphere (Figure 35) has principle curvatures which lie on circles at each point, such that the surface has constant nonzero mean and Gaussian curvature. Unlike the cone, which offered a natural choice for the introduction of a discontinuity so as to permit the surface to unfold onto the plane, no such choice exists for the hemisphere. While it is possible to introduce a series of cuts along longitudes of the hemisphere at regular intervals, thereby reducing the total curvature along continuous curves of constant latitude, such cuts destroy the continuity of the map. Thus, we attempted the mapping of the hemisphere onto the plane without introducing any such discontinuities. Since it is possible for such surfaces of constant nonzero Gaussian curvature to be mapped conformally, but not isometrically, onto the plane, we expected that angles would be preserved while lengths would be distorted. As a consequence, we expected that the orthogonal intersections of longitude and latitude on the hemisphere would be preserved on the map. In addition, we expected that lengths on the map would on average have zero deviation from their corresponding lengths on the hemisphere, with a central region of compression and a peripheral region of expansion. As in the case of the cone, the map is determined up to an arbitrary uniform translation, rotation, or reflection.

We mapped a hemispherical surface using a sample of nodes consisting of a polar grid having 8 nodes per latitude and 32 nodes per longitudinal cycle (Figure 35), from which the values of the reference lengths and angles were determined. Except for the use of this hemispherical surface to determine the reference lengths and angles, the mapping was identical to that used in the case of the cone described above. The result of applying the mapping to the hemisphere is shown in Figure 36. The map was initialized with all nodes at the origin, and the temperature increased so as to allow the nodes to "evaporate" from this singularity (Figure 36A). The temperature was then increased so as to allow at least 95 percent of all displacements to occur, resulting in a rapid expansion of the nodes from the origin (Figure 36B,C). This continued until the map had attained a distorted but nevertheless topographic conformation (Figure 36D). As the temperature was exponentially reduced, these distortions progressively decreased, resulting in a contraction of the map to an approximately correct size (Figure 36E). Finally, when the temperature had been reduced to the point where less than 0.1 percent of displacements are accepted, the map is considered finished and the annealing halted, to give the final result (Figure 36F).

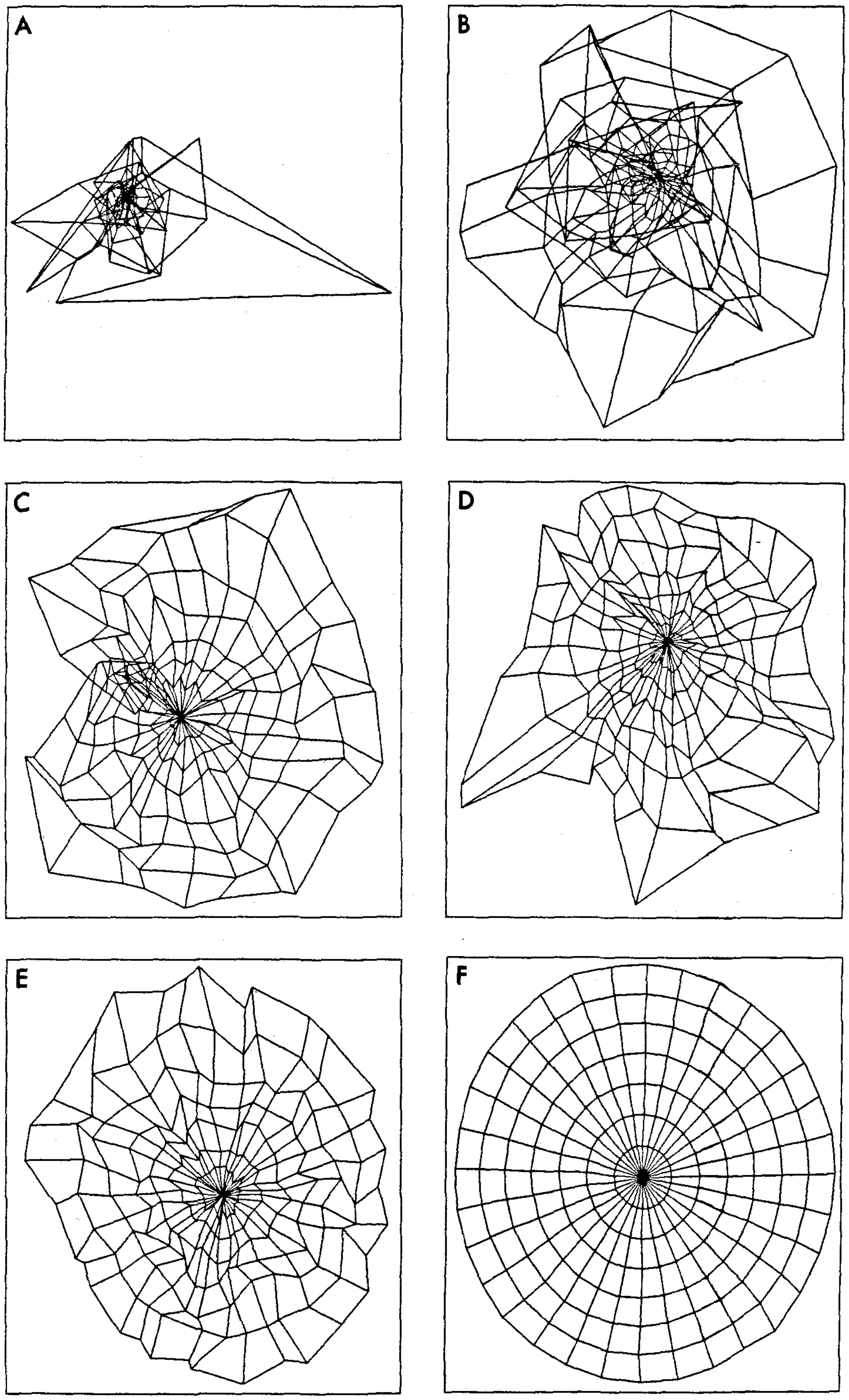


Figure 36. Map of the hemisphere. **A–F**: Conformations 1, 10, 50, 100, 200, and 400 are shown.

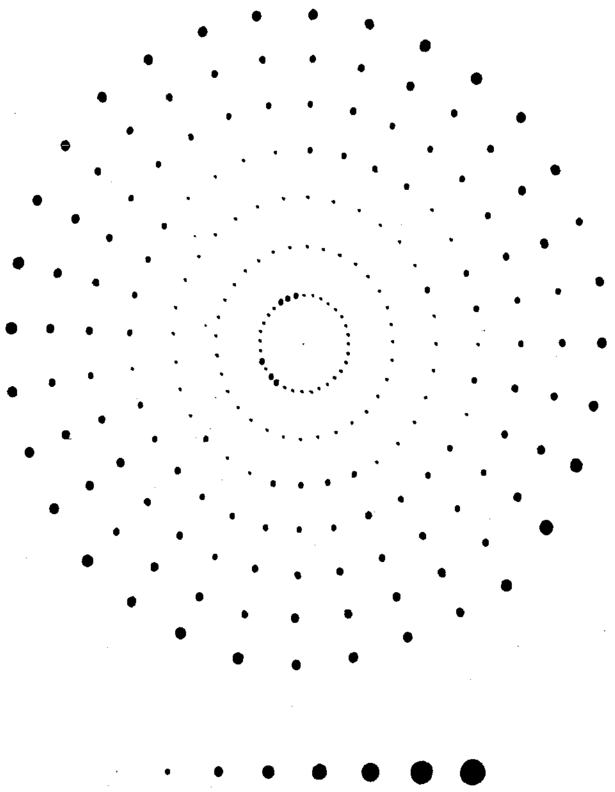


Figure 37. Distribution of linear and angular distortion for the map of the hemisphere. The distribution of distortion over the map is illustrated by plotting a filled circle whose radius is proportional to the combined local linear and angular distortion at the position of each node. The scale (below) shows RMS distortions of 0.0 to 0.70, in increments of 0.10, with larger dots signifying greater linear or angular distortion. Here the dots are barely visible, representing the almost zero distortion of this map.

The distortion is plotted as a function of location on this map (Figure 37). We found an average of 0.053 linear RMS distortion and 0.054 angular RMS distortion per node for this map. This modest linear and angular distortion demonstrates that the algorithm is capable of obtaining mappings near the theoretical minimum of distortion, even for intrinsically curved surfaces. If we examine the spatial distribution of the distortion (Figure 37), we note that it is largest at the periphery, resulting from extension of distances on the map relative to those on the surface of the hemisphere. Thus, the map is like an elastic model of the hemisphere which has been stretched starting from its edges in order to flatten it onto the plane. Clearly, more distortion is required in order to flatten the edges than the center of such an elastic model. We note that the intersections of lines of lattitude and longitude of this map are approximately orthogonal, just as they were on the surface of the hemisphere, in agreement with our expectation that the map should be conformal. In fact, this mapping is qualitatively similar to the stereographic projection of the hemisphere, which is a conformal mapping (Hilbert and Cohn–Vossen, 1952).

Since the above mappings were obtained using the same mapping algorithm with the same choice of parameters and approximately the same number of nodes sampled from the reference surfaces, only the differences in measurements of length and angle from those surfaces can account for the differences in maps obtained. Since the longitudinal and torsional elasticity is independent of the choice of particular node sampling, we do not expect differences in the position or density of nodes to substantially alter the outcome, provided that the sampling of a surface by the nodes is sufficient to accurately represent the geometry of a surface. Finally, we note that our mapping algorithm is adaptive to the surface being mapped, producing an approximately isometric map when possible, and an approximately conformal map when it was not.

3.2 Macaque Visual Cortex

Having observed that our computational algorithms were capable of producing reasonable mappings of such geometric surfaces, we next attempted to obtain a mapping of layer IV of the entire striate cortex (area V1) of the macaque monkey. This was to be the first computational mapping of this particular area, and the first such mapping of any area of mammalian cortex in its entirety.

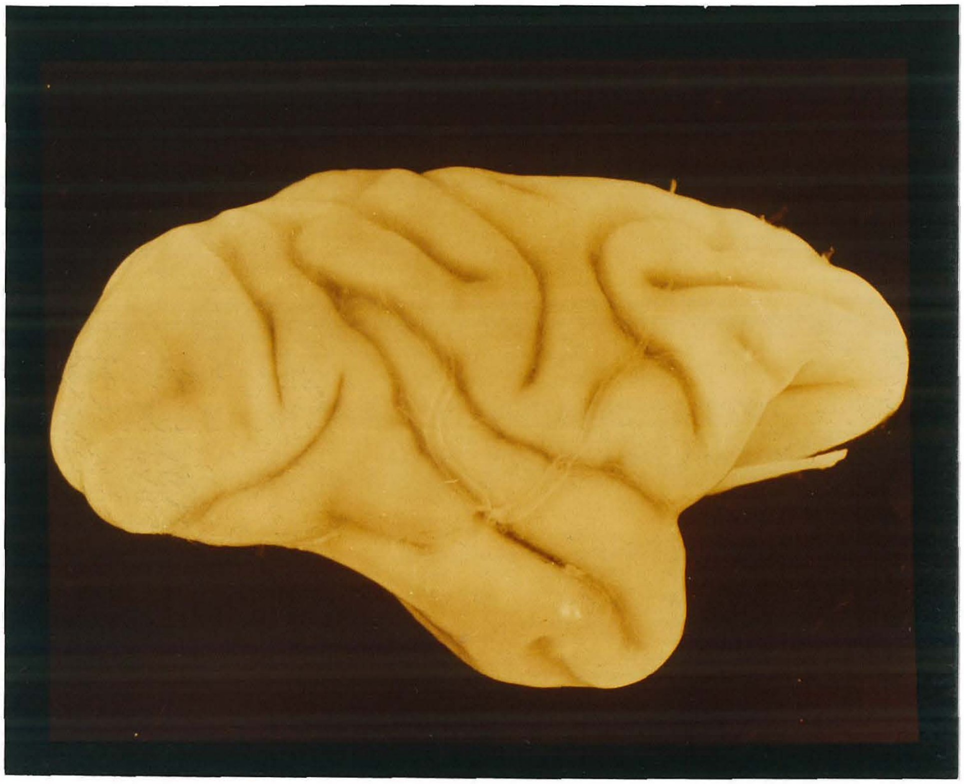


Figure 38. Lateral view of right hemisphere of macaque cortex. The striate cortex is located posteriorly (left), and includes the smooth operculum (shown) as well as a substantial infolding of the medial surface known as the calcarine sulcus (not shown).

This cortical area presents a variety of local surface characteristics, including a large, relatively flat region on the exposed posterior surface of the cortex known as the operculum, as well as the highly infolded calcarine sulcus (Figure 38). We chose to map the surface described by layer IV through this region of cortex, in part because it is thought to be the least likely of all layers to undergo compression or expansion due to folding of the cortex into sulci and gyri (Van Essen and Maunsell, 1980), and it is readily identified in Nissl stained histological sections of the monkey brain. The layer IV contours of a series of horizontal sections spaced 500 microns apart on average were digitized, smoothed, aligned, and sampled at intervals of 250 microns. The resultant 3,369 nodes were then reconstructed into a three–dimensional reference surface. For each such node, measures of length and angle were obtained to characterize the local geometry in the neighborhood of the node. The mapping was performed using the same code and parameters as were used to map the geometric surfaces above, with the exception that κ^{o} and λ^{o} were both set to unity.

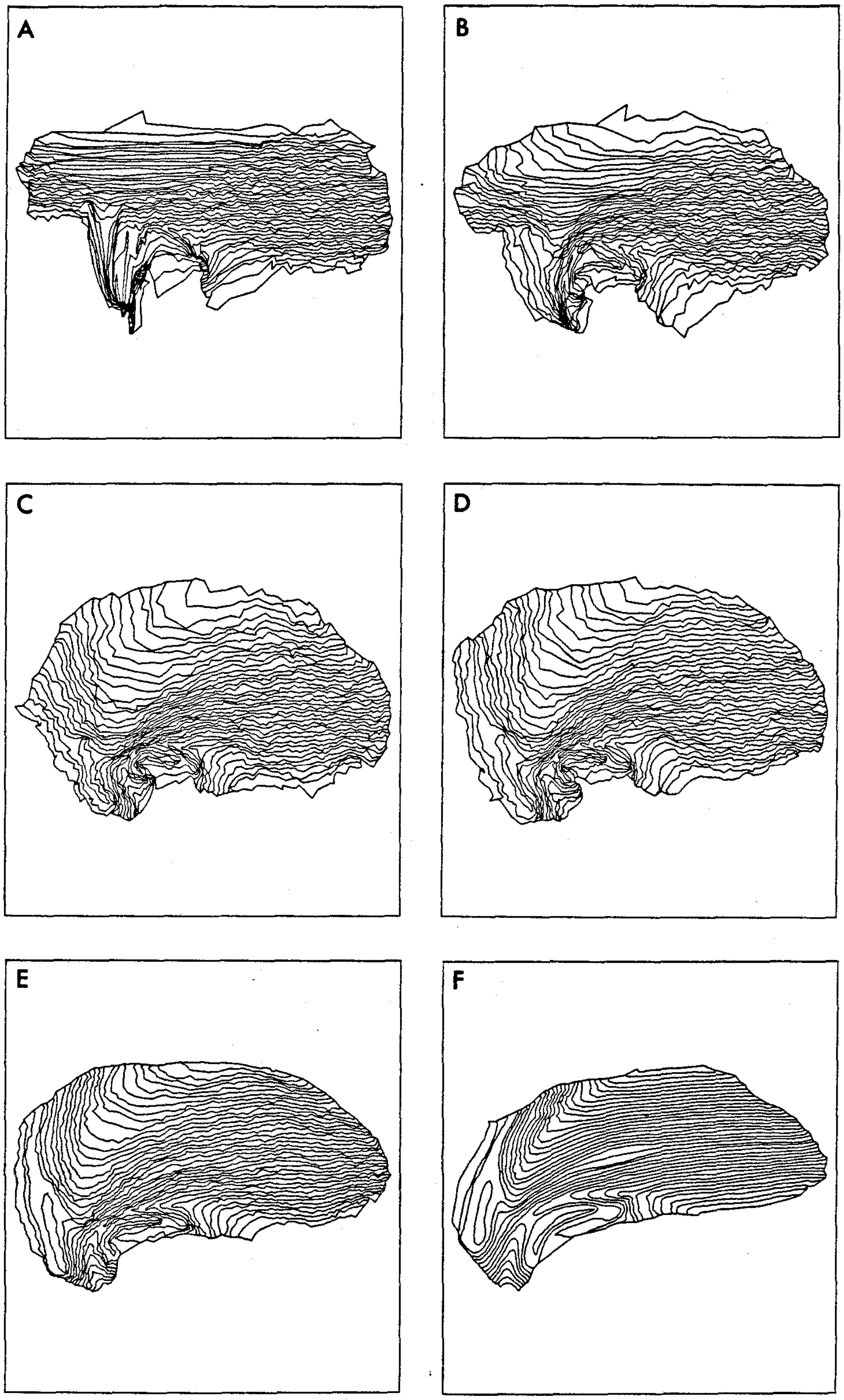


Figure 39. Map of macaque striate cortex (V1). **A–F**: Conformations 1, 10, 50, 100, 150, and 250 are shown. The contours shown are the images of the contours of section. This was the first computational map ever made of the entire macaque striate cortex.

Rather than begin the mapping with all points at the origin, as was done for the mapping of the geometric surfaces above, here we chose to begin with an estimate of the unfolded map (Figure 39A). This estimate was generated by using the topological sorting algorithm described in the **Methods** to move each point on the map so as to obtain correct local topography. Beginning from this initial estimate of the map, the temperature was exponentially increased until 95 percent of all displacements were accepted. During this heating, the map "melts" from the initial estimate, undergoing significant conformational change as it does so (Figure 39B). Once the map equilibrates at this elevated temperature, annealing begins with an exponential decline in temperature (α = 0.95). Over the course of the mapping, further changes in conformation of the map can be seen as the regions corresponding to the operculum and to the dorsal and ventral aspects of the calcarine sulcus become clearly defined (Figure 39C,D,E). It is noteworthy that during the mapping the large scale structure or organization is the first to emerge, followed by medium scale structure, and so on, until the positions of the nodes are determined at the smallest scale. The emergence of structure in such a coarse–to–fine fashion is characteristic of annealing, which finds optimal configurations across a continuum of scales by allowing the system to wander over configurations having energy differences of no more than $k_B T$. As this annealing progresses, the progressively lower temperature results in the selection of configurations which have progressively lower energy. When the temperature had been reduced to the point that less than 3 percent of the proprosed displacements were accepted, the map was considered to have crystallized and annealing is halted.

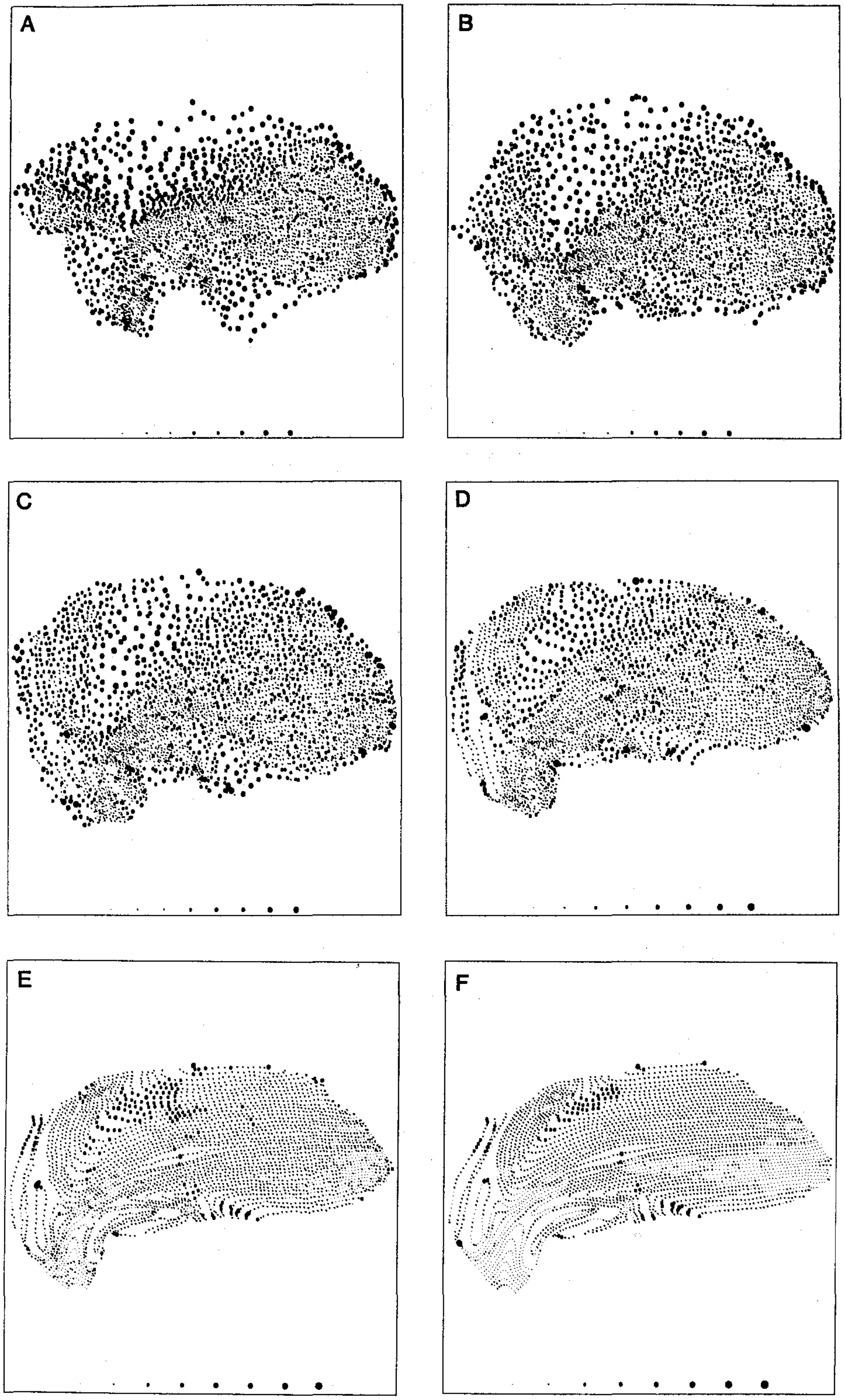


Figure 40. Change in local distortion over the course of annealing. See next page for caption.

Figure 40. Change in local distortion over the course of annealing. **A – F**: The distribution of distortion over the map for successive conformations is illustrated by plotting a filled circle whose radius is proportional to the combined local linear and angular distortion at the position of each node. Large dots signify large linear or angular distortion, while small dots signify low distortion. The same scale has been used in each plot, representing RMS distortions of 0.0 to 2.50, in increments of 0.357 . Distortion initially spreads from the boundaries to the interior during heating from conformation 10 to 50 (**A, B**), and then decreases during subsequent annealing over conformations 100, 150, 200, and 250 (**C – F**).

The gradual spread and reduction of distortion of local geometry of the map over the course of annealing is shown in Figure 40. The distortion is initially largest at the boundaries of the maps, and can be seen to spread to the interior during the course of the initial heating as the energies of nodes on the interior equilibrate with those on the boundary of the map, so as to randomize the initial geometry. The distortion is seen to progressively decrease during the subsequent annealing as the temperature is exponentially decreased. The distribution of distortion is plotted as a function of location on the map in Figure 41. The contributions of linear and angular errors to the combined distortion are also shown. Note how the linear and angular contributions can occur at the same or at different locations on the map, illustrating how this mapping technique is capable of adaptively combining linear and angular distortions so as to obtain an overall minimum of distortion across the map. If we examine the spatial distribution of the distortion (Figure 41D), we note that it is largest along the "crease" corresponding to the boundary of the operculum with the dorsal calcarine cortex. We suspect that at least some of this distortion arises from undersampling of the surface in this acutely folded region. Such undersampling can lead to inaccuracies in the reconstruction, and concommitant distortions in the subsequent unfolding.

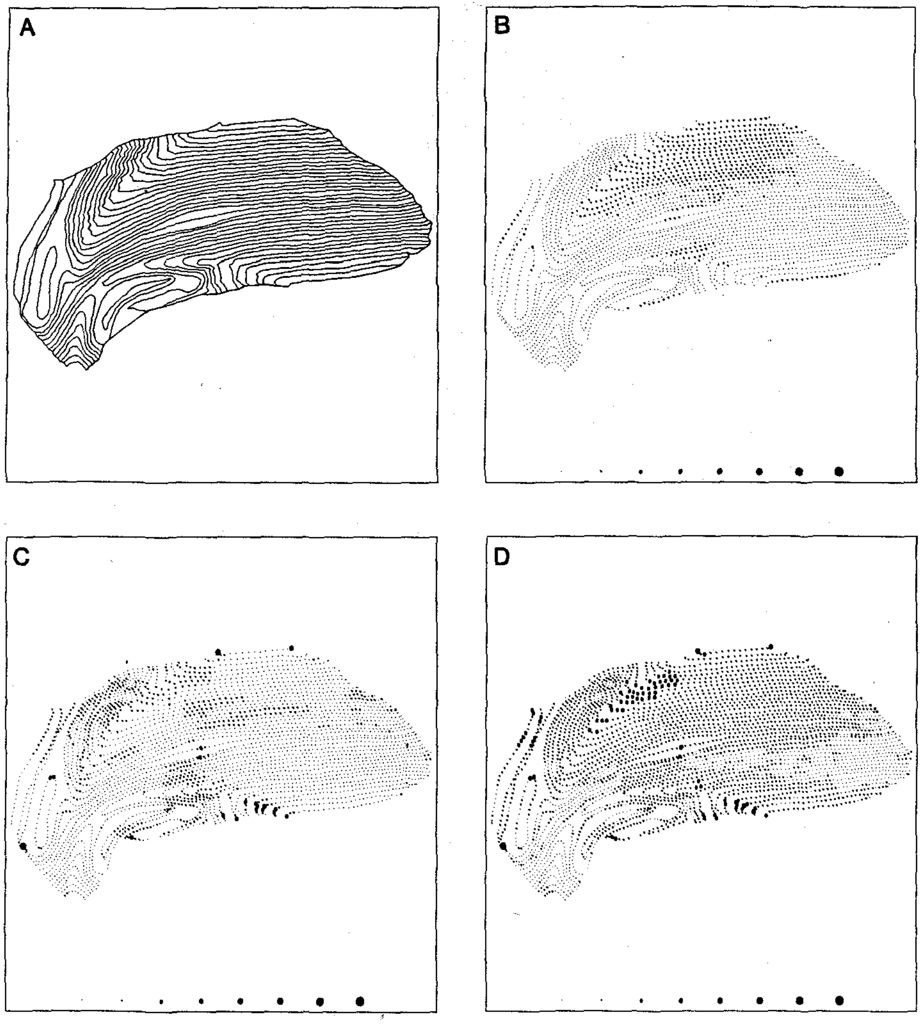


Figure 41. Contributions of linear and angular distortion to the distribution of distortion on the final map. (A) Final conformation of the map, showing the contours of section. (B – D) Distribution of distortions plotted as a function of node position, as in previous Figure. (B) Distribution of linear distortion. (C) Distribution of angular distortion. (D) Distribution of combined linear and angular distortion. The same scale has been used in each plot, representing RMS distortions of 0.0 to 2.50, in increments of 0.357 .

The final conformation of the map has the lowest distortion. Since this surface is believed to have low intrinsic curvature (Van Essen and Maunsell, 1980), we expected the map to have little distortion. We found an average of 0.13 linear RMS distortion and 0.27 angular RMS distortion per node for this map. This distortion represents a factor of two to four reduction in distortion as compared with comparable estimates of distortion for maps produced by the manual technique of Van Essen and Maunsell (1980).

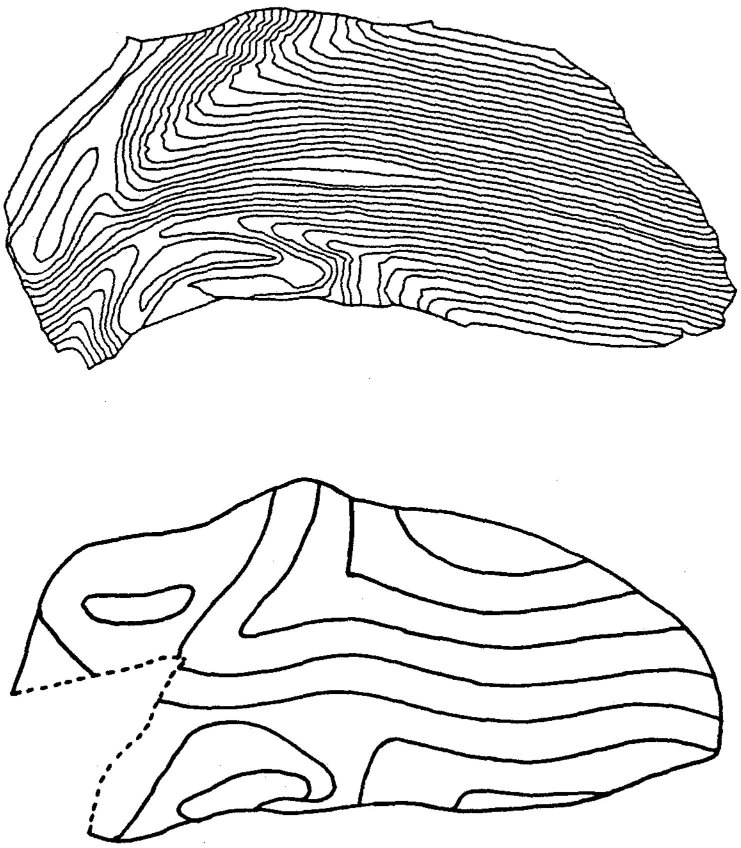


Figure 42. Comparison of computed and manual maps. The final conformation of the computed map (above) resembles the final version of a manual mapping of the same data (below, from Le Vay *et al.*, 1985). The solid lines show the contours of section and boundaries of the area in both maps. The dashed lines on the manual map represent a discontinuity introduced in order to facilitate manual unfolding, which accounts for much of the difference in shape of the two maps.

More significantly, we observe that the configurations of the contours of section of this map are comparable to those obtained with the manual technique (Figure 42). This result reaffirms the validity of that technique while demonstrating our ability to improve upon it through quantitative mapping.

3.3 Human Visual Cortex

Knowledge of the organization of the brain provides a basis for the understanding of its normal function and pathology. Much of this knowledge has come from the study of the brains of subhuman primates, whereas study of the structure and function of the human brain has been largely limited to clinical or postmortem evaluations (Allman, 1988; Allman and McGuiness, 1988; Van Essen, 1985). However, large differences in the organization of the cortex have been discovered among the primates, suggesting that such indirect and limited study of the human brain will not be sufficient to fully and independently characterize its organization (Allman, 1988; Sereno and Allman, 1989). More directed study of the human brain has been made possible by the recent development of two noninvasive brain imaging techniques: magnetic resonance imaging (MRI), which reveals the anatomical structure of the brain, and positron emission tomography (PET), which allows us to measure changes in cerebral blood flow believed to be correlated with the activity of neurons (see Brownell *et al.*, 1982 for a review). By applying these two techniques to individual subjects, it is possible to determine not only the change in physiological activity due to, say, visual stimulation, but to determine the distribution of that activity within that subject's brain (Mora, Carman, and Allman, 1989).

MRI produces images based on the magnetic properties of nuclei with an odd number of protons or neutrons, of which hydrogen is the most abundant in biological tissues (Norman and Brant-Zawadzki, 1985). When an external magnetic field is imposed, the magnetic moments of such nuclei can assume a low energy alignment with the field or a high energy alignment against the field. These moments can then be perturbed from this alignment by applying an alternating magnetic field in the form of radiofrequency waves, which can lead to coherent oscillation and resonance of the nuclei with the alternating field. The frequency at which this resonance occurs will depend on the strength of the constant component of the external field, permitting a spatial gradient in the external magnetic field to produce a spatial gradient of resonant frequencies within the tissue. When the radiofrequency waves are terminated, the coherent oscillation of the nuclei decays rapidly, and is accompanied by the emission of radiowaves whose frequency will depend on the spatial location of the nuclei and whose magnitude will depend on the density of nuclei per unit volume. Thus by recording the emitted radiowaves and applying Fourier decomposition, one can determine the density of hydrogen nuclei at each spatial location, provided certain other tissue-specific parameters are known (Brownell *et al.*, 1982; Norman and Brant-Zawadzki, 1985; Oldendorf, 1985; Sokoloff, 1985).

When applied to the human head, the MR images obtained show a wealth of internal structure. These images are displayed using a gray scale such that regions with largest magnitude, corresponding to the greatest density of hydrogen, appear white, while regions with the smallest magnitude, corresponding to the least density of hydrogen, appear dark. For the spin-echo imaging technique used in acquiring data for this study, these differences in hydrogen density are mostly due to differences in the amount of water the tissues contain. Thus, the grey matter of the brain generally appears lighter than the white matter, which in turn appears lighter than the cerebrospinal fluid or bone (Bradley *et al.*, 1983). Using such images, it is possible to obtain outlines of the external surface of cortex, and in many cases, the contour of the boundary between the grey and white matter. Provided that the MRI images have sufficient contrast and resolution, they can provide contours representing the intersection of such cortical surfaces which are equivalent to those that would be obtained from histological sectioning of the brain.

Just as MRI provides the capability of noninvasively imaging the anatomical structure of the human cortex, PET makes it possible to noninvasively detect changes in regional cerebral blood flow (rCBF) believed to be correlated to changes in neuronal activity (Raichle *et al.*, 1976; Lassen *et al.*, 1978; Fox *et al.*, 1984; Fox and Raichle, 1984; Fox and Raichle, 1986; Paulson and Newman, 1987). Cerebral blood flow is monitored by injecting positron-emitting compounds such as $H_2{}^{15}O$ into the blood circulation. Positrons emitted by such compounds travel short distance before anihilating with an electron to produce a pair of gamma rays travelling in opposite directions. If these gamma rays intersect detectors consisting of scintillation crystals coupled to photomultiplier tubes, their line and time of travel can be determined and this information used to localize the anihilation event in space. The frequency of these events from a given unit volume will be proportional to both the activity of the radionuclide and rCBF. By use of the washout curve and partition coefficients, rCBF can be determined accurately. By arranging detectors in rings around the head, rCBF can be determined for several PET image planes simultaneously. The low spatial resolution of PET scanners (typically greater than 1 cm) can be improved through use of differential imaging, wherein scans obtained during control trials are subtracted from scans obtained during experimental trials, permitting localization of changes in rCBF to within distances on the order of 1 mm in some cases (Brownell *et al.*, 1982; Fox *et al.*, 1986).

By applying the techniques of MRI and PET to the same subject, we can noninvasively obtain correlated information about the anatomical structure and the physiological response of the brain to a variety of sensory stimuli. These measures of structure and function can be combined by bringing the MRI and PET scans into registration, and superposing the images produced by each technique using methods which have been described elsewhere (Mora, Carman, and Allman, 1989; see Appendix II). The result is shown in Figures 43 to 46, which display anatomical structure observed with MRI as an intensity or gray scale (vertical axis of color map shown on the right of each image), and the change in cerebral blood flow observed with PET as a pseudocolor scale (horozontal axis of color map). Such superpositions allow us to determine not only the changes in physiological activity due to, in this case, visual stimulation, but also to determine the region or regions of the subject's brain at which such changes take place (Mora, Carman, and Allman, 1989).

However, the extensive folding of the human cortex into a series of sulci and gyri impedes both the visualization and analysis of the organization of such physiological responses on the cortex to visual stimulation. The algorithms described previously in connection with the computational mapping of macaque cortex can be used to unfold human cortex as well. Such a mapping provides a challenging test of our cortical mapping technique, as the sulci and gyri of the human cortex are both deeper and more numerous than in any other species. In the following, I will describe the production of such an unfolded map of the

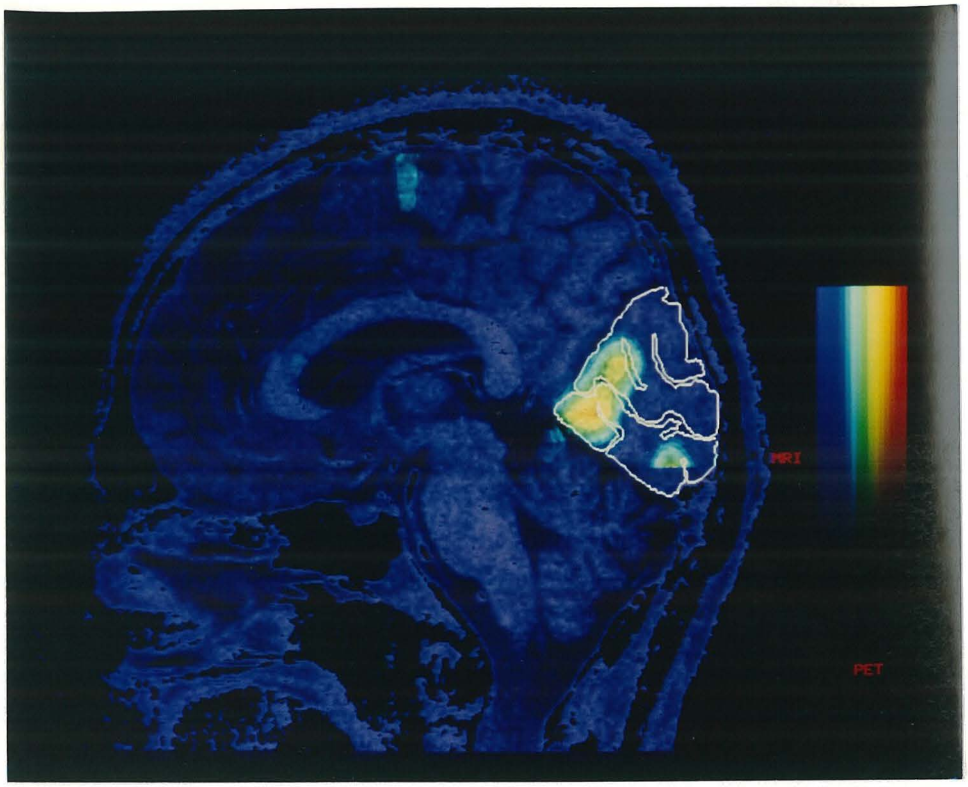


Figure 43. Superposition of anatomical and physiological measures obtained noninvasively from a single human subject. The color map used here presents anatomical structure in shades of gray, and physiological response to visual stimulation as a pseudocoloring. A medial view of the right hemisphere is shown. Blue indicates no change in cerebral blood flow, while red color indicates increased blood flow in response to visual stimulation relative to flow during fixation point controls. The response shown here was to upper field hemiannulus stimulation (see Figure 49), and had a maximum value of 13.2 ml per gm per minute after normalization. See Fox *et al.*, 1987 and Mora, Carman, and Allman, 1989 for details.

human visual cortex through the application of these techniques. While our specific application is based on data obtained from MRI and PET data, the techniques apply generally to any sources of anatomical or physiological data, and we will consider such alternate sources of data to illustrate the generality of our approach.

Several sources of information about the anatomical structure of the human

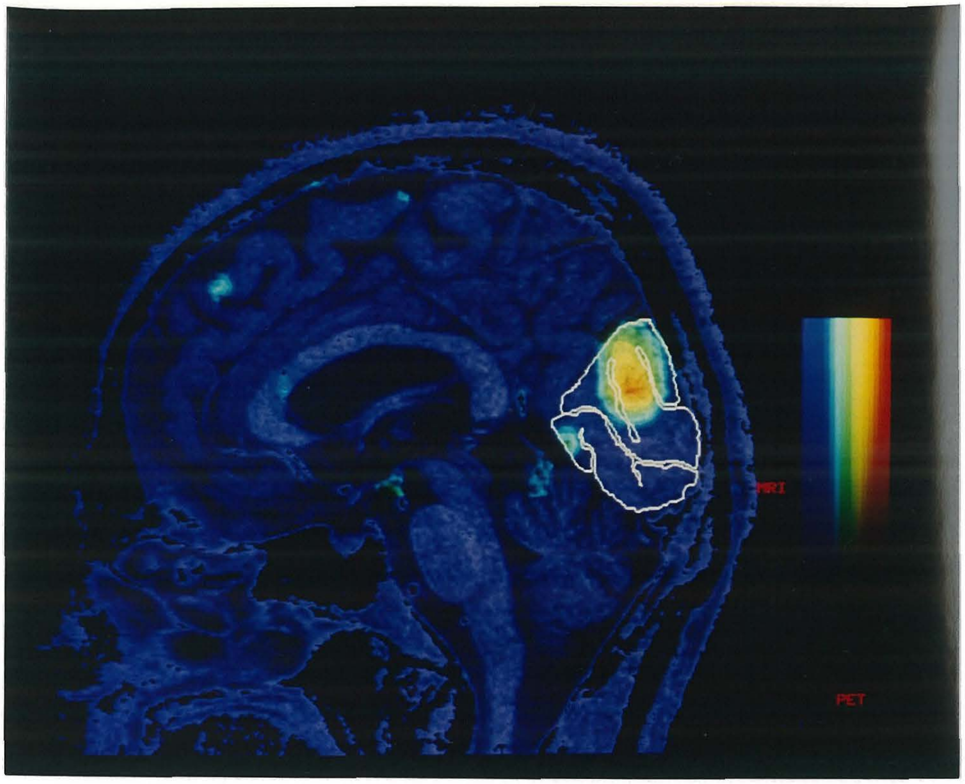


Figure 44. Superposition of anatomical and physiological measures obtained noninvasively from a single human subject. Color map and view as in previous image. The response shown here was to lower field hemiannulus stimulation (see Figure 49), and had a maximum value of 11.8 ml per gm per minute after normalization.

brain are available for use in producing an unfolded map of human cortex. For example, numerous atlases exist which contain sections of human cortex which many be suitable for mapping part of all of the human brain (Talairach and Szikla, 1967; Roberts *et al.*, 1987). In addition, human cadavar brains are available which can be sectioned and used as the source of contours of a chosen layer of the cortex, much as is done in the monkey. However, large variability in brain structure and organization between individuals limits the usefulness of maps prepared from anatomical data in one subject for the purposes of presentation and analysis of physiological data obtained from another (Schaltenbrand and Bailey, 1959; Talairach and Szikla, 1967; Stensaas *et al.*,

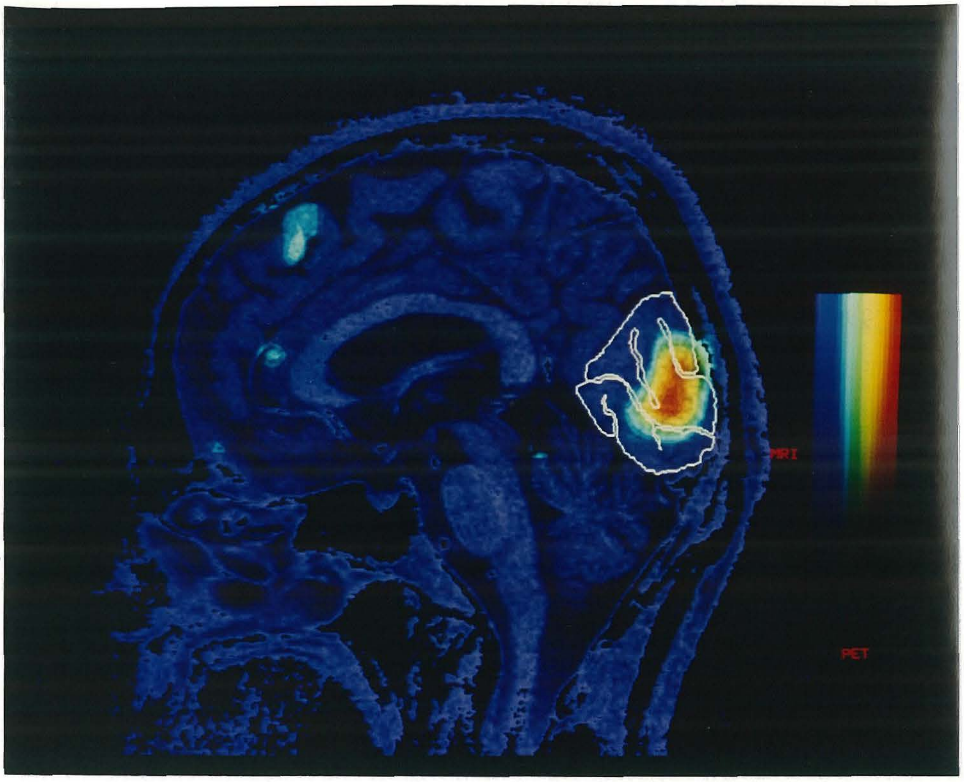


Figure 45. Superposition of anatomical and physiological measures obtained noninvasively from a single human subject. Color map and view as in previous image. The response shown here was to perimacular stimulation (see Figure 49), and had a maximum value of 13.3 ml per gm per minute after normalization.

1974). In order to minimize errors due to such individual differences, one would like to obtain anatomical and physiological data for the same subject.

Through combined application of MRI and PET in the same subject, it is possible to both noninvasively measure changes in physiological activity and to localize those changes upon images of the anatomical structure. As illustrated previously, this can be achieved through MR imaging of the subject from which physiological data is obtained. Contours of section can be digitized directly from the MRI images by use of a grapics cursor tied to a mouse, so as to obtain a sample of the chosen surface of cortex analogous to that obtained by

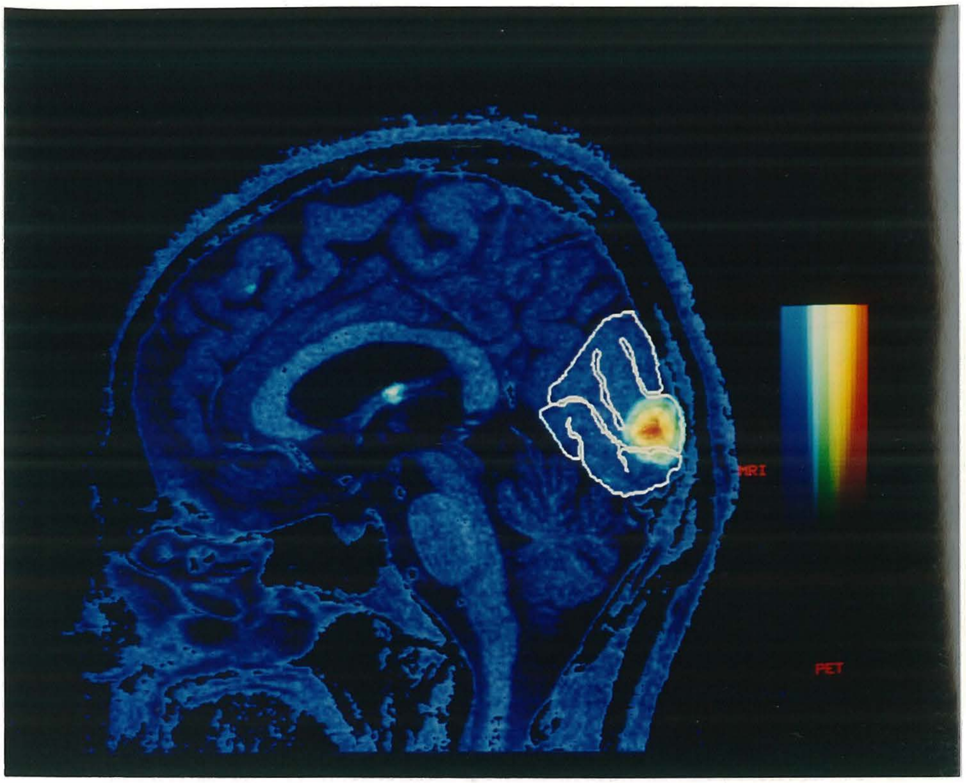


Figure 46. Superposition of anatomical and physiological measures obtained noninvasively from a single human subject. Color map and view as in previous image. The response shown here was to macular stimulation (see Figure 49), and had a maximum value of 13.4 ml per gm per minute after normalization.

digitization of contours from histological brain sections. A series of such contours, obtained from MR in parasaggital image planes spaced 2.7 mm apart, can be used as a basis for reconstructing the surface of human cortex (Figure 29). However, the precision with which these contours can be digitized is limited by both the contrast and resolution of the MRI imaging process, and by the spacing between the MRI image planes, all of which create difficulties for the reconstruction of the surface to be mapped. As a result, this reconstruction requires a greater degree of interaction than mappings of macaque cortex, whose surfaces can be sampled with greater precision from histological sections. Nevertheless, the final reconstruction of the surface of human cortex compares well with its known structure as observed in atlases or specimens of the human brain, and is therefore sufficient to provide a reference surface for producing an unfolded map of the human cortex.

The same algorithms used for the mapping of the macaque cortex can be applied to data obtained from such noninvasive proceedures as MRI and PET, allowing us to obtain the first *in vivo* mapping of the human cortex. The ability to produce such a mapping noninvasively provides a significant advantage over other techniques, such as manual unfolding (Tootell and Silverman, 1985; Olavarria and Van Sluyters, 1985), which can only be performed postmortem, and which yield data on anatomical but not physiological organization. Since the pattern of activation in response to visual stimulation was already known for the medial wall of the occipetal cortex, we decided to map the more lateral expanse of this cortex burried within adjacent sulci. We proceeded by reconstructing the pial surface of this cortex from the contours of section shown in Figure 29. The resulting reference surface, shown in Figures 30 and 31, consisted of a band around the medial wall of the occipetal cortex, which we believe consisted mostly of extrastriate cortex, as well as an portion of the intracalcarine striate cortex. This band was cut through its most anterior aspect to permit unfolding. We then proceeded to map this region with a more efficient coding of the algorithm used to compute the aforementioned maps. The parameters κ^o and λ^o were both set to unity, and the exponential base α was set to 0.95 as before.

The result of the mapping is shown in Figure 47. The elongated appearance of this map results from the numerous foldings of this circummedial band of cortex. The map is approximately 400 mm long by 17 mm wide, for a total area of

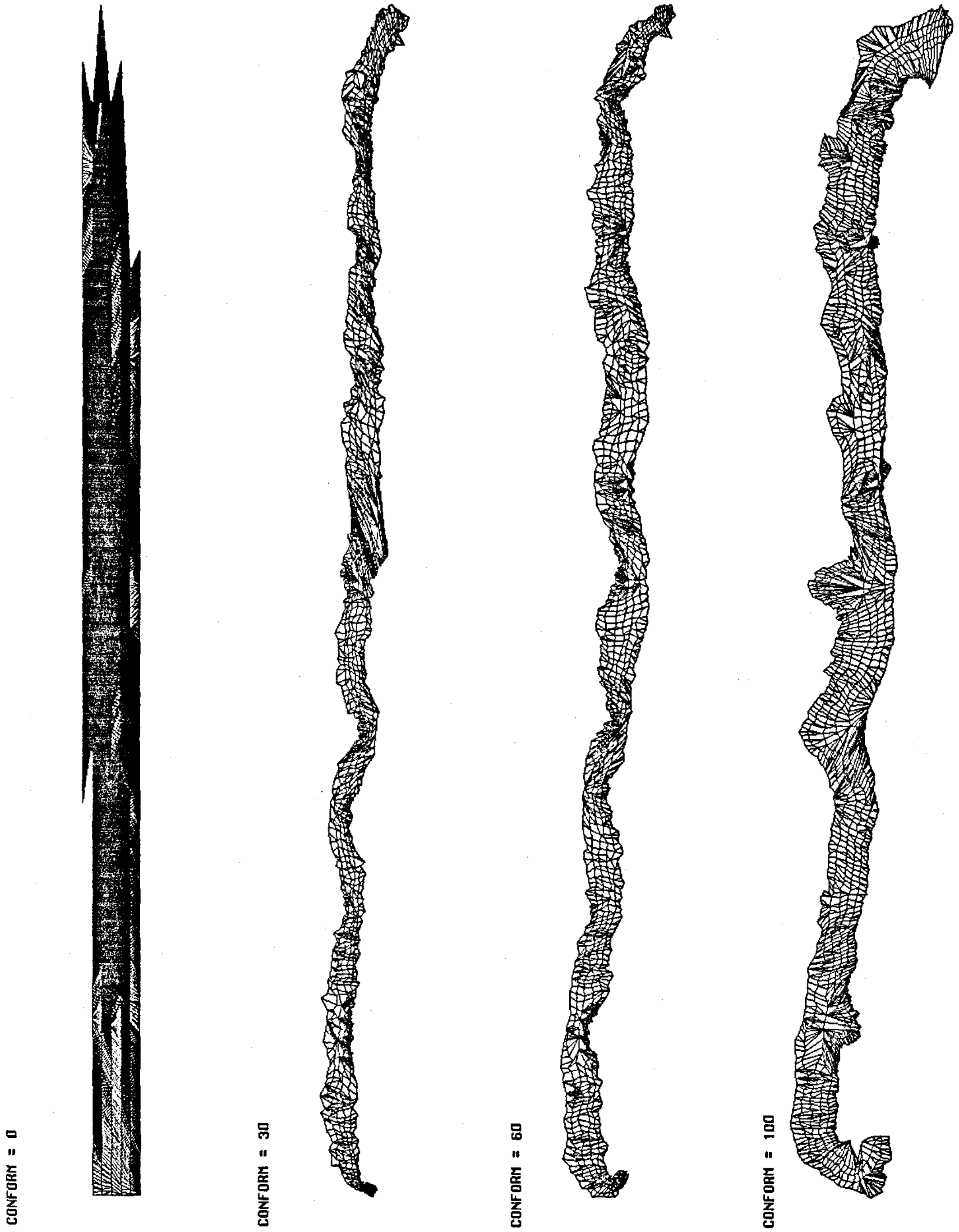


Figure 47. Map of a portion of human occipetal cortex. Conformations 0, 30, 60, and 100 are shown. The progressive reduction of distortion during the course of the mapping produces the changes in local geometry which can be observed by comparing one conformation with the next.

approximately 6480 square mm, making this the largest computational unfolding yet undertaken. The map is over five times larger than the map of the macaque striate cortex described earlier.

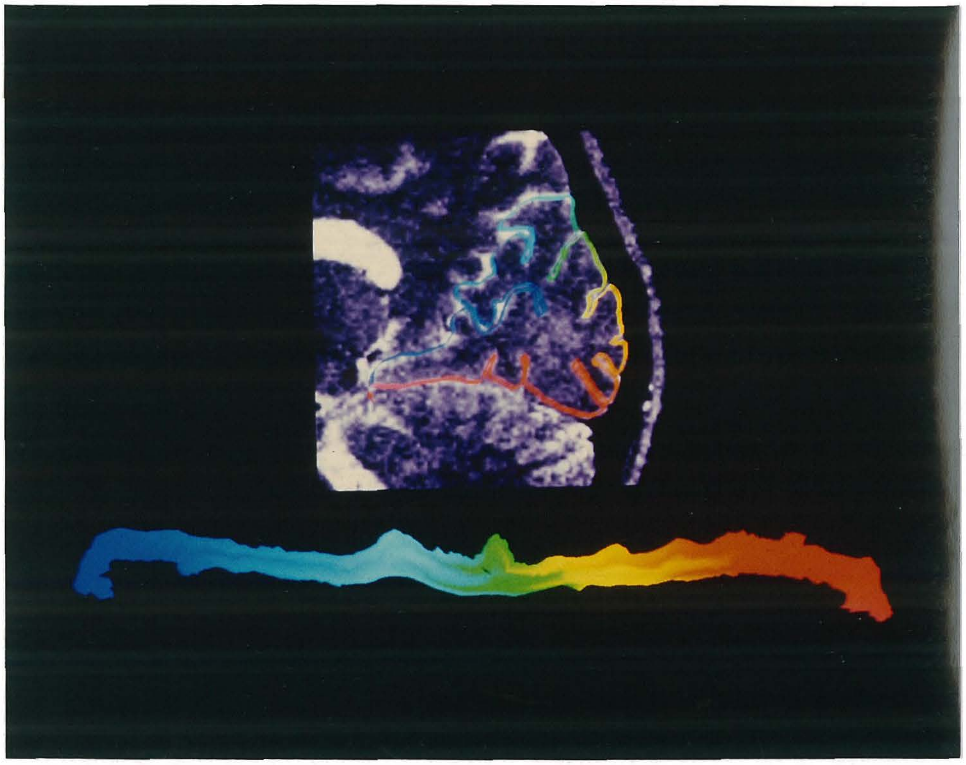


Figure 48. Relationship between the map and the surface of cortex it represents. All the contours of section have been colored according to the length along the contour (the middle contour is shown). These colors were then transferred to the unfolded map so as to display the relationship between positions on the surface of cortex and positions on the map.

The relationship between position within the occipetal cortex and position on the map is depicted in Figure 48. The extensive change in the relative extent of different sulci and gyri is evident in the lateral blending of color which can be seen on this map. Such rapid changes in shape and size of the cortical surface makes the mapping of such regions starting from such sparse samples challenging. Nevertheless, the final conformation of the map had a average linear distortion of only 0.21 and an average angular distortion of 0.31 .

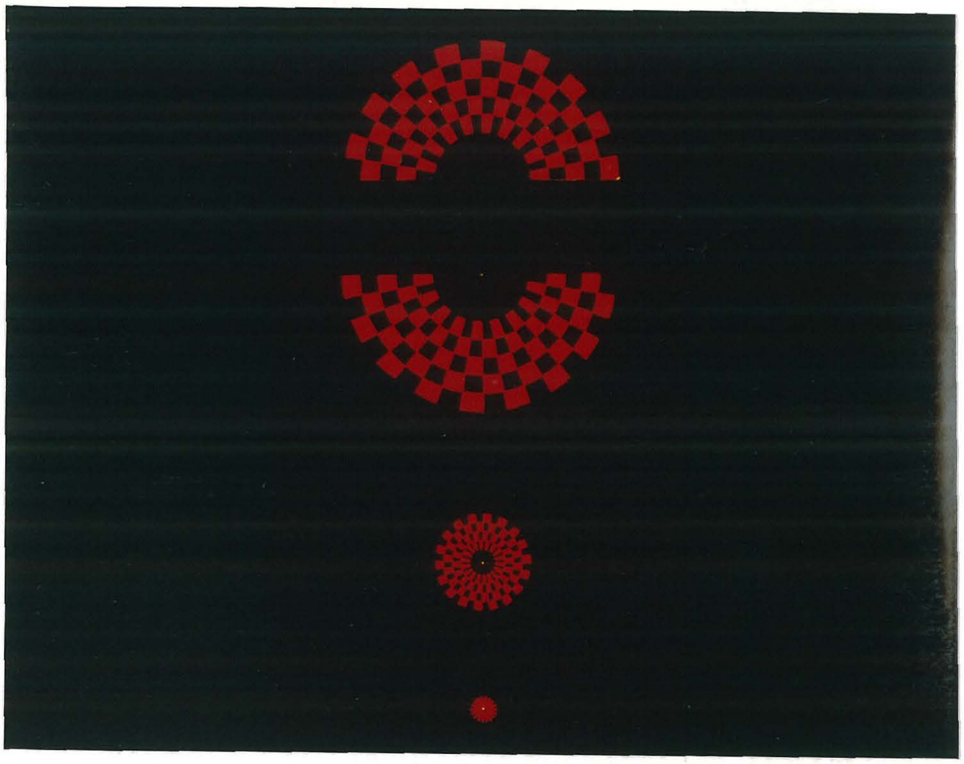


Figure 49. Stimuli consisted of red and black 10 Hz counterphased annuli. The stimuli were presented in the *upper field* (superior 5.5 to 15.5 degrees) the *lower field* (inferior 5.5 to 15.5 degrees), the *perimacular* region (1.5 to 5.5 degrees) and the *macular* region (0.1 to 1.5 degrees).

Once this map had been computed, it was colored with data obtained from experiments which have been described elsewhere (Fox *et al.*, 1987; Mora, Carman, and Allman, 1989). Changes in cerebral blood flow in response to viewing the four stimuli depicted in Figure 49 were determined relative to fixation point controls, and these differences were used to color the map of the corresponding cortical regions. We used the same color scale in all cases, so that the maps can be directly compared with one another. Differences between these colorings of the map cannot be attributed to differences in the particular surface which is mapped, since the same portion of cortex is mapped in all cases. Note also that we have plotted decreases as well as increses in cerebral blood flow in these colorings of the map, so as to provide a more balanced view of the response to these stimuli.

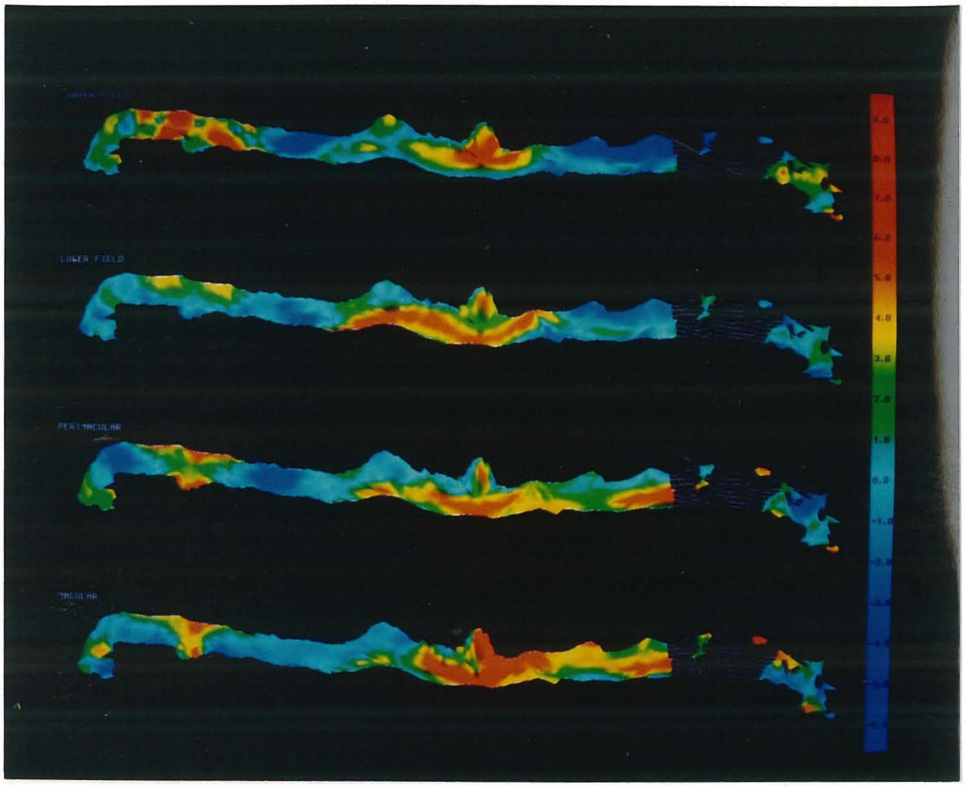


Figure 50. Change in cerebral blood flow displayed on an unfolded map of a portion of human occipetal cortex. Going from top to bottom, the maps show the response to upper field, lower field, perimacular, and macular stimulation. The scale gives the change in cerebral blood flow in ml per gm per minute, after normalization. These are the first computational unfoldings mappings of human cortex *in vivo*, and were based entirely on noninvasive techniques.

These are the first *in vivo* computational mappings of the human visual cortex based on noninvasive anatomical and physiological imaging (Figure 50). Numerous features of these maps suggest that they contain information relevant to the topographic organization of both the striate and extrastriate regions of the human. For example, complementary patterns of increased and decreased blood flow can be observed under some stimulus conditions. A careful analysis of these patterns must await a more complete reconstruction and mapping of the occipetal lobe before it will be possible to give reasonable interpretations to such data. Nevertheless, the present maps provide a clear demonstration that such noninvasive computational mapping has great potential to reveal both the anatomical and physiological organization of the human cortex *in vivo*.

4. Discussion

This work has been an attempt to attain two separate objectives, involving the development of empirical technique in parallel with the development of theoretical understanding. The first objective was to develop a series of algorithms to permit the production of unfolded, two–dimensional maps of the cerebral cortex. The second objective was to determine the general nature of the algorithms used by biological vision systems to obtain solutions to the *reconstruction problem* and the *topological problem* described earlier in this work. We have succeeded in both objectives, insofar as demonstrated by the following main results:

1. The first computational unfolding of the entire striate cortex of the macaque, as presented at the 1985 convention of the Society for Neurosciences (Carman and Van Essen, 1985);
2. The first computational mapping of physiological measures of regions of the striate and extrastriate cortices of the human, as presented at the 1989 convention of the Society for Neurosciences (Carman and Mora, 1989);
3. The development of a model for computation with flows of information based on the complex harmonic potentials; and,
4. The simulation of fusion of images differing by affine transformations using such a flow model of the primate visual pathway.

We will discuss various aspects of these empirical and theoretical advances in turn.

We note that the techniques which we have developed are by no means the only ones capable of producing unfolded, two-dimensional maps of the cerebral cortex. As yet, no quantitative comparison of these techniques has been attempted since at present the details of many of these techniques remain unpublished. Furthermore, few results from the study of anatomy or physiology of the cerebral cortex require quantitative precision of the mapping to reveal their essential nature. Typically, other factors such as the variability between animals and the small amount of data which can be obtained from each have limited more quantitative approaches (Connolly and Van Essen, 1984; Van Essen *et al.*, 1984; Tootell *et al.*, 1988). Nevertheless, such approaches might reveal a great deal of information about the successive transformations or maps within the sensory pathways of individual animals, if the proper empirical techniques are brought to bear. The quantiative techniques developed here represent a contribution in that direction. Combined with advances in anatomical and physiological techniques, such approaches may eventually permit and support functional interpretations of the successive transformations within a given sensory pathway (Knudsen *et al.*,1987; Lund, 1989; Carman, 1990).

Nevertheless we can make a few qualitative comparisons of our techniques to others which are presently available. Three others described previously in the **Introduction** have been selected as representative of the techniques currently available. We note that all of these approaches can be regarded as instances of a general cartographic problem of mapping from spaces of N dimensions to M dimensions, where M is less than N, which was first addressed by Sammon (1969). Whereas he used a gradient descent algorithm to obtain his mappings, which can be trapped in relative minima of distortion, our use of simulated annealing guarantees that our maps can proceed to global distortion minima in principle, and will attain near global minima in practice.

The techniques described by Van Essen and Maunsell (1980) rely largely on human judgement and approximation for the preservation of geometry, and thus are not conducive to the quantitative work envisioned above. The only mechanism for preservation of geometry intrinsic to this method is the preservation of length along the contours of section, which permits substantial distortions due to shearing of one contour relative to another. This can introduce significant anisotropy which will depend on the plane of section which produced the contours of the mapped surface. Also, the small random distortions which are introduced as an artifact of the mapping proceedure typically result in a "wavy" appearance of maps made in this fashion (e.g., LeVay *et al.*, 1985), making comparison of maps within or between individual animals difficult. Finally, the construction of a map of cortical area V1 can require several days of tedious effort, even for those experienced with this manual technique. In comparison, we computed an unfolded map of cortical area V1 using our algorithms in only six hours. In that time, the computer made over 49 million comparisons of lengths and angles in the course of attempting over 6 million point movements to improve the preservation of geometry on the map. In principle, the distortion of such computed maps can be made as small as possible given the intrinsic curvature of the surface being mapped. We estimate that our algorithm is thus about one to two orders of magnitude faster, and perhaps two to three orders of magnitude more accurate, than the manual technique of Van Essen and Maunsell.

The "brainprints" of Jouandet *et al.* (1988) have the distinction of providing the first topological mapping of entire cerebral hemispheres in humans. Like the techniques of Van Essen and Maunsell (1980), they rely in part on the preservation of length along contours of section to help obtain some preservation of geometry. However, their contours are restricted to remain one–dimensional, so that they cannot undergo any change in conformation as permitted by our technique. Thus, the brainprints are thus best regarded as preserving topology, with the preservation of geometry being limited, variable, and imprecisely defined in terms of the various kinds of mappings. Nevertheless, brainprints are likely to provide a useful mapping for qualitative or very approximate quantitative comparisons of the locations and extent of large lesions in human patients. However, the detailed mapping of small lesion data for the purpose of inferring the functional organization of the cerebral cortex in man will require more precise quantitation than the techniques of Jouandet *et al.* can provide.

Schwartz and Merker (1985, 1986) used an algorithm very similar to that of Sammon (1969) to produce mappings of the smooth, exposed opercular portion of macaque striate cortex. Like Sammon, they use a local gradient descent method in order to minimize the error between distances on the map and a set of global distances measured from the surface. This method can fail to obtain minimum distortion maps if it becomes trapped in relative minima of the function they seek to minimize. While it has not been possible to compare our methods directly, the susceptability of their algorithm to such relative minima could be seen in a film of the progressive unfolding of their opercular map shown during one presentation of their work (Schwartz and Merker, 1985). One could clearly observe several folds develop in their map, whose removal required the addition of additional nodes well outside the boundaries of their map. Given the relative smoothness of the operculum, it is likely that their algorithm will require supervision and intervention of this kind to correct such foldings as they arise in the course of mapping regions of the cortex which are larger or have more intrinsic curvature. Furthermore, the mapping of Schwartz and Merker only attempts to preserve some subset of lengths measured on a triangular tesselation, which limits them to reconstruction of isometric approximations which are unsuitable to regions of cortex which have intrinsic curvature. In contrast, our algorithms attempts to preserve both lengths and angles measured on a quadrilateral tesselation or on surface patches fit to such a tesselation. This offers us the ability to independently control the weight given to the preservation of lengths and angles, so that either isometric or conformal mappings may be obtained as is appropriate, and to obtain better estimates of surface geometry. Finally, we have the capability of producing other maps, such as an area–preserving map, with only a minor change in code.

Our technique is not without limitations, however. The guarantee of attaining the near theoretical minimum distortion provided by the use of simulated annealing also means that the technique is computationally expensive. It is clear, however, that annealing is necessary for at least some maps if they are reconfigured solely within the plane. We demonstrated this in the course of early efforts to find a suitable optimization method by setting the temperature to zero during the annealing, so as to produce a stochastic optimization which always attempts to decrease the distortion, instead of allowing some probability of increasing or decreasing distortion. The maps which resulted often became caught in bizarrely folded configurations symptomatic of relative minima, comparable to the foldings seen in the map of Schwartz and Merker (1985). When the surfaces were mapped using annealing, however, no such folds appeared. We believe the occurrence of folding to be due to the inability of the map to find a configuration wherein measures of surface geometry are all mutually satisfied. This *frustration* is to be expected whenever intrinsically curved regions of surfaces are mapped onto the plane. We note that such frustration is bound to occur regardless of how the map is reconfigured, provided that the objective of the mapping is an unfolded, two-dimensional representation of such surfaces. We are developing alternative approaches to address these issues while at the same time improving the speed with which the maps can be computed. Fortunately, the computational demands of our present techniques are quite reasonable, and are by no means a limiting factor in such work. In practice, it is a small cost to pay for the assurance that the maps will attain near-global minima of distortion.

In practice, most distortions which occur in the final maps can be traced to errors in the input data. This calls attention to the need for the development of calibration and rectification proceedures to correct distortions arising from histological processing, and improved sampling techniques involving video digitization of histological data. Further developments in nondestructive and noninvasive sampling by confocal microscopy, magnetic resonance imaging, and positron-emission tomography will all assist in achieving such objectives.

Our current use of sampling of single contours of section can be generalized to accomodate sparse volumetric data, such as a distribution of labelled neurons not restricted to a single lamina or diffuse patterns of activation which may extend across several laminae. In such cases, either the data of interest can be projected onto a single surface which is then mapped, or the position of data can be recorded with respect to two surfaces (the boundaries of the gray matter), both of which are then mapped together. In this manner, the entire cortical mantle can be mapped as a "thick sheet" to obtain the laminar pattern of extended or distributed data, which can then be "tangentially sectioned" to reveal how it varies with depth.

Perhaps more significant than the development of these quantitative techniques for the mapping of the cortex has been the elucidation of a model, based on the complex harmonic potentials, for producing stable flows of information which correspond to solutions of the topological problem and other isomorphic problems of vision. As a final flourish to these results, we note that plotting the spatial component of the complex potentials involved in generating these flows reproduces many aspects of the topography of the primate visual pathway (Figure 51) (Carman, 1990). The qualitative similarity of these plots to the known typical shape and topographic organization of these stages of the visual pathway is remarkable in many regards (Malpeli and Baker, 1975; Connolly and Van Essen, 1984; Van Essen *et al.*, 1984; Tootell *et al.*, 1988). The first component of this mapping transforms the visual hemifield into the representation seen in the lateral geniculate nucleus. This is nothing more than a conformal mapping to permit combination of input from the two eyes, and to transform the hemifield into a disc, as the first step in construction of a Green's function. The second component of this mapping transforms the retinotopic organization of the lateral geniculate nucleus into that of V1, the primary visual cortex. This is a conformal mapping which completes the

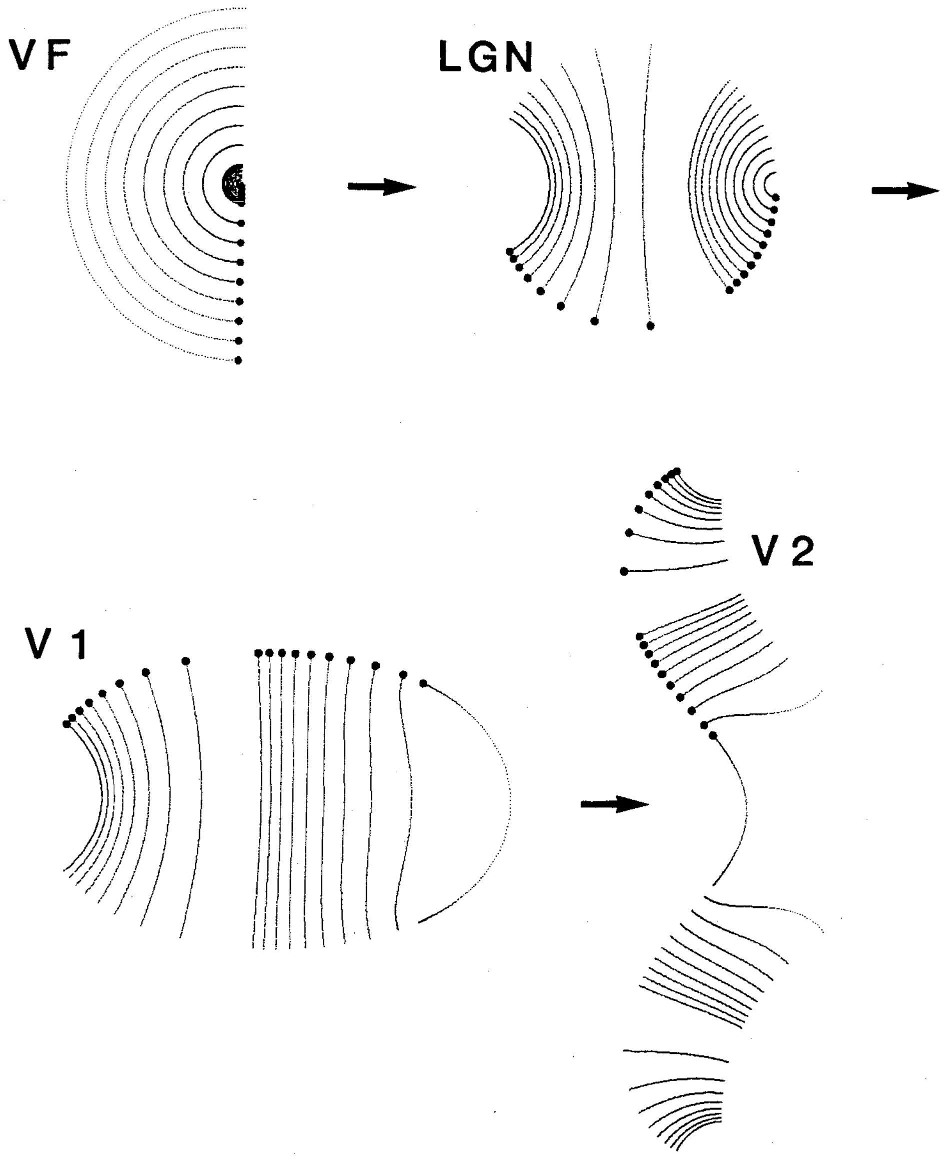


Figure 51. Plots of the spatial components of the complex potentials for computing flows for affine invariance. Isoeccentricities are plotted every 1 degree for the most central 10 degrees, and every 10 degrees thereafter. VF: visual hemifield; LGN: lateral geniculate nucleus; V1: cortical area V1; V2: cortical area V2.

construction of the Green's function used in computing the potential governing translations across spatial scales through through solution of the Dirichlet problem (Kantorovich and Krylov, 1964). The third component of this mapping transforms the topography of V1 to that of V2, the second visual area. This is also a conformal mapping, which repeats the construction of a Green's function so as to compute the potential governing the rotations and scaling across spatial scales. The combination of these two potentials permit the determination of a continuous scale similarity transform for fusion of image pairs in stereopsis or kineopsis. Thus, the shape and organization of the visual pathway may be interpreted as a precomputation of the spatial components of the complex harmonic potentials subserving the generation of flows of information useful for visual processing. The mappings of the cerebral cortex can thus be regarded as embeddings of such flows in Euclidean space, whose existence is guaranteed by the embedding theorem (Whitney, 1936), and whose number is as large as the set of all analytic functions, providing a large repetoire of possible computations from which evolution could select when "designing" sensory pathways in this manner. If this is a general feature of cortical organization, it may be possible to determine the functions performed by other cortical areas by quantitative mapping of their topography and determination of the analytic functions which produce them. If this proves to be the case, then eventually mathematical theory will supplant phenomenology as the foundation for the empirical study of mappings and functional organization of the cerebral cortex.

5. Bibliography

Abraham, R.H. and Shaw, C.D. *Dynamics -- The Geometry of Behavior* Parts I, II, III, and IV. (Santa Cruz, CA: Aerial Press, 1982).

Acton, F. S. *Numerical Methods That Work*. (New York: Harper and Row, 1970).

Adelson, E.H., Anderson, C.H., Bergen, J.R., Burt, P.J., and Ogden, J.M. Pyramid Methods in Image Processing. *RCA Engineer* **29**, 1984.

Ahuja, N. Dot Pattern Processing Using Voronoi Neighborhoods. *IEEE Trans. Patt. Anal. Mach. Intell.* **PAMI−4**: 336−343, 1982.

Alexandroff, P. *Elementary Concepts of Topology* (New York: Dover, 1961).

Algazi, V.R., B.W. Reutter, W.L.G. van Warmerdam, and C.C. Liu. Three−Dimensional Image Analysis and Display by Space−Scale Matching of Cross Sections. *J. Opt. Soc. Amer. A.* **6**: 890−899, 1989.

Allman, J.M. Variations in Visual Cortex Organization in Primates. *Neurobiology of Neocortex*, Rakic, P. and Singer, W., eds. (New York: John Wiley & Sons, Ltd., 1988), pp. 29−40.

Allman, J.M. and E.R. McGuiness. In: *Comparative Primate Biology, Volume 4: The Neurosciences*, Steklis, H.J. and Erwin, J., eds. (New York: Alan R. Liss, 1988), pp. 279−326.

Barnsley, M.F. and Elton, J.H. A New Class of Markov Processes for Image Encoding. *Adv. Appl. Prob.* **20**: 14−32, 1988.

Barnsley, M.F. and Hardin, D.P. A Mandelbrot Set Whose Boundary is Piecewise Smooth. *Trans. Amer. Math. Soc.* **315**: 641−659, 1989.

Barnsley, M.F., Elton, J.H., and Hardin, D.P. Recurrent Iterated Function

Systems. *Constr. Approx.* **5**: 3–31, 1989.

Barr, A. Superquadrics and Angle–Preserving Transformations. *IEEE Computer Graphics and Applications*, pp. 11–23, January, 1981.

Barr, A. Global and Local Deformations of Solid Primatives. *Computer Graphics* **18**: 21–30, 1984.

Barsky, B.A. A Description and Evaluation of Various 3D Models. In: Kunii, T.L. (ed.), *Computer Graphics -- Theory and Applications* (New York: Springer–Verlag, 1983).

Bassett, K. Numerical Methods for Map Analysis. *Progress in Geography* **4**: 218 – 254, 1972.

Batnitzky, S., Price, H.I., Cook, P.N., Cook, L.T., and Dwyer, S.J. III. Three–Dimensional Computer Reconstructions from Surface Contours for Head CT Examinations. *J. Computer Assisted Tomography* **5**: 60–67, 1981.

Berge, C. *Graphs*, Second Edition. (Amsterdam: North–Holland, 1985).

Beusmans, J.M.H., Hoffman, D.D., and Bennett, B.M. Description of Solid Shape from Occluding Contour. *J. Opt. Soc. Amer. A*, **4**: 1155–1167, 1987.

Binford, T.O. Visual Perception by Computer. *Proceedings IEEE Conference on Systems and Control*, Miami, Florida, December 1971.

Boissonnat, J.D. Shape Reconstruction from Planar Cross Sections. *Computer Vision, Graphics, and Image Processing* **44**: 1–29, 1988.

Bounds, D.G. New Optimization Methods from Physics and Biology. *Nature* **329**: 215–219, 1987.

Bradley, W.G., Newton, T.H., and Crooks, L.E. In: *Modern Neuroradiology, Volume*

2: Advanced Imaging Techniques, Newton, T.H. and Potts, D.G., eds. (New York: Claradel Press, 1983).

Brady, M., Ponce, J., Yuille, A., and Asada, H. Describing Surfaces. *Comp. Vis. Graph. Image Proc.* **32**: 1–28, 1985.

Bresler, Y., Fessler, J.A., Macovski, A. A Bayesian Approach to Reconstruction from Incomplete Projections of a Multiple Object 3D Domain. *IEEE Trans. Patt. Anal. Mach. Intell.* **11**: 840–858, 1989.

Brevdo, L., Sideman, S., Beyar, R. A Simple Approach to the Problem of 3D Reconstruction. *Computer Vision, Graphics, and Image Processing* **37**: 420–427, 1987.

Brownell, G.L., Budinger, T.P., Lauterbur, P.C., and McGeer, P.L. Positron Emission Tomography and Nuclear Magnetic Resonance Imaging. *Science* **215**(4533): 619–626, 1982.

Carman, G.J. Reduction of Visual Processes to A Common Algorithm. *Soc. Neurosci. Abstr.* **13**: 630, 1987.

Carman, G.J. Conformal Mapping of the Visual Pathway. *Soc. Neurosci. Abstr.* **16**: in press, 1990.

Carman, G.J. and Van Essen, D.C. Computational Algorithms for the Production of Two-Dimensional Maps of the Cerebral Cortex. *Soc. Neurosci. Abstr.* **11**: 1243, 1985.

Carman, G.J. and Mora, B.N. Noninvasive Computational Cartography of Human Visual Cortex Based on Magnetic Resonance Imaging and Positron Emission Tomography. *Soc. Neurosci. Abstr.* **15**: 1106, 1989.

Christiansen, H.N. and Sederberg, T.W. *Computer Graphics* **12**: 187–192, 1978.

Connolly, M. and Van Essen, D.C. The Representation of the Visual Field in Parvicellular and Magnocellular Layers of the Lateral Geniculate Nucleus of the Macaque Monkey. *J. Comp. Neurol.* **226**: 544–564, 1984.

Cook, P.N. A Study of Three-Dimensional Reconstruction Algorithms. *Automedica* **4**: 3–12, 1981.

Coxeter, H.S.M. *Introduction to Geometry* (New York: John Wiley, 1961).

Crutchfield, J.P. Space-time Dyanmics in Videofeedback. *Physica D* **10**: 229–245, 1984.

Crutchfield, J.P., Farmer, J.D., Packard, N.H., and Shaw, R.S. Chaos. *Scientific American* **255**: 46–57, 1986.

Curtis, S.R. and Oppenheim, A.V. Reconstruction of Multidimensional Signals from Zero Crossings. *J. Opt. Soc. Amer.* **4**: 221–231, 1987.

Curtiss, D.R. *Analytic Functions of a Complex Variable* (La Salle, Illinois: Open Court Publishing Co., 1948).

DeFloriani, L. A Pyramid Data Structure for Triangle-based Surface Description. **IEEE Computer Graphics & Applications**, March, pp. 67–78, 1989.

DeFloriani, L., Falcidieno, B., Nagy, G., Pienovi, C. A Hierarchical Structure for Surface Approximation. *Comput. & Graphics* **8**: 183–193, 1984.

Delaunay, B. *Bull. Acad. Sci. USSR (VII) Cl. Sci. Mat.* 793, 1906.

Dierckx, P., Suetens, P., Vandermeulen, D. *J. Comp. Appl. Math.* **23**: 367–388, 1988.

Dorst, L. and Smeulders, A.W.M. Length Estimators for Digitized Contours.

Computer Vision, Graphics, and Image Processing **40**: 311–333, 1987.

Dunham, J.G. Optimum Uniform Piecewise Linear Approximation of Planar Curves. *IEEE Trans. Patt. Anal. Mach. Intell.* PAMI–8: 67–75, 1986.

Flanders, H. *Differential Forms with Applications to the Physical Sciences.* (Toronto: Dover, 1989).

Foley, J.D. and Van Dam, A. *Fundamentals of Interactive Computer Graphics* (Reading, Massachusetts: Addison–Wesley, 1982).

Forbes, D.J. and Petry, R.W. Computer Assisted Mapping with the Light Microscope. *J. Neurosci. Methods* **1**: 77–94, 1979.

Fox, P.T., Mintun, M.A., Raichle, M.E., Miezin, F.M., Allman, J.M., and Van Essen, D.C. Mapping Human Visual Cortex with Positron Emission Tomography. *Nature* **323**(6091): 806–809, 1986.

Fox, P.T. and Raichle, M.E. Stimulus Rate Dependence of Regional Cerebral Blood Flow in Human Striate Cortex, Demonstrated by Positron Emission Tomography. *J. Neurophysiol.* **51**(5): 1109–1120, 1984.

Fox, P.T. and Raichle, M.E. Focal Physiological Uncoupling of Cerebral Blood Flow and Oxidative Metabolism During Somatosensory Stimulation in Human Subjects. *Proc. Natl. Acad. Sci. USA* **83**:1140–1144, 1986.

Fuchs, B.A. and Shabat, B.V. *Functions of a Complex Variable, Volume I* (Oxford: Pergamon Press, 1964).

Fuchs, H., Kedem, Z.M., Uselton, S.P. Optimal Surface Reconstruction from Planar Contours. *Communications of the ACM* **20**(10): 693–702, 1977.

Funnell, W.R.J. On the Choice of a Cost Function for the Reconstruction of Surfaces by Triangulation Between Sections. *Computers and Structures* **18**:

23-26, 1984.

Gabor, D. Theory of Communication. *J. Inst. Electr. Eng.*, Part 3, **93**: 429-457, 1946.

Ganapathy, S. and Dennehy, T.G. A New General Triangulation Method for Planar Contours. *Computer Graphics* **16**: 69-75, 1982.

Gardner, M. On Map Projections (with Special Reference to Some Inspired Ones). *Scientific American*, November, 1975, pp. 120-125.

Gauss, K.F. *Theorema egregium* in: *Disquisitiones Circa Superficies Curvas*, 1827.

Gattass, R., Gross, C.G., and Sandell, J.H. Visual Topography of V2 in the Macaque. *J. Comp. Neurol.* **201**: 519-539, 1981.

Geman, S. and Geman, D. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Trans. Pat. Anal. Mach. Intell.* PAMI-**6**(6): 721-741, 1984.

Gilbert, E.N. Distortion in Maps. *SIAM Review* **16**(1): 47-62, 1974.

Glazer, F. Multileval Relaxation in Low-Level Computer Vision. In: Rosenfeld, A., ed., *Multiresolution Image Processing and Analysis* (New York: Springer-Verlag, 1984).

Grimson, W.E.L. A Computational Theory of Visual Surface Interpolation. *Phil. Trans. R. Soc. London, Series B* **298**: 395-427, 1982.

Grunbaum, B. and Shephard, G.C. *Tilings and Patterns: An Introduction.* San Francisco: W.H. Freeman and Company, 1989).

Harary, F. *Graph Theory* (Reading, Massachusetts: Addison-Wesley, 1969).

Herbin, M., Venot, A., Devaux, J.Y., Walter, E., Lebruchec, J.F., Dubertret, L., and Roucayrol, J.C. Automated Registration of Disimilar Images: Application to Medical Imagery. *Comp. Vis. Graph. Image Proc.* **47**: 77–88, 1989.

Hilbert, D. and Cohn–Vossen, S. *Geometry and the Imagination*, translated by P. Nemenyi (New York: Chelsea Publishing Co., 1952).

Hildebrandt, S. and Tromba, A. *Mathematics and Optimal Form* (New York: Scientific American Library, 1984).

Hildreth, E.C. *The Measurement of Visual Motion. (Cambridge: MIT Press, 1984).*

Hoffman, D.D. and Richards, W.A. Parts of Recognition. Cognition **18**: 65–96, 1984.

Hogg, T. and Huberman, B.A. Attractors on Finite Sets: The Dissipative Dynamics of Computing Structures. *Phys. Rev. A* **32**: 2338–2346, 1985.

Horaud, R. and Brady, M. On the Geometric Interpretation of Image Contours. *Artificial Intelligence* **37**: 333–353, 1988.

Horn, B.K.P. *Robot Vision.* (Cambridge, Mass.: MIT Press, 1986).

Imig, T.J. and Brugge, J.F. Sources and Terminations of Callosal Axons Related to Binaural and Frequency Maps in Primary Auditory Cortex of the Cat. *J. Comp. Neurol.* **182**: 637–660, 1978.

Jouandet, M.L., Tramo, M.J., Herron, D.M., Hermann, A., Loftus, W.C., Bazell, J., and Gazzaniga, M.S. Brainprints: Computer–Generated Two–Dimensional Maps of the Human Cerebral Cortex in Vivo. *J. Cog. Neurosci.* **1**(1): 88–117, 1988.

Julesz, B. *The Foundations of Cyclopean Perception* (Chicago: University of Chicago Press, 1971).

Kantorovich, I.V. and Krylov, V.I. *Approximate Methods of Higher Analysis.* (New York: John Wiley & Sons, 1964).

Kehtarnavaz, N. and de Figueiredo, R.J.P. A Framework for Surface Reconstruction from 3D Contours. *Comp. Vis. Graph. Im. Proc.* **42**: 32–47, 1988.

Kehtarnavaz, N., Simar, L.R., and de Figueiredo, R.J.P. A Syntactic/Semantic Technique for Surface Reconstruction from Cross-Sectional Contours. *Comp. Vis. Graph. Im. Proc.* **42**: 399–409, 1988.

Keppel, E. Approximating Complex Surfaces by Triangulation of Contour Lines. *IBM J. Res. Develop.* **2**: 2–11, 1975.

Kirkpatrick, S. Frustration and ground-state degeneracy in spin glasses. *Phys. Rev. B.* **16**(10): 4630–4641, 1977.

Kirkpatrick, S., Gelatt, C.D. Jr., and Vecchi, M.P. Optimization by Simulated Annealing. *Science* **220**(4598): 671–680, 1983.

Knudsen, E.I., du Lac, S., and Esterly, S. Computational Maps in the Brain. *Ann. Rev. Neurosci.*, **10**: 41–65, 1987.

Koch, C., Poggio, T. and Yuille, A. Analog Neuronal Networks in Early Vision. *PNAS* **83**: 4263–4267, 1986.

Koenderink, J.J. The Structure of Images. *Biol. Cybern.* **50**: 363–370, 1984.

Koenderink, J.J. The Structure of the Visual Field. Preprint, November 1986.

Koenderink, J.J. An Internal Representation for Solid Shape Based on the Topological Properties of the Apparent Contour. In: Richards, W. and Ullman, S., eds., *Image Understanding 1985–86* (Norwood, NJ: Ablex, 1987).

Koenderink, J.J. and Van Doorn, A.J. The Representation of Space–Time in the Human Visual System. Preprint, May 1987.

Lancaster, P. and Salkauskas, K. *Curve and Surface Fitting*. (London: Academic Press, 1986).

Landau, H.J. Sampling, Data Transmission, and the Nyquist Rate. *Proc. IEEE* **55**: 1701–1706, 1967.

Lang, S. *Linear Algebra* (Reading, MA: Addison–Wesley, 1970).

Lassen, N.A., Ingvar, D.H., and Skinhoj, E. Brain Function and Blood Flow. *Scientific American*, October 1978, pp. 62 – 71.

Lee, D.T. and Schachter, B.J. Two Algorithms for Constructing a Delaunay Triangulation. *Int. J. Comp. Inf. Sci.* **9**: 219–242, 1980.

Leith, A., Marko, M., Parsons, D. Computer Graphics for Cellular Reconstruction. *IEEE Computer Graphics and Applications*, September, 1989, pp. 16–23.

Le Vay, S., Connolly, M., Houde, J., and Van Essen, D.C. The Complete Pattern of Ocular Dominance Stripes in the Striate Cortex and Visual Field of the Macaque Monkey. *J. Neurosci.* **5**: 486–501, 1985.

Levinthal, C. and Ware, R. Three–Dimensional Reconstruction from Serial Sections. *Nature* **236**: 207–210, 1972.

Lin, W.–C., Chen, S.–Y., and Chen, C.–T. A New Surface Interpolation Technique for Reconstructing 3D Objects from Serial Cross–Sections. *Comp. Vis. Graph. Im. Proc.* **48**: 124–143, 1989.

Livadas, P.E. A Reconstruction of an Unknown 3–D Surface From a Collection of Its Cross Sections. *Intern. J. Computer Math.* **26**: 143–160, 1989.

Logan, B.F. Jr. Information in the Zero–crossings of Bandpass Signals. *Bell Syst. Tech. J.* **56**: 487–510, 1977.

Lorensen, W.E. and Cline, H.E. Marching Cubes: A High Resolution 3D Surface Reconstruction Algorithm. *Computer Graphics* **21**: 163–169, 1987.

Lund, J.S. Mapping Strategies of Monkey Primary Visual Cortex. In: *Sensory Processing in the Mammalian Brain*, J.S. Lund, ed. (New York: Oxford University Press, 1989), pp. 209–225.

Malpeli, J.G. and Baker, F.H. The Representation fo the Visual Field in the Lateral Geniculate Nucleus of *Macaca mulatta*. *J. Comp. Neurol.* **161**:569–594, 1975.

Mandelbrot, B.B. *The Fractal Geometry of Nature* (San Francisco: W.H. Freeman and Co., 1982).

Marko, M., Leith, A., and Parsons, D. Three–Dimensional Reconstructions of Cells and Whole–Cell Mounts Using Multilevel Contouring of Stereo Micrographs. *J. Electron Microscopy Technique* **9**: 395–411, 1988.

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

Marr, D. and Nishihara, H.K. Representation and Recognition of the Spatial Organization of Three–Dimensional Shapes. *Proc. R. Soc. Lond. B.* **200**: 269–294, 1978.

Marr, D. and Poggio, T. Cooperative Computation of Stereo Disparity. *Science* **194**: 283–287, 1976.

Marr, D., Ullman, S., and Poggio, T. Bandpass Channels, Zero–crossings, and Early Visual Information Processing. *J. Opt. Soc. Am.* **69**: 914–916, 1979.

Marvasti, F.A. Reconstruction of Two–Dimensional Signals from Nonuniform Samples or Partial Information. *J. Opt. Soc. Amer. A.* **6**(1): 52–55, 1989.

Merickel, M. 3D Reconstruction: The Registration Problem. *Comp. Vis. Graph. Image Proc.* **42**: 206–219, 1988.

Metropolis, N., Rosenbluth, A., Teller, A., and Teller, E. Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.* **21**: 1087–1092, 1953.

Milios, E.E. Shape Matching Using Curvature Processes. *Comp. Vis. Graph. Im. Proc.* **47**:203–226, 1989.

Mitchell, J.S.B., Mount, D.M., and Papadimitriou, C.H. The Discrete Geodesic Problem. *SIAM J. Comput.* **16**: 647–668, 1987.

Mora, B., Carman, G.J., and Allman, J. *In vivo* Functional Localization of the Human Visual Cortex Using Positron Emission Tomography and Magnetic Resonance Imaging. *Trends in Neurosci.* **12**(8): 282–284, 1989.

Morganthaler, D.G. and Rosenfeld, A. Surfaces in Three–Dimensional Digital Images. *Information and Control* **51**: 227–247, 1981.

Munkres, J.R. *Elements of Algebraic Topology* (Menlo Park, CA: Benjamin/Cummings, 1984).

Norman, D. and Brant–Zawadzki, M. Magnetic Resonance Imaging of the Central Nervous System. In: *Brain Imaging and Brain Function*, Sokoloff, L., ed. (New York: Raven Press, 1985).

Olavarria, J. and Van Sluyters, R.C. Unfolding and Flattening the Cortex of Gyrencephalic Brains. *J. Neurosci. Meth.* **15**: 191–202, 1985.

Oldendorf, W.H. Principles of Imaging Structure by NMR. In: *Brain Imaging and Brain Function*, Sokoloff, L., ed. (New York: Raven Press, 1985).

O'Rourke, J. Polyhedra of Minimal Area as 3D Object Models. *Proc. Int. Joint Conf. Artif. Intell.*, pp. 664–666, 1981.

Paulson, O.B. and Newman, E.A. *Science* **237**: 896–898, 1987.

Pavlidis, T. Segmentation of Plane Curves. *IEEE Trans. Comp.* **C–23**: 860–870, 1974.

Pavlidis, T. A Review of Algorithms for Shape Analysis. *Computer Graphics and Image Processing* **7**: 243–258, 1978.

Phelps, M.E. and Mazziotta, J.C. Positron Emission Tomography: Human Brain Function and Biochemistry. *Science*, **228**: 799–809, 1985.

Poggio, T., Torre, V., Koch, C. Computational Vision and Regularization Theory. *Nature* **317**:314–319, 1985.

Press, W.H., Flannery, B.P., Teukolsky, S.A., and Vetterling, W.T. *Numerical Recipes in C* (Cambridge: Cambridge University Press, 1988).

Raichle, M.E. Neuroimaging. *Trends in Neurosciences*, **9**: 525–529, 1986.

Raichle, M.E., Grubb, R.L., Mokhtar, H.G., Eichling, J.O., and TerPogossian, M.M. *Arch. Neurol.* **33**: 523–526, 1976.

Reale, R.A. and Imig, T.J. Tonotopic Organization in Auditory Cortex of the Cat. *J. Comp. Neurol.* **192**: 265–291, 1980.

Requicha, A.A.G. Zeros of Entire Functions: Theory and Engineering Applications. *Proc. IEEE* **68**: 308–328, 1980.

Richardus, P. and Adler, R.K. *Map Projections for Geodesists, Cartographers, and Geographers* (Amsterdam: North–Holland, 1972).

Richman, D.P., Stewart, M.R., Hutchinson, J.W., and Caviness, V.S. Jr. Mechanical Model of Brain Convolutional Development. *Science,* **189**: 18–21, 1975.

Riedinger, R., Habar, M., Oelhafen, P., Guntherodt, H.J. About the Delaunay–

Voronoi Tesselation. *Journal of Computational Physics* **74**: 61–72, 1988.

Ritter, G.X. and Tou, J.T. The Encoding of Arbitrary Surfaces in 3–Dimensional Space. *Pattern Recognition* **17**: 593–606, 1984.

Roberts, M.P., Hanaway, J., and Morest, D.K. *Atlas of the Human Brain in Section* (Philadelphia: Lea and Febiger, 1987).

Rogers, C.A. *Packing and Covering.* (Cambridge University Press: London, 1964).

Samet, H. The Quadtree and Related Hierarchical Data Structures. *Computing Surveys* **16**: 187–260, 1984.

Sammon, J.W., Jr. A Nonlinear Mapping for Data Structure Analysis. *IEEE Transactions on Computers* **C–18**: 401–409, 1969.

Sander, P.T. and Zucker, S.W. Inferring Differential Structure from 3–D Images: Smooth Cross Sections of Fibre Bundles. *McGill Res. Cent. Intell. Machines Tech. Rep. TR–CIM–88–6* (Montreal, Canada: McGill University, 1988).

Schaltenbrand, G. and Bailey, P. *Introduction to Stereotaxis with an Atlas of the Human Brain.* (Thieme, 1959).

Schwartz, E.L. and Merker, B. Flattening Cortex: An Optimal Computer Algorithm and Comparisons with Physical Flattening of the Opercular Surface of Striate Cortex. *Soc. Neurosci. Abstr.* **11**(2): 1243, 1985. Presented at the 15th annual meeting of the Society for Neurosciences, Dallas, TX, October 1985.

Schwartz, E.L. and Merker, B. Computer–aided Neuroanatomy: Differential Geometry of Cortical Surfaces and an Optimal Flatening Algorithm. *Computer Graphics and Applications*, March 1986, pp. 36–44.

Sereno, M.I. and Allman, J.M. Cortical Visual Areas in Mammals. Unpublished manuscript, 1989.

Shannon, C.E. A Mathematical Theory of Communication. *Bell System Technical Journal* **27**: 379–423, 1948.

Shannon, C.E. Communication in the Presence of Noise. *Proc. Inst. Rad. Eng.* **37**: 11–21, 1949.

Sloan, S.W. and Houlsby, G.T. An Implementation of Watson's Algorithm for Computing 2–Dimensional Delaunay Triangulations. *Adv. Eng. Software* **6**: 192–197, 1984.

Sokolnikoff, I.S. *Mathematical Theory of Elasticity* (New York: McGraw–Hill, 1946).

Sokolnikoff, I.S. and Redheffer, R.M. *Mathematics of Physics and Modern Engineering*, second edition. (New York: McGraw–Hill, 1958).

Sokoloff, L., ed. *Brain Imaging and Brain Function* (New York: Raven Press, 1985).

Solina, F. and Bajcsy, R. Recovery of Parametric Models from Range Images: The Case for Superquadrics with Global Deformations. *IEEE Trans. Patt. Anal. Mach. Intell.* **12**: 131–147, 1990.

Srihari, S.N. Representation of Three–Dimensional Images. *Computing Surveys* **13**: 399–424, 1981.

Stensass, S.S., Eddington, D.K., and Dobelle, W.H. The topography and variability of the primary visual cortex in man. *J. Neurosurg.* **40**: 747–755, 1974.

Stewart, I. *Does God Play Dice? The Mathematics of Chaos.* (Oxford: Basil Blackwell, Ltd., 1989).

Talairach, J. and Szikla, G. *Atlas d'Anatomie Stereotaxique du Telencephale: Etudes Anatomo–Radiologiques.* (Paris: Masson and Co., 1967).

Talbot, S. A. and Marshall, W.H. Physiological Studies on Neural Mechanisms of Visual Localization and Discrimination. *Am. J. Ophthal.* **24**: 1255–1264, 1941.

Talbot, S. A. and Marshall, W.H. Physiologic Studies on Neural Mechanisms of Visual Localization and Discrimination. *Archs. Ophthal.* **27**: 213–215, 1942.

Terzopoulos, D. Multilevel Computational Processes for Visual Surface Reconstruction. *Comp. Vision Graphics Image Proc.* **24**: 52–96, 1983.

Terzopoulos, D. On Matching Deformable Models to Images. *Topical Meeting on Machine Vision*, Technical Digest Series, volume 12 (Washington, D.C.: Optical Society of America, 1987), pp. 160–167.

Terzopoulos, D. and Fleischer, K. Modelling Inelastic Deformation: Viscoelasticity, Plasticity, and Fracture. *Computer Graphics,* **22**: 269–278, 1988.

Terzopoulos, D., Platt, J., Barr, A., and Fleischer, K. Elastically Deformable Models. *Computer Graphics* **21**: 205–214, 1987.

Terzopoulos, D., Witkin, A., and Kass, M. Constraints on Deformable Models: Recovering 3D Shape and Nonrigid Motion. *Artificial Intelligence* **36**: 91–123, 1988.

Todd, J.T. and Reichel, F.D. Ordinal Structure in the Visual Perception and Cognition of Smoothly Curved Surfaces. *Psychological Review*, **96**: 643–657, 1989.

Tootell, R.B.H. and Silverman, M.S. Two Methods for Flat–Mounting Cortical Tissue. *J. Neurosci. Methods* **15**: 177–190, 1985.

Tootell, R.B.H., Switkes, E., Silverman, M.S., and Hamilton, S.L. Functional Anatomy of Macaque Striate Cortex. II. Retinotopic Organization. *J. Neurosci.*

8: 1531–1568, 1988.

Tusa, R.J., Palmer, L.A., and Rosenquist, A.C. The Retinotopic Organization of Area 17 (Striate Cortex) in the Cat. *J. Comp. Neurol.* **177**: 213–236, 1978.

Ullman, S. *The Interpretation of Visual Motion* (Cambridge, Mass: MIT Press, 1979).

Van Essen, D.C. Functional Organization of Primate Visual Cortex. In: *Cerebral Cortex*, Volume 3, Poters, A., and Jones, E.G., eds. (New York: Plenum Publishing, 1985).

Van Essen, D.C. and Maunsell, J.H.R. Two-dimensional maps of the cerebral cortex. *J. Comp. Neurol.* **191**:255–281, 1980.

Van Essen, D.C., Newsome, W.T., and Maunsell, J.H.R. The Visual Field Representation in Striate Cortex of the Macaque Monkey: Asymmetries, Anisotropies, and Individual Variability. *Vision Res.* **24**: 429–448, 1984.

Van Essen, D.C. and Zeki, S.M. The Topographic Organization of Rhesus Monkey Prestriate Cortex. *J. Physiol.* **277**: 193–226, 1978.

van Oosterom, A. Triangulating the Human Torso. *The Computer Journal* **21**: 253–258, 1977.

Voronoi, G. *J. Reine Angew. Math.* **134**: 198, 1908.

Watson, D.F. Computing the N-Dimensional Delaunay Tesselation with Application to Voronoi Polytopes. *Computer Journal* **24**: 167–172, 1981.

Watson, D.F. and Philip, G.M. Systematic Triangulations. *Comp. Vis. Graph. Image Proc.* **26**: 217–223, 1984.

Weinstein, M. and Castleman, K.R. Reconstructing 3-D Specimens from 2-D

Section Images. *Proc. SPIE* **26**: 131–138, 1971.

Whitney, H. Differentiable Manifolds. *Annals of Mathematics* **37**(3): 645–680, 1936.

Whitney, H. On Singularities of Mappings of Euclidean Spaces. I. Mappings of the Plane into the Plane. *Annals of Mathematics* **62**: 374–410, 1955.

Wilford, J.N. *The Mapmakers* (Vintage Books, New York, 1982).

Witkin, A.P. Scale Space Filtering. In : *Proceedings of the Eighth International Joint Conference on Artificial Intelligence*, A. Bundy, ed. (Los Altos, CA: Morgan Kauffman, 1983), pp. 1019–1021.

Witkin, A.P., Terzopoulos, D., Kass, M. Signal Matching Through Scale Space. *Int. J. Comp. Vision*, **1**: 133–144, 1989.

Woolsey, C. N. Comparative Studies on Cortical Representation of Vision. *Vision Res. Supplement No. 3*, 365–382, 1971.

Woolsey, T.A. and Van der Loos, H. The Structural Organization of layer IV in the Somatosensory Region (SI) of Mouse Cerebral Cortex. *Brain Res.* **17**: 205–242.

Xu, S.B. and Lu, W.X. Surface Reconstruction of 3D Objects in Computerized Tomography. *Comp. Vis. Graph. Image Proc.* **44**: 270–278, 1988.

Yuille, A.L. and Poggio, T. A Generalized Ordering Constraint for Stereo Correspondence. *Massachusets Institute of Technology A.I. Memo 777*, 1984.

Appendix I

Approximation of Torsional Springs By Longitudinal Spring Pairs

The action of a torsional spring on a rigid beam can be approximated by the action of a pair of longitudinal springs attached to a fixed base (e.g., another beam) and which act on a beam which pivots about its endpoint. We choose the two springs to have resting lengths l_{io} such that neither spring is compressed or streched when the beam is at an angle Θ_0 with respect to the fixed base. This angle corresponds to the resting angle of the torsional spring.

Deflection of the beam from the angle Θ_0 by an angle Θ will stretch one spring and compress the other. The springs will be stretched or compressed by an amount:

$$r \sin(\Theta) = r\,\Theta$$

assuming the spring is confined to a circular cam or if Θ is small. The stored energy e of such a spring pair is:

$$\begin{aligned}
e &= \frac{1}{2}\,\kappa\,(l_1 - l_{10})^2 + \frac{1}{2}\,\kappa\,(l_2 - l_{20})^2 \\
&= \frac{1}{2}\,\kappa\,(-r\sin(\Theta - \Theta_0))^2 + \frac{1}{2}\,\kappa\,(r\sin(\Theta - \Theta_0))^2 \\
&= \kappa\,(r\sin(\Theta - \Theta_0))^2 \\
&= \kappa\,(r(\Theta - \Theta_0))^2 \\
&= \kappa\,r^2\,(\Theta - \Theta_0)^2 \\
&= \frac{1}{2}\,\lambda\,(\Theta - \Theta_0)^2
\end{aligned}$$

where $\lambda = 2\kappa\,r^2$. The last equation has the form of the energy of a torsion spring of resting angle Θ_0, demonstrating the desired equivalence. Note that this approximation is used only to determine values for coefficients κ and λ which are equivalent in terms of energy per unit displacment. Actual calculations of energies due to angular distortion involve no approximations.

Appendix II

Publications

Sleep, 7(4):347–355

Patterns in the Distribution of REM Sleep in Normal Human Sleep

*†George J. Carman, *‡Linda Mealey, *§Susan T. Thompson,
and *§Mark A. Thompson

**Cornell University, Ithaca, New York; †California Institute of Technology, Pasadena, California; ‡University of Texas, Austin, Texas; and §University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, U.S.A.*

Summary: Results are reported of four analyses of the distribution of REM sleep across nights of two subjects who slept for 50 consecutive nights on a regimented, but normal, sleep schedule. We found (a) a strong phase-setting effect of sleep onset on the distribution of REM sleep within nights, (b) no systematic change in the phase of the distribution of REM sleep across nights, and (c) a relationship between the nocturnal temperature minimum and the distribution of REM sleep within nights. Our results are consistent with the notion that the nightly distribution of REM sleep may be determined by an oscillatory process, the phase of which is reset at sleep onset, but which may be subject to other influences, such as the circadian temperature rhythm. These results are in general agreement with those found by investigators studying subjects on free-running or other abnormal sleep schedules. **Key Words:** Sleep—REM sleep—Ultradian rhythms—Circadian rhythms—Body temperature.

Although the existence of a nightly REM-NREM sleep cycle has been known for almost three decades (1), the variability of the REM-NREM pattern both between and within individuals across nights remains both unpredictable and unexplained. Numerous hypotheses have been put forth in an attempt to explain this REM-NREM patterning: Globus (2) and Othmer et al. (3) suggested that the REM-NREM cycle was a manifestation of a 24 h basic-rest-activity-cycle (BRAC) (4) and was linked to clock time; Webb and Agnew (5) hypothesized that it was the manifestation of a free-running BRAC, unentrained, and therefore shifting in real time; Schulz et al. (6) proposed that the first REM period (REMP) of the night was linked to sleep onset, but that later REMPs shifted as if linked to an unentrained BRAC; Moses et al. (7) and Johnson (8) claimed that the REM-NREM cycle was a sleep-dependent process, completely distinct from the BRAC; and McPartland and Kupfer (9) found evidence for both a sleep-independent (BRAC) component and a sleep-dependent component in the REM-NREM rhythm. More recently, a series of investigators have reported a relationship between the circadian rhythm in body temperature and sleep

---

Accepted for publication August 1984.

Address correspondence and reprint requests to Dr. Mealey at Department of Psychology, University of Texas, Austin, Texas 78712, U.S.A.

onset, sleep offset, sleep duration, REM latency, and REM duration in both free-running (10–12) and enforced, irregularly scheduled sleep conditions (13,14).

This article reports the results of analyses of the REM-NREM sleep distribution of two subjects who slept 50 consecutive nights under an enforced, but normal sleep schedule. It is hoped that these analyses may provide useful baseline information for those who are studying free-running, disrupted or otherwise abnormal sleep in human subjects.

METHODS

Protocol

Three male undergraduates (EA, JA, and TP, aged 18, 19, and 20, respectively) volunteered to participate in the study as part of an independent study project in psychology. Each slept in the laboratory for 50 consecutive nights, maintaining a constant schedule of lights out and awakening as selected by each subject at the outset of the study. The subjects were restricted from drugs, alcohol, and naps throughout the study, but otherwise maintained their normal routine. Each night electroencephalography, electrooculography, and electromyography were monitored using standard techniques on Grass model 5D (JA) and Model 7 (EA and TP) polygraphs. Two subjects (EA and TP) were also monitored for rectal temperature using isolated telethermometer probes (Yellow Springs Instruments Model 401) interfaced to a Harvard chart recorder (model 480 chart mover and model 350 recording modules) through a custom-designed Wheatstone bridge circuit. Temperature was sampled automatically at 30 s intervals.

Analyses were performed using data only from the last 42 nights of the study to ensure that adaptation to the lab and to the schedule was complete. All three subjects completed the study, but one (EA) developed a cold with apparent sequelae in his sleep and temperature data; hence results from this subject were excluded from the present report.

Scoring

Sleep episodes were coded and scored blindly into ½ min epochs, using the criteria of Rechtschaffen and Kales (15). The direct Fourier transform was computed from the minute-by-minute REM-NREM time series derived from each subject-night's data (D. F. Kripke, personal communication; 16). Using test periodicities of 70 to 140 min at 1 min increments, the period of maximal mean power across nights was determined for each subject (105 min for JA and 98 min for TP). For each night, we obtained the phase angle of the cosine having the period of maximal mean power for that subject, thus providing a stationary approximation for each subject-night's REM-NREM distribution. These phase angles were subsequently used in three analyses of the REM sleep distribution of each subject (see Fig. 1 for a pictorial description of this methodology).

Temperature recordings were scored for the occurrence of nightly minima using two methods: an absolute minimum was scored for each night, and a minimum was derived using a least-squares fit of a cosine of 24 h period (16). Our preliminary report (17) discussed results based on the use of the absolute temperature minima; the present report is based on results using the fitted temperature minima.

RESULTS

The REM-NREM sleep distribution of each subject is depicted in Fig. 2. Visual examination of these data suggests an entrainment of REMP 1 to sleep onset but is ineffective at elucidating a pattern among later REMPs. Our analyses were made in an attempt to

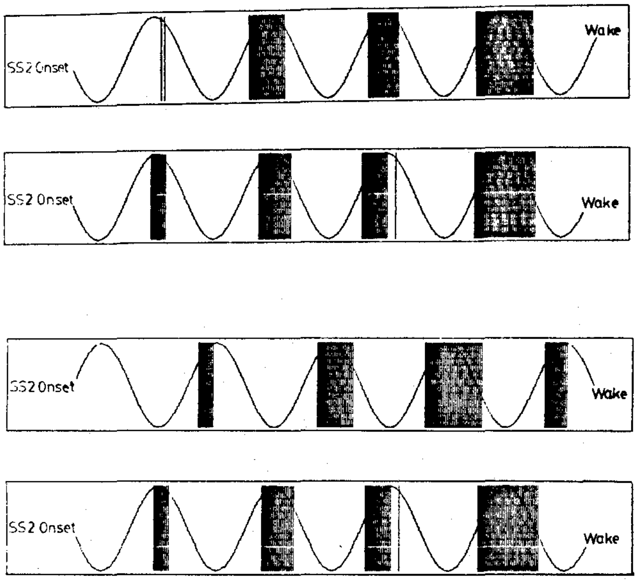


FIG. 1. Pictorial description of the quantification of the phase of each night's REM-NREM distribution. The shaded regions represent REM periods and the white regions NREM periods, which appear as a function of time. Each curve is a cosine, the period and phase of which are obtained by application of the direct Fourier transform as described in the text. **Above:** A pair of nights with similar REM-NREM distributions yield nearly identical phases with respect to the origin set at sleep stage 2 (SS2) onset. **Below:** A pair of nights with dissimilar REM-NREM distributions yield different phases. In this example the phases differ by 180°.

confirm this apparent relationship between sleep onset and REMP 1 and to further decipher any pattern in the later REMPs.

Our first analysis consisted of simple computation of the means, standard deviations, and correlations between timing of: lights out; sleep stage 1 and 2 onsets; REMP latencies; and the latency of the temperature minimum (for subject TP). Since timing of lights out, sleep stage 1 onset, and sleep stage 2 onset were so close and varied little in real time (Table 1), REMP latency correlations are reported only with respect to sleep stage 2 onset (Table 2), and all future reference to sleep onset refers to sleep stage 2 onset.

The descriptive statistics confirm our initial impression: as indicated by the standard deviations of REMP latencies, REMP 1 was significantly more strongly tied to the timing of sleep onset than were later REMPs. Additionally, the lack of significant negative correlations between sleep latency and REMP latencies indicates no tendency for REMPs to be tied to clock time.

TABLE 1. *Statistics of controlled sleep regimen*

	JA		TP	
	$\bar{x}$	SD	$\bar{x}$	SD
Lights out	12:33 a.m.	3.42	11:19 p.m.	4.04
SS 1 onset	12:42 a.m.	8.19	11:30 p.m.	6.01
SS 2 onset	12:49 a.m.	8.42	11:33 p.m.	6.55
T minimum	NA	NA	3:51 a.m.	49.95
Wake	7:30 a.m.	3.49	6.59 a.m.	1.25

Variation in lights out, sleep stage 1 (SS 1) onset, and sleep stage 2 (SS 2) onset was minimal under the controlled sleep regimen. Mean times are given as hours:minutes Eastern standard time. Standard deviations are given in minutes.

TABLE 2. *REM period (REMP) latencies and correlations*

	JA			TP		
	Latency	SD	r	Latency	SD	r
REMP 1	62.78	11.09[a]	−0.18	73.83	10.77[b]	0.03
REMP 2	170.50	25.90	−0.01	164.06	14.69	0.05
REMP 3	272.18	24.68	0.12	259.14	20.80	0.09
REMP 4	361.74	27.35	−0.03	355.26	29.62	0.17

REMP latencies are reported in minutes from sleep stage 2 (SS 2) onset. Standard deviations are reported in minutes. Correlations (r) between REMPs and SS2 onset are computed using latency of SS2 from midnight and latencies of REMPs from SS2 onset.

[a] Significantly smaller than that of REMPs 2, 3, and 4 at $p < 0.01$.

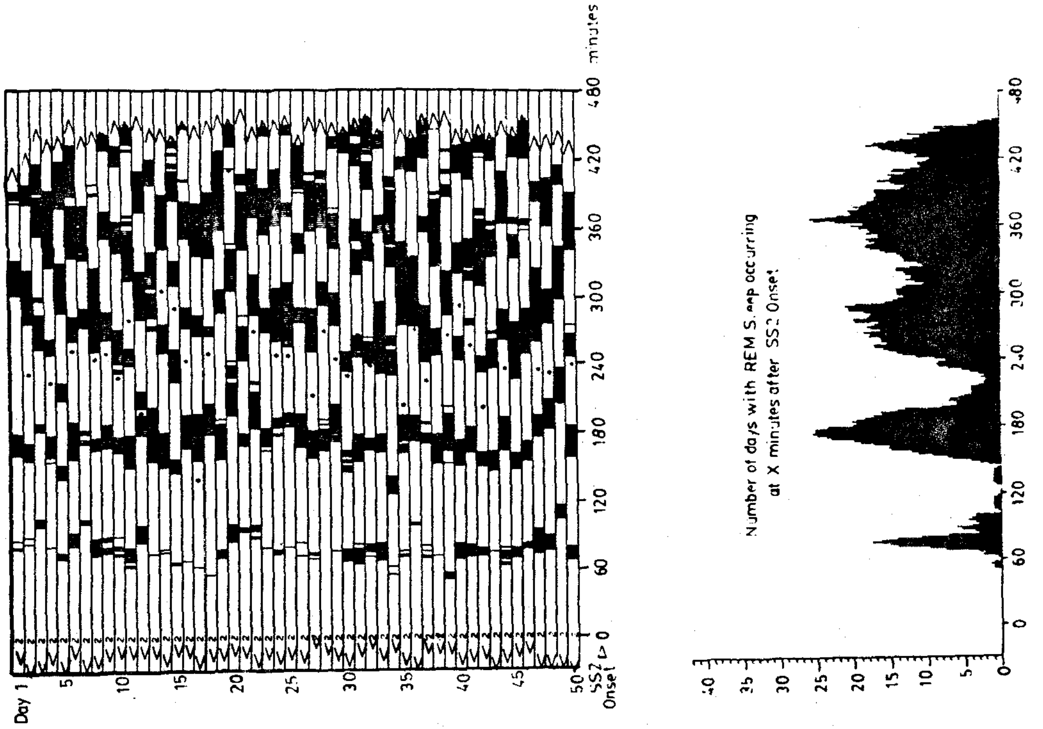
[b] Significantly smaller than that of REMP 3 at $p < 0.05$ and REMP 4 at $p < 0.01$.

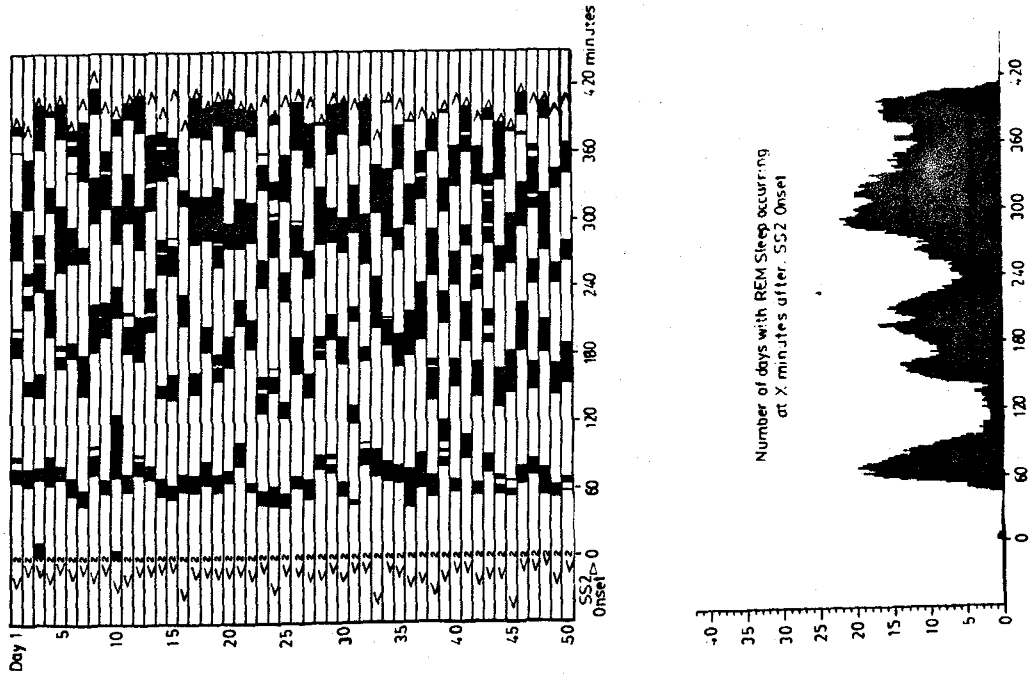
The second analysis examined the distribution across nights of the phase angles of the REM-NREM sleep distribution obtained for each night. For each subject, these phase angles were plotted on a circle with the origin (phase angle of zero) set at sleep onset (Fig. 3). For both subjects, the Rayleigh test (18) indicated a nonrandom distribution ($p < 0.05$ for JA; $p < 0.01$ for TP), corresponding to an average phase angle across nights. The average angle was similar for the two subjects (−90° for JA; −83° for TP). We also tested the distribution of phase angles obtained by weighting each vector according to the point-biserial correlation between each night's REM-NREM distribution and the cosine approximating that distribution.[1] Test results with the weighted vectors were even stronger than those using the nonweighted distributions ($p < 0.005$ for JA; $p < 0.001$ for TP) using the cosinor zero-amplitude test (19) (see also Fig. 3). These results further substantiate the entrainment effect of sleep onset on the distribution of REM sleep through the night.

---

[1] The length of the vector representing a given night's phase angle was set equal to the point-biserial correlation between that night's minute-by-minute values of sleep state (REM versus NREM) and the fitted cosine. Thus, if REM sleep occurred equally during the positive and negative phases of the cosine, a zero correlation (and a zero-length vector) would be obtained; perfect alignment of REMPs with the positive, and NREMPs with the negative phase of the curve, would yield a correlation of approximately 0.90. Since REM periods are generally shorter than their associated NREM periods, a correlation of 0.90 would not be expected even for REMPs that are perfectly aligned with the peaks of the cosine function. Rapid eye movement periods that are thus aligned, yet which last one-fourth to one-half as long as the adjoining NREMPs yield correlations of approximately 0.65 to 0.80. As the cosinor analysis assumes a maximum theoretical vector length of 1.00, the p values obtained in these analyses are thus conservative. The mean point-biserial correlations for our subjects were 0.58 (for JA) and 0.50 (for TP). All vectors are plotted on circles of unit radius.

FIG. 2. Above: REM-NREM sleep distributions with sleep stage 2 (SS2) onsets (2) aligned (JA left; TP right). Shaded portions are REM; white portions are NREM. Brackets represent lights out (<) and awakening (>). **Below:** Incidence of REM sleep over nights 9–50, as a function of latency (minutes) from SS2 onset (JA left; TP right).


Day 1
5
10
15
20
25
30
35
40
45
50
SS2 Onset
0
60
120
180
240
300
360
420
480 minutes
Number of days with REM Sleep occurring at X minutes after SS2 Onset
0
5
10
15
20
25
30
35
40


Day 1
5
10
15
20
25
30
35
40
45
50
SS2 Onset
0
60
120
180
240
300
360
420 minutes
Number of days with REM Sleep occurring at X minutes after SS2 Onset
0
5
10
15
20
25
30
35
40

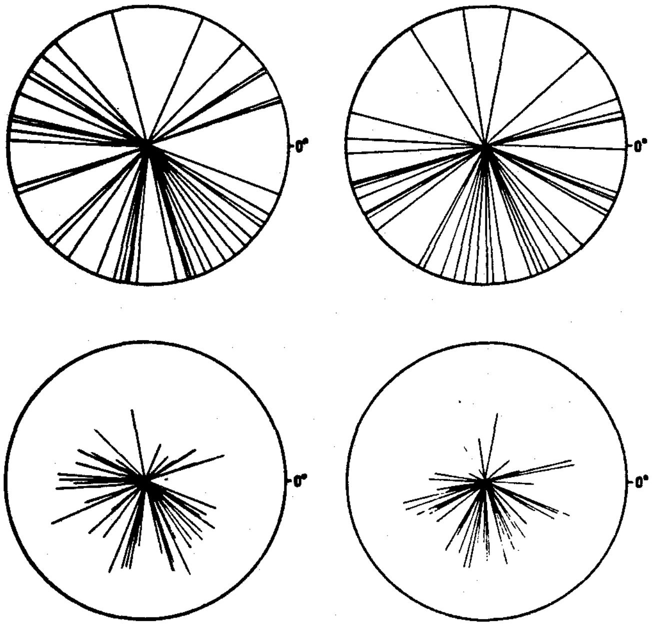


FIG. 3. **Above:** Phase of the REM-NREM sleep distribution for each night with respect to the origin at sleep stage 2 onset. For JA (left) the resultant mean vector has a magnitude of 0.28, with a direction of −90° ($p < 0.05$, Rayleigh test). For TP (right) the resultant mean vector has a magnitude of 0.37 and a direction of −83° ($p < 0.01$, Rayleigh test). **Below:** The same phase angles as above are plotted on unit circles with each vector having a magnitude proportional to the point-biserial correlation between that night's REM-NREM distribution and its cosine curve. (See footnote in text for further explanation.) For JA (left) the resultant mean vector has a magnitude of 0.19 with an angle of −97° ($p < 0.005$, cosinor zero amplitude test). For TP (right) the resultant mean vector has a magnitude of 0.21 and an angle of −84° ($p < 0.001$, cosinor zero amplitude test).

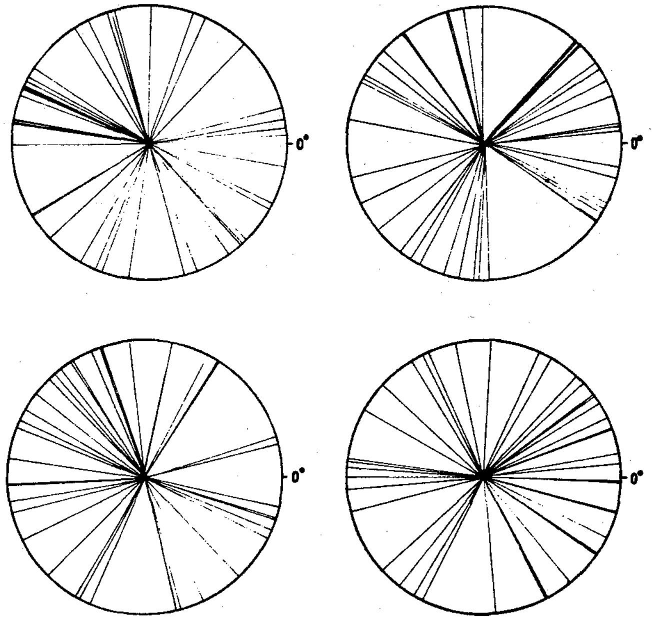


FIG. 4. **Above:** Change in the phase of the REM-NREM sleep distribution between pairs of consecutive nights, with phases determined with respect to origins in clock time (JA, left; TP, right). The distributions were not significantly different from random (Rayleigh test), indicating no systematic shift in the phase of REM sleep across nights. **Below:** Change in the phase of the REM-NREM sleep distribution between pairs of consecutive nights, with phases determined with respect to origins at sleep onset (JA, left; TP, right). As above, the distributions were not significantly different from random (Rayleigh test).

Next, the possibility of a shift in the phase of the distribution of REM sleep across nights was investigated by computation of the difference between phase angles for consecutive pairs of nights (Fig. 4). The resultant phase differences did not indicate a consistent shift in phase across nights, with the results of the Rayleigh test being consistent with a random change in the phase of REM sleep across nights ($p > 0.05$). This was found to be the case whether the phase differences were computed from phase angles obtained for the entire REM-NREM distribution for each night or from those obtained for the latter part of each night (e.g., all REM periods and intervening NREM periods excepting the first).

Last we examined the relationship of the REM-NREM sleep distribution to the circadian temperature minima. Aligning graphical representations of the REM-NREM sleep distribution for each night so that all temperature minima are aligned results in the sleep pattern portrayed in Fig. 5. Visual examination of the corresponding histogram reveals two distinct peaks in the occurrence of REM sleep across nights—one approximately 20 min after the temperature minimum, another approximately 100 min later.

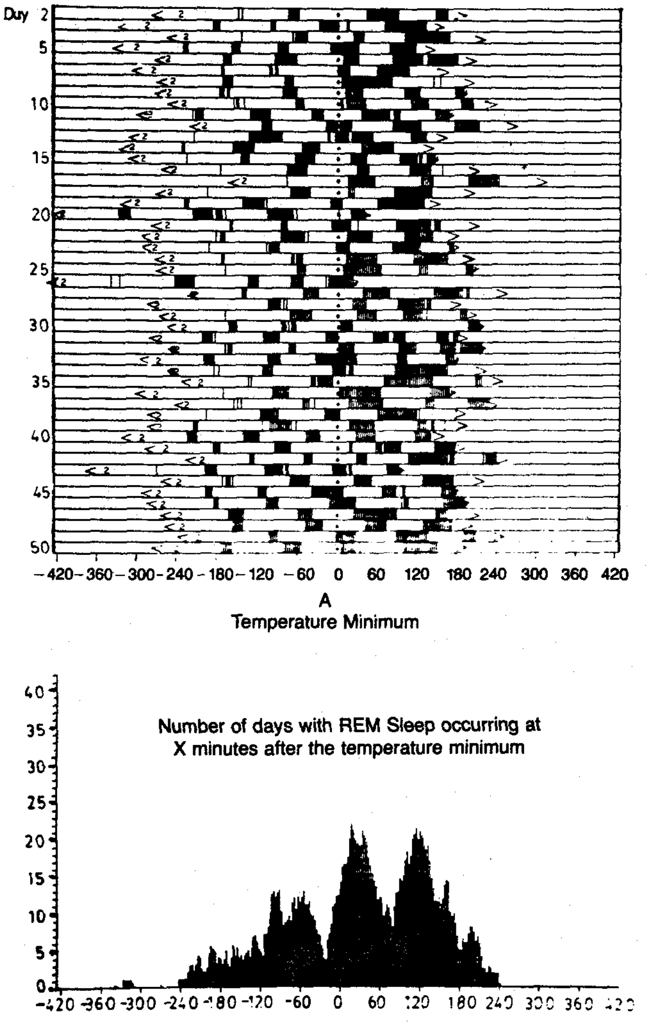


FIG. 5. Above: REM-NREM sleep distribution with temperature minima (*) aligned (subject TP). Shaded portions are REM, white portions are NREM. Brackets represent lights out (<) and awakening (>); (2) represents sleep stage 2 onset. No temperature data available for night 1. **Below:** Incidence of REM sleep over nights 9–50, as a function of latency (minutes) from the temperature minimum. This histogram differs from the one previously published (17) as a result of using temperature minima obtained by a fitting technique (see Methods) as opposed to using absolute minima as had been previously done.

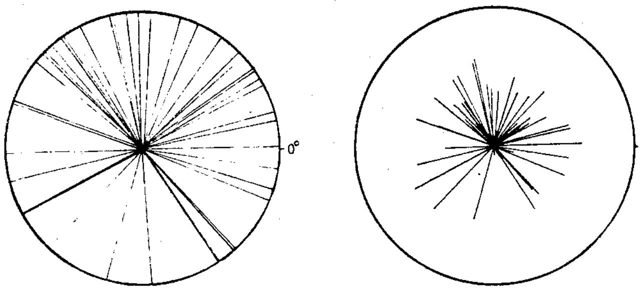


FIG. 6. Left: Phase angles for each night's REM-NREM distribution with respect to the origin at the temperature minimum (subject TP). The resultant mean vector has a magnitude of 0.28 and a direction of +79° ($p < 0.05$, Rayleigh test). **Right:** The same phase angles plotted on unit circles with each vector having a magnitude proportional to the point-biserial correlation between each night's REM-NREM distribution and the cosine approximating that distribution. (See footnote in text for further explanation.) The resultant mean vector has a magnitude of 0.13 and an angle of 89° ($p < 0.05$, cosinor zero amplitude test).

To determine the significance of this relationship, we examined the distribution of phase angles describing each night's REM-NREM sleep distribution with respect to an origin at the temperature minimum (Fig. 6). Using either unweighted or weighted vectors, the Rayleigh and cosinor tests respectively yield significance levels of $p < 0.05$.

DISCUSSION

The results of this study can be summarized as follows.

1. Even under conditions in which sleep onset time varies only slightly, the timing of all REMPs, especially the first REMP, is significantly determined by the timing of sleep onset. This result extends the results of Moses et al. (7) obtained from napping subjects and is consistent with the hypothesis that the timing of REMPs is sleep-dependent, rather than a manifestation of an ongoing BRAC.

2. Rapid eye movement periods exhibit no systematic drift across nights, even when the first or second REMP of the night is removed from the analysis. These results represent a failure to replicate the findings of Schulz et al. (6) and again are consistent with the sleep-dependent model.

3. There appears to be a relationship between the timing of the circadian temperature minimum and the timing of REMPs, such that the probability of REM sleep occurring is greatest approximately 20 min after the temperature minimum, and again 100 min subsequently. Although we have obtained this result for only one subject, it is similar to the findings Czeisler et al. obtained from free-running subjects (10,11) and to those of Carskadon and Dement from subjects on a 90 min sleep-wake schedule (13). The similarity of findings using three sleep-wake regimens suggests either that the REM-NREM cycle is influenced by the circadian temperature cycle or that a third factor influences them both. To the extent that the temperature cycle is entrained by other circadian cues, this relationship might explain previous findings that have suggested a sleep-independent component of the REM-NREM distribution (9).

We thus propose that the nightly distribution of REM and NREM sleep may be determined by at least two circadian events: the initial sleep onset and the subsequent nocturnal temperature minumum. We hope that these findings from normal subjects on a normal schedule will provide useful baseline data for investigations of abnormal or disordered sleep and contribute to the study of the sleep process.

Acknowledgment: These data were obtained at Cornell University, Ithaca, NY, by two undergraduate groups under the auspices of Dr. Barbara L. Finlay, and using facilities made available to us by Dr. James B. Maas, both of the Department of Psychology. We are indebted to Dr. Finlay for her advice and encouragement throughout the study. We are grateful for the assistance of our colleagues, who helped monitor the subjects and collect the data: Matthew Adler, Joe Adriano, Eric Aronson, Jeffrey Gottlieb, Sheri Kapel, Mark Kim, Barry Kushelowitz, Thomas Petrie, Diane Piraino, Robert Rene, Abbey Rosso, Anne Updegrove, David Wrobel, and Keith Zeno. We would also like to thank Dr. Daniel F. Kripke for providing us with instruction on methodology, as well as criticism and encouragement, during the course of our study. This research was supported in part by grants to LM and GJC from the Dean of the College of Arts and Sciences of Cornell University and by a National Science Foundation Graduate Fellowship awarded to GJC.

REFERENCES

1. Dement W, Kleitman N. Cyclic variations in EEG during sleep and their relations to eye movements, body motility, and dreaming. *Electroencephalogr Clin Neurophysiol* 1957;9:673–90.
2. Globus G. Rapid eye movement cycle in real time. *Arch Gen Psychiatry* 1966;15:654–9.
3. Othmer E, Hayden M, Segelbaum R. Encephalic cycles during sleep and wakefulness. *Science* 1969;164:447–9.
4. Kleitman N. *Sleep and wakefulness*. Chicago: University of Chicago Press, 1963.
5. Webb W, Agnew H. A superdian rhythm and its consequences. *Sleep Res* 1972;1:179.
6. Schulz H, Dirlich G, Zulley J. Phase shift in the REM sleep rhythm. *Pflugers Arch* 1975;358:203–12.
7. Moses J, Lubin A, Johnson L, Naitoh P. Rapid eye movement cycle is a sleep-dependent rhythm. *Nature* 1977;265:360–1.
8. Johnson L. The REM cycle is a sleep-dependent rhythm. *Sleep* 1980;2:299–307.
9. McPartland RJ, and Kupfer DJ. REM sleep cycle, clock time and sleep onset. *Electroencephalogr Clin Neurophysiol* 1978;45:178–85.
10. Czeisler C, Weitzman E, Moore-Ede M, Zimmerman J, Knauer R. Human sleep: its duration and organization depend on its circadian phase. *Science* 1980;210:1264–7.
11. Czeisler C, Zimmerman J, Ronda J, Moore-Ede MC, Weitzman ED. Timing of REM sleep is coupled to the circadian rhythm of body temperature in man. *Sleep* 1980;2:329–46.
12. Zulley J, Wever R, Aschoff J. The dependence of onset and duration of sleep on the circadian rhythm of rectal temperature. *Pflugers Arch* 1981;391:314–8.
13. Carskadon M, Dement WC. Distribtion of REM sleep on a 90-minute sleep-wake schedule. *Sleep* 1980;2:309–17.
14. Gillberg M, Harstedt T. Body temperature and sleep at different times of day. *Sleep* 1982;5:378–88.
15. Rechtschaffen A, Kales A, eds. *A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects*. Washington, DC: US Government Printing Office, 1968.
16. Halberg F, Johnson EA, Nelson W, Runge W, Sothern R. Autorhythmometry procedures for physiologic self-measurements and their analysis. *Physiol Teacher*. 1972;1.
17. Carman GJ, Mealey L. Relationship of sleep onset and the circadian temperature minimum to the distribution of REM sleep during normal human sleep. *Sleep Res* 1982;11:189.
18. Batschelet E. *Circular statistics in biology*. New York: Academic Press, 1981.
19. Nelson W, Liang Tong Y, Lee J, Halberg F. Methods for cosinor rhythmometry. *Chronobiologia* 1979;6:305–23.

Animal Learning & Behavior
1987, 15 (2), 124-129

Attempts to demonstrate magnetic discrimination by homing pigeons in flight

GEORGE J. CARMAN, MICHAEL M. WALKER, and ANDY K. LEE
California Institute of Technology, Pasadena, California

Eight homing pigeons, trained to fly between two elevated feeders within a flight tunnel, were tested for their ability to discriminate between two magnetic field stimuli and two acoustic stimuli, using a unitary discrete-trials procedure with successive presentation of stimuli. Magnetic stimuli consisted of the ambient magnetic field and a reduced magnetic field in which the vertical component of the field was reduced to 50% of its ambient value. Acoustic stimuli consisted of an ambient white noise and the white noise plus a tone. Stimuli were paired with food reward and either a time penalty (Experiment 1) or electric shock (Experiment 2). Although subjects could discriminate sounds with our procedures, none of the subjects demonstrated discrimination of magnetic fields. The failure of pigeons to discriminate magnetic stimuli is discussed as a consequence of either the failure to provide conditions sufficient for such discrimination or the absence of a magnetic sense in these animals.

Although behavioral experiments in the field have suggested the use of a magnetic sense in the navigation of homing pigeons (Frei & Wagner, 1976; Keeton, 1971; Walcott, 1978; Walcott & Green, 1974), various attempts to demonstrate sensitivity to magnetic stimuli in the laboratory have failed (Alsop, 1987; Beaugrand, 1976; Delius & Emmerton, 1978; Griffin, 1982; Kreithen & Keeton, 1974; McIsaac & Kreithen, 1987; Meyer & Lambe, 1966; Moore, Stanhope, & Wilcox, 1987; Orgel & Smith, 1954). If a magnetic sense exists in these animals, such failures to demonstrate magnetic discrimination in the laboratory could be due to the failure to satisfy any one of several prerequisites for such a demonstration. The suggestion that flight may be necessary for discrimination of magnetic stimuli by pigeons (Kreithen & Keeton, 1974) might explain the lack of discrimination of magnetic stimuli in conventional keypecking or cardiac conditioning paradigms. Such an explanation was supported by the experiments of Bookman (1977), who reported that mated pairs of homing pigeons discriminated magnetic stimuli only when they "fluttered" (performed sustained hovering, jumping, rapid turning, or short flights) before entering one or the other of two goalboxes located at one end of a flight tunnel.

We began our investigations with an attempt to replicate the experiments of Bookman (1977), duplicating his apparatus and using similar magnetic stimuli and procedures (M. A. Bookman, personal communication, 1983). Two mated pairs of homing pigeons were first pretrained to allow for adaptation to the flight tunnel and shaping of their responses. During the subsequent magnetic discrimination training, we observed only random performance over a course of 307 trials for one pair and 149 trials for the other pair (Carman & Mahowald, 1984). Random performance was observed regardless of whether we examined all responses made by each pair or only the responses made during trials accompanied by the "fluttering" behavior that Bookman (1977) had claimed to be correlated with discrimination of magnetic stimuli.

Next we attempted control experiments using a variety of light and sound stimuli to determine whether the two-choice paradigm used by Bookman was capable of demonstrating discrimination via known sensory modalities. To facilitate data collection, we automated the experiments, thereby eliminating the handling of birds by experimenters and the associated possibility of inadvertent cues. In addition, we used individual homing pigeons, rather than mated pairs of birds, the latter having often been observed to engage in a variety of behaviors other than the desired goal-directed response. Although we observed discrimination of discrete light stimuli contiguous with the goalboxes and discrimination of diffuse light stimuli presented between the goalboxes, we did not observe discrimination of either discrete, contiguous sound stimuli or diffuse sound stimuli in any of 4 birds tested (Carman & Mahowald, 1984). The failure of this paradigm to demonstrate discrimination of such known sensory stimuli raised serious doubts about its suitability for demonstrating the existence of a hypothetical magnetic sense.

Such discriminative choice experiments, in which stimuli are presented successively, with only one stimulus present during each trial, are difficult discrimination tasks and frequently fail with well-understood, salient stimuli (Bitterman, 1979; Mackintosh, 1974). The failure

---

We thank J. Kirschvink and M. Konishi of the California Institute of Technology for providing laboratory space and partial support for these experiments, and M. E. Bitterman of the University of Hawaii for valuable advice and criticism. We also thank Michelle Mahowald for assisting in the construction of the apparatus and in the replication of Bookman's experiments, and Senior Instrument Specialist Herb Adams of Caltech's Central Engineering Services for the design and construction of the feeder platforms. This research was supported in part by NIH BRSG Grant RR07003 and by an NSF Graduate Fellowship to G.J.C. The authors' mailing address is: Division of Biology 216-76, California Institute of Technology, Pasadena, CA 91125.

of the above experiments to produce discrimination not only of magnetic stimuli but also of acoustic stimuli led us to employ a unitary discrimination training procedure developed by Woodard and Bitterman (1974) in our subsequent experiments. Such unitary procedures, in which the rate of a single response to each of two successively presented stimuli is measured, yield discrimination more readily than choice procedures (Mackintosh, 1974), while requiring multiple responses during trials improves discrimination (Bitterman, 1979). In our experiments, individually trained birds were required to shuttle repeatedly between two cylindrical feeders positioned at opposite ends of a flight tunnel during each trial. Depending on the magnetic field stimulus present during a trial, this response was either rewarded with food or punished with a time penalty (Experiment 1) or electric shock (Experiment 2) at the end of the trial.

MATERIALS AND METHODS

Apparatus

The present experiments were performed in an elevated flight tunnel (interior measurements: 1.07 m wide × 0.89 m high × 3.40 m long) constructed from wood and nonferrous hardware (Figure 1). Within the tunnel were located two cylindrical feeders (20.3 cm in diameter and 40.6 cm high), also constructed of nonferrous materials. Atop each feeder was a spring-supported platform coupled to three microswitches wired in series, which allowed for the automatic collection of data. Each feeder contained a pneumatic piston-driven timed-access feeder that delivered split peas through a central hole (2.54 cm in diameter) located at the center of each platform and equipped with an infrared photobeam to detect feeding. Atop each platform were a set of eight concentric aluminum rings connected to a remote high-voltage pulse generator that could deliver trains of constant power shocks to the feet of the pigeons while they stood atop the feeder. The interior of the tunnel was illuminated during trials by two lamps powered by a remote 12-V dc supply, providing an average luminance of 1.6 fL. Access to the interior of the tunnel was provided through hinged side panels, three of which made up each side.

For magnetic discrimination training, the magnetic field within the tunnel was manipulated by three pairs of circular coils, each coil consisting of 150 turns of 22-ga wire wound on a 1.11-m-diam styrofoam core (Figure 1). The coils were powered by a custom-designed three-channel current source capable of producing steady state magnetic fields with less than ±11 gamma of noise (1 gamma = 10^{-5} G). Using this apparatus, we produced two magnetic field conditions for use as stimuli: (1) the ambient magnetic field (AMF) due to the local geomagnetic field, and (2) the reduced magnetic field (RMF), in which the field produced by the coils reduced the vertical component of the local geomagnetic field to approximately 50% of its normal value. In contrast, the horizontal component of the magnetic field for the two conditions differed on the average by less than 1% in magnitude and in direction (see Table 1). Measurements of the magnetic field were made with a Develco fluxgate magnetometer (Model 9200C) and were accurate to within ±1%.

During both magnetic and acoustic discrimination training, white noise was used to mask environmental sounds throughout the sessions. The white noise was delivered from a speaker centered beneath the flight tunnel at an intensity of 75 dB SPL measured in the flight tunnel. For acoustic discrimination training, we also used a tone with audible harmonic distortion. This tone was produced by driving a second speaker centered beneath the flight tunnel with a 440-Hz square wave at an intensity of 95 dB SPL measured in the flight tunnel. Two acoustic conditions were used as stimuli: (1) the ambient sound consisting of the white noise (AS), and (2) the altered sound consisting of the white noise and the tone (RS).

A microcomputer located in a nearby room controlled all aspects of the experimental procedure, including the timing of trials and intertrial intervals, the sequence of magnetic or acoustic stimuli, the delivery of reward and punishment, and the automatic recording of behavior. A closed-circuit television camera (Sony AVC-3260) allowed us to monitor and record the activity of the pigeons within the flight tunnel on video cassettes to provide an auxiliary record of behavior.

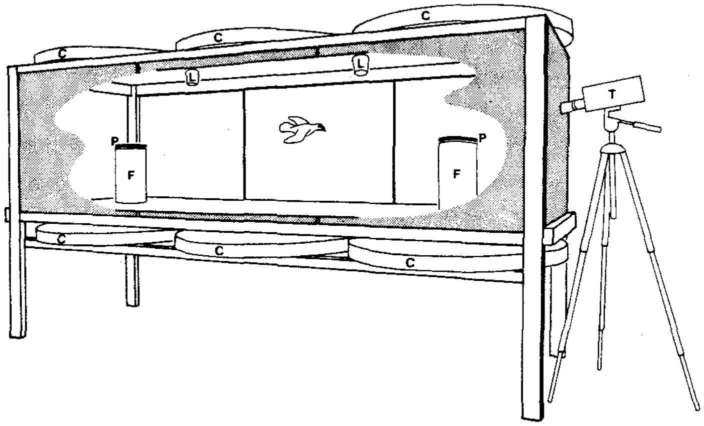


Figure 1. Perspective view of the flight tunnel used in the present experiments. Portions of the side panels have been cut away to reveal the interior. F, cylindrical feeders; P, platforms atop feeders with concentric aluminum rings (not shown) for the delivery of electric shock; C, coils for the control of magnetic stimuli; L, tunnel lights; and T, television camera. See text for details.

Table 1
Mean and Standard Deviation of Magnetic Field Components as a Function of Magnetic Field Condition

Magnetic Field Condition	Horizontal Component				Vertical Component	
	Magnitude (gamma)		Direction (degrees)		Magnitude (gamma)	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
AMF	20,010	±2,870	69.2	±3.11	−40,250	±1,580
RMF	19,820	±2,470	68.9	±3.51	−19,700	±2,540
[AMF−RMF]	192	±534	0.35	±1.63	−20,550	±1,450

Note—Means and standard deviations were calculated for 10 measurements taken at equidistant locations along the flight path from one feeder to the other under each magnetic field condition (AMF and RMF), and for the 10 differences in each measurement between the two conditions ([AMF−RMF]). Note that 1 gamma = 10^{-5} G. The direction of the horizontal component is given relative to the long axis of the tunnel, as measured counterclockwise from the TV camera end when viewed from above.

Subjects and Experimental Procedure

The subjects were 8 experimentally naive homing pigeons of the gray check and wine check varieties, aged 6 to 24 months. All birds had been raised at the local Caltech loft and were given the opportunity for daily flight experience. All birds had demonstrated their homing ability in at least three homing trials over distances of at least 8 km.

The subjects were isolated in individual cages, given free access to water and grit, and switched from their normal mixed-grain diet to limited amounts of split peas. Pretraining began when the birds were reduced to about 90% of their ad-lib weights. With the flight tunnel darkened, each bird was placed upon the feeders, which were initially located together at the midpoint of the flight tunnel. When the tunnel was illuminated, a few split peas placed near the central hole in the feeder platform attracted the attention of the birds to the hole. The feeder mechanisms then were operated alternately, leading the birds back and forth from one feeder to the other. As the distance between the feeders was progressively increased over a period of days, the birds rapidly learned first to step, then to hop, and finally to fly between the feeders. During this time, the birds also learned to mount the feeders from the floor of the tunnel.

Once each bird had learned to shuttle between the feeders, the trial procedure was incorporated into the pretraining. Each daily session consisted of 10 trials, each trial beginning with the illumination of the tunnel and the presentation of the magnetic field associated with reinforcement (S+). During each trial, each bird flew back and forth between the feeders, with each complete shuttle (i.e., leaving one feeder and arriving at the other) being counted as a response. At the first response after 90 sec, the bird was given 15 sec access to food, after which the trial was terminated, the tunnel lights turned off, and the AMF continued or restored for the duration of the 210-sec intertrial interval. This procedure continued until each bird had completed at least 3 days of pretraining with the feeders separated by 229 cm, the distance used during discrimination testing. Body weight was maintained at a fixed percentage (±5 g) between 70% and 80% of ad lib for each bird by supplemental feeding when necessary after each session.

Experiment 1. During discrimination training, each bird was given 20 trials in each daily session, 10 with the reinforced stimulus (S+) and 10 with the nonreinforced stimulus (S−). Stimuli were presented in a balanced, quasi-random order (Gellermann, 1933). Training was balanced for stimuli, with 2 birds receiving AMF as S+ and RMF as S− and 2 birds receiving RMF as S+ and AMF as S−. On S+ trials, at the first response after 90 sec, the bird was given 15 sec access to food, after which the trial was terminated. For S− trials, the bird was given a time penalty whose duration was determined by a timer that was reset by subsequent responses either until the bird ceased responding for 15 sec, thus terminating the trial, or until a total of 180 sec of penalty time had accrued, at which time the trial was terminated. During the intertrial interval, whose duration was varied pseudorandomly from 30 to 90 sec with a mean of 60 sec, the tunnel lights were out and the AMF was present.

Experiment 2. Discrimination training of 4 additional birds proceeded as in Experiment 1 except for the substitution of electric shock for the time penalty during S− trials. For these trials, responses after 90 sec resulted in an electric shock to the bird for 15 sec or for as long as it remained atop the conductive rings on either feeder platform, after which the trial was terminated. Alternatively, no shock was given if no responses occurred between 90 and 105 sec, at which time the trial was terminated.

At least 10 sessions (200 trials) of magnetic field discrimination training were followed by at least 10 sessions (200 trials) of acoustic discrimination training as a control for both Experiment 1 and Experiment 2. The procedures used for acoustic discrimination training were identical to those used for magnetic discrimination training, except that the magnetic stimuli were replaced by the acoustic stimuli (AS replaced AMF and RS replaced RMF). The AMF was present in the tunnel throughout these control experiments.

We chose this stimulus order (first magnetic, then acoustic) in order to avoid the possibility that experience with the ambient magnetic field during both S+ and S− acoustic discrimination trials might interfere with subsequent magnetic discrimination training had the alternative stimulus order been chosen. Since the acoustic discrimination training was performed only as a test of the effectiveness of our procedures, we did not attempt to balance the order of the two stimulus types, but chose instead to optimize the conditions for magnetic discrimination. Although this resulted in a confounding of the stimulus type (magnetic or acoustic) with the stimulus order (first or second), we considered it extremely unlikely that discrimination would depend on the order of the stimuli rather than on the type of stimulus used.

RESULTS

Experiment 1

Data consisted of the number of responses within the first 90 sec of each trial during the last 10 sessions of magnetic discrimination training and the last 10 sessions of acoustic discrimination training for each subject. Using *these data, we calculated the mean and standard error for* both the S+ and the S− trials of each session (Figure 2). These data were also used in an analysis of variance (BMDP8V ANOVA, equal cell size and mixed models,

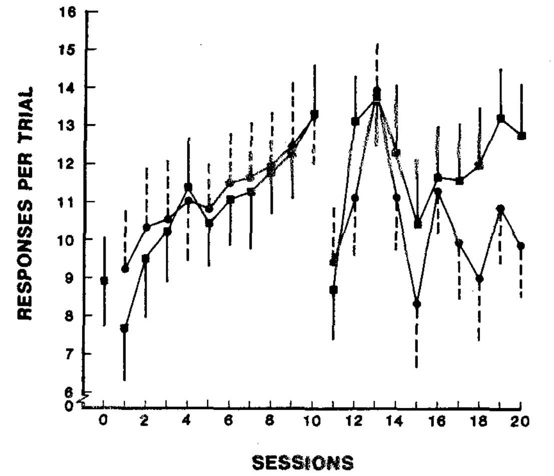


Figure 2. Mean and standard error of the number of responses for S+ trials (squares and solid error bars) and S− trials (circles and dashed error bars) for the 4 subjects given a time penalty as punishment on S− (Experiment 1). Each point represents the mean of 10 trials from a single daily session for each of 4 subjects. Session 0: last day of pretraining (using S+ alone). Sessions 1 to 10: magnetic discrimination training. Sessions 11 to 20: acoustic discrimination training.

BMDP Statistical Software, Inc., Los Angeles, CA 90025) in order to examine the main effects and interactions of various factors in our experimental design (Table 2).

No bird demonstrated discrimination of the magnetic stimuli (Figure 2, Sessions 1 to 10). This is confirmed in the analysis of variance by the absence of a significant main effect of reinforcement and the absence of a significant interaction of reinforcement with sessions (experience), which would be expected if subjects had discriminated the stimuli and learned their association with reinforcement or nonreinforcement (Table 2, sources R and R × Se). The only significant effect found for magnetic discrimination training was a steady increase in the rate of responses to both magnetic stimuli during the course of training (Table 2, source Se). Such an effect would be expected if the birds had not reached their maximum rates of response by the beginning of discrimination training.

Table 2
Results of ANOVA for Experiment 1

		Magnetic Stimuli		Acoustic Stimuli	
Source	*df*	*F*	*p*	*F*	*p*
R	1	1.30	0.3718	9.30	0.0928
St	1	0.01	0.9298	0.08	0.8010
Se	9	2.58	0.0413*	1.06	0.4319
R×St	1	0.01	0.9452	0.08	0.8070
R×Se	9	0.56	0.8117	3.37	0.0136*
St×Se	9	1.73	0.1550	0.66	0.7351
R×St×Se	9	0.39	0.9247	1.56	0.2008

Note—Key to source: R = *main effect of reinforcement (reward or* punishment); St = main effect of stimuli (AMF or RMF; AS or RS); Se = main effect of sessions (experience); R×St = interaction of reinforcement with stimuli; R×Se = interaction of stimuli with sessions (learning); St×Se = interaction of stimuli with sessions; R×St×Se = interaction of reinforcement with stimuli and sessions (interaction of stimulus conditions and learning). *Probabilities less than the .05 criterion value, indicating significant main effects or interactions.

In contrast, comparison of the number of responses for S+ and S− trials over the course of acoustic discrimination training shows a progressive separation of the response rates to the two acoustic stimulus conditions (Figure 2, Sessions 11 to 20). Although the analysis of variance of these data revealed no significant main effect of reinforcement, there is a significant interaction of reinforcement with sessions (experience) (Table 2, sources R and R × Se). This result would be expected in the case of discrimination that is acquired partway through the course of training.

Experiment 2

As was observed in Experiment 1, no bird in Experiment 2 demonstrated discrimination of the magnetic field stimuli (Figure 3, Sessions 1 to 10). In contrast, however, we observed a decline in the rate of response to both magnetic stimulus conditions during training. We attributed this decline in response rate to an avoidance of the feeders due to the experience of shock. However, neither this decline in rate nor any other effect proved statistically significant in the analysis of variance (Table 3).

Separation of the response rates to the reinforced and punished acoustic stimulus conditions as well as a partial recovery of response rates were observed in the second half of Experiment 2 (Figure 3, Sessions 11 to 20). Two of the birds (those given AS as S+ and RS as S−) achieved and maintained discrimination of the acoustic stimuli to the end of training. The remaining 2 birds (those given RS as S+ and AS as S−) initially achieved discrimination, but failed to maintain this discrimination due to a feeder malfunction, during Session 17, which prevented the delivery of the reinforcement under S+. The effect of this malfunction can be observed in the mean of the rates of response of all 4 birds (Figure 3, Session 17). Although this malfunction occurred for only one session, it resulted in a markedly reduced separation of the response rates under the two acoustic stimulus conditions for the remainder of the sessions for these 2 birds.

These results are reflected in the analysis of variance of the data, which showed neither a stimulus main effect nor a learning interaction (Table 3, sources R and R × Se). There was, however, a significant interaction of reinforcement, stimuli, and experience (Table 3, source R × St × Se). Although this interaction could be interpreted as an effect of the pairing of stimuli with reinforcement upon learning during training, it probably represents contamination of a learning effect by the decreased performance of the 2 birds that experienced the feeder malfunction, both of which were given the same pairing of stimuli and reinforcement.

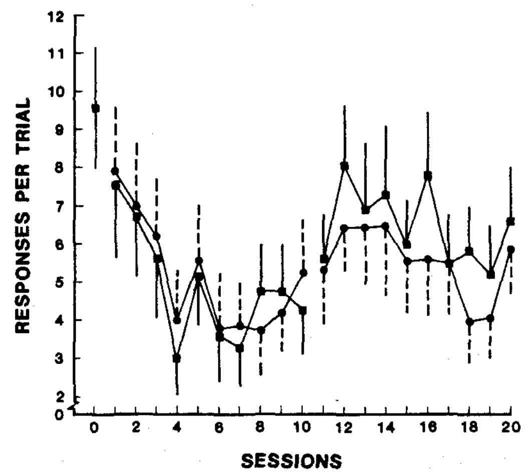


Figure 3. Mean and standard error of the number of responses for S+ trials (squares and solid error bars) and S− trials (circles and dashed error bars) for the 4 subjects given an electric shock as punishment on S− (Experiment 2). Each point represents the mean of 10 trials from a single daily session for each of 4 subjects. Session 0: last day of pretraining (using S+ alone). Sessions 1 to 10: magnetic discrimination training. Sessions 11 to 20: acoustic discrimination training.

DISCUSSION

A successful demonstration of sensory discrimination depends on the simultaneous satisfaction of a number of prerequisites, including (1) the use of a species suspected of possessing the sensory modality in question, (2) the selection of sensory stimuli thought to be discriminable to the animal, (3) the determination of a conditioned response that the animal is capable of associating with such stimuli, and (4) the design of experimental apparatus and procedures that allow both for the detection of stimuli and for their association to the conditioned behavioral response. The failure to demonstrate discrimination of magnetic field stimuli in the present experiments could be due to one or more of these prerequisites' not having been satisfied, a possibility for which there are precedents in other sensory modalities of the pigeon (Kreithen, 1978; Ossenkopp & Barbeito, 1978). The results of our control experiments, which used diffuse acoustic stimuli, suggest that our design of apparatus and our procedures were adequate for demonstrating discrimination within one of the more challenging modalities for such experiments. Nevertheless, allowing pigeons to "flutter" or training them to fly back and forth through different magnetic fields was not sufficient to permit discrimination of magnetic stimuli, as had been previously suggested (Bookman, 1977; Kreithen & Keeton, 1974).

Table 3
Results of ANOVA for Experiment 2

		Magnetic Stimuli		Acoustic Stimuli	
Source	*df*	*F*	*p*	*F*	*p*
R	1	1.02	0.4183	6.67	0.1229
St	1	0.01	0.9217	0.11	0.7749
Se	9	0.96	0.4995	1.08	0.4201
R×St	1	0.13	0.7546	0.18	0.7152
R×Se	9	0.93	0.5244	1.50	0.2207
St×Se	9	0.45	0.8899	1.13	0.3895
R×St×Se	9	1.09	0.4191	3.05	0.0212*

Note—See Table 2 for key to source.

The two magnetic fields we used as stimuli were chosen for their similarity to those used by Bookman (1977), in that they differed mostly in their vertical intensity (see Table 1). Despite considerable care taken in the positioning of the flight tunnel within the laboratory space so as to obtain the most uniform ambient magnetic field within the tunnel, there was a residual spatial variation or gradient of magnetic field intensity that was similar under the two stimulus conditions. The average difference between the two magnetic field stimuli was 10 times larger than the variation within each stimulus due to this gradient. However, it is possible that this gradient may have prevented discrimination of the two stimuli, either because the spatial pattern of magnetic field intensity common to the two stimuli may have been more salient to the animal than the difference in mean vertical intensity between the stimuli or because the resultant variation interfered with a magnetic sense. Support for this interpretation comes from reports of disorientation of pigeons released at magnetic anomalies exhibiting comparable variation in magnetic intensity over much larger distances (Frei & Wagner, 1976; Wagner, 1983; Walcott, 1978). Thus, the use of magnetic field stimuli that have either different spatial gradients and equal mean intensities or no gradients and different mean intensities might yield discrimination in future experiments.

The isolation and control of stimulus variables that can be achieved in the laboratory would be difficult, if not impossible, to duplicate in the field. However, such laboratory experiments may not adequately simulate the behavioral context of homing to be useful in exploring all the sensory mechanisms used in navigation. By holding other stimuli constant and restricting the range of behavior during magnetic field discrimination training, we present the animal with an impoverished environment that may not adequately approximate that encountered by the homing pigeon released at an unfamiliar site to motivate or permit the use of a navigational magnetic sense. Such use of magnetic cues in the field and their nonuse in the laboratory could be more a matter of unconscious reflex than of conscious choice on the part of the animal. If this were the case, it might be very difficult, if not impossible, for the animal to form associations between magnetic stimuli and arbitrary behaviors under laboratory conditions, a circumstance that also has its precedents (LoLordo, 1979). This interpretation suggests that it would not be possible to demonstrate discrimination under laboratory conditions, not because the animal could not sense magnetic fields, but rather because its use of magnetic information was restricted to navigational contexts.

If a magnetic sense exists in these animals, then our experiments either serve to constrain the conditions under which magnetic sensitivity can be demonstrated or suggest that such laboratory experiments may fail for reasons beyond experimental control. Although the failure to demonstrate discrimination of magnetic fields does not prove that a magnetic sense does not exist, our findings are also consistent with the absence of magnetic sensitivity in homing pigeons. There now are numerous reports of failure to demonstrate magnetic discrimination, whereas the only two claims of magnetic discrimination (Bookman, 1977; Reille, 1968) have not withstood replication. Regardless of whether or not these animals possess a magnetic sense, these results make it less likely that magnetic discrimination by these animals will be demonstrated under laboratory conditions. Therefore, it may be more appropriate to attempt to replicate and extend the field studies that suggested the existence of a magnetic sensitivity in homing pigeons in the first place.

REFERENCES

Alsop, B. (1987). A failure to obtain magnetic discrimination in the pigeon. *Animal Learning & Behavior*, **15**, 110-114.

Beaugrand, J. P. (1976). An attempt to confirm magnetic sensitivity in the pigeon, *Columba livia*. *Journal of Comparative Physiology*, **110**, 343-355.

Bitterman, M. E. (1979). Discrimination. In M. E. Bitterman, V. M. LoLordo, J. B. Overmier, & M. E. Rashotte (Eds.), *Animal learning: Survey and analysis* (pp. 413-443). New York: Plenum Press.

Bookman, M. A. (1977). Sensitivity of the homing pigeon to an earth-strength magnetic field. *Nature*, **267**, 340-342.

Carman, G. J., & Mahowald, M. (1984). [Attempted replication of M. A. Bookman's (1977) experiments.] Unpublished data.

Delius, J. D., & Emmerton, J. (1978). Sensory mechanisms related to homing in pigeons. In K. Schmidt-Koenig & W. T. Keeton (Eds.), *Animal migration, navigation, and homing* (pp. 35-41). New York: Springer-Verlag.

Frei, U., & Wagner, G. (1976). Die Anfangsorientierung von Brieftauben im erdmagnetisch gestörten Gebiet des Mont Jorat. *Revue Suisse de Zoologie*, **83**, 891-897.

Gellermann, L. W. (1933). Chance orders of alternating stimuli in visual discrimination experiments. *Journal of Genetic Psychology*, **42**, 206-208.

Griffin, D. R. (1982). Ecology of migration: Is magnetic orientation a reality? *Quarterly Review of Biology*, **57**, 293-295.

Keeton, W. T. (1971). Magnets interfere with pigeon homing. *Proceedings of the National Academy of Sciences*, **68**, 102-106.

Kreithen, M. L. (1978). Sensory mechanisms for animal orientation—can any new ones be discovered? In K. Schmidt-Koenig & W. T. Keeton (Eds.), *Animal migration, navigation, and homing* (pp. 25-34). New York: Springer-Verlag.

Kreithen, M. L., & Keeton, W. T. (1974). Attempts to condition homing pigeons to magnetic stimuli. *Journal of Comparative Physiology*, **91**, 355-362.

LoLordo, V. M. (1979). Constraints on learning. In M. E. Bitterman, V. M. LoLordo, J. B. Overmier, & M. E. Rashotte (Eds.), *Animal learning: Survey and analysis* (pp. 473-504). New York: Plenum Press.

Mackintosh, N. J. (1974). *The psychology of animal learning*. New York: Academic Press.

McIsaac, H. P., & Kreithen, M. L. (1987). Attempts to condition homing pigeons to magnetic cues in an outdoor flight cage. *Animal Learning & Behavior*, **15**, 118-123.

Meyer, M. E., & Lambe, D. R. (1966). Sensitivity of the pigeon to changes in the magnetic field. *Psychonomic Science*, **5**, 349-350.

Moore, B. R., Stanhope, K. J., & Wilcox, D. (1987). Pigeons fail to detect low-frequency magnetic fields. *Animal Learning & Behavior*, **15**, 115-117.

Orgel, A. R., & Smith, J. C. (1954). Test of the magnetic theory of homing. *Science*, **120**, 891-892.

Ossenkopp, K.-P., & Barbeito, R. (1978). Bird orientation and the geomagnetic field: A review. *Neuroscience & Biobehavioral Reviews*, **2**, 255-270.

Reille, A. (1968). Essai de mise en évidence d'une sensibilité du pigeon au champ magnétique à l'aide d'un conditionnement nociceptif. *Journal de Physiologie* (Paris), **60**, 85-92.

Wagner, G. (1983). Natural geomagnetic anomalies and homing in pigeons. *Comparative Biochemistry & Physiology*, **76A**, 691-700.

Walcott, C. (1978). Anomalies in the earth's magnetic field increase the scatter of pigeons' vanishing bearings. In K. Schmidt-Koenig & W. T. Keeton (Eds.), *Animal migration, navigation, and homing* (pp. 143-151). New York: Springer-Verlag.

Walcott, C., & Green, R. P. (1974). Orientation of homing pigeons altered by a change in the direction of an applied magnetic field. *Science*, **184**, 180-182.

Woodard, W. T., & Bitterman, M. E. (1974). A discrete-trials/fixed-interval method of discrimination training. *Behavior Research Methods & Instrumentation*, **6**, 389-392.

(Manuscript received December 11, 1985;
revision accepted for publication July 1, 1986.)

In vivo *functional localization of the human visual cortex using positron emission tomography and magnetic resonance imaging*

Bassem N. Mora, George J. Carman and John M. Allman

Bassem N. Mora, George J. Carman and John M. Allman are at the Division of Biology 216–76, California Institute of Technology, Pasadena, CA 91125, USA.

Positron emission tomography (PET) and magnetic resonance imaging (MRI) are two recently developed methods for imaging the human brain in vivo. *One application of PET measures stimulus-evoked changes in cerebral blood flow while MRI provides a detailed anatomical map of the brain. Here we report the combined application of these two techniques in the same human subject. Subtracted PET scans of a brain receiving visual stimulation were superimposed upon MRI images of the same brain. The PET scans were converted into the MRI coordinate space before superposition, which allowed for a more precise correlation between MRI anatomical data and PET physiological data. Responses were localized in striate and extrastriate visual areas as well as in the posterior thalamus.*

The application of physiological and anatomical techniques[1–3] has led to the discovery in non-human primates of a large number of cortical visual areas, some of which have definite functional specializations. Knowledge of the organization of human visual cortex has been mainly limited to the inferences gained from the observation of patients afflicted by brain injuries[4–6]. However, the advent of new brain imaging techniques has made it possible to study experimentally the organization of human visual cortex. Stimulus-evoked changes in regional cerebral blood flow (rCBF) monitored with positron emission tomography (PET) can be used to investigate the physiological responses of the brain. The anatomical structure of the brain can be obtained by proton nuclear magnetic resonance imaging (MRI). In this study, we have sought to combine the physiological-localizing ability of PET with the anatomical-resolving power of MRI in parallel observations made in the brain of the same subject.

PET measures local concentrations of positron-emitting compounds injected into living tissue. Cortical blood flow is measured by monitoring changes in the concentration of water labeled with oxygen-15 ($H_2{}^{15}O$) in the brain. Increases in neural activity within a brain region lead to increases in rCBF in that region[7–9]. Recently, the retinotopic organization of primary visual cortex has been mapped in normal human volunteers using stimulus-evoked changes in rCBF[10]. This was accomplished by first obtaining a control PET scan, where the subject is looking at a small fixation point on a computer screen. This control scan measures background brain activity. The presentation of different stimuli on the computer screen then activates cortical areas, which leads to increased rCBF, which is detected by PET. The stimulus PET scan therefore contains background brain activity in addition to cortical activity due to the stimulus. Subtraction of the control from the stimulus PET scans results in a map of cortical activity due to the visual stimulus[11].

The PET scans used in this study were performed at Washington University School of Medicine, MI, USA, by Raichle, Fox, Miezin, Allman and co-workers. The PETT VI system was used, which simultaneously acquires seven parallel horizontal brain slices[12]. Each horizontal slice consisted of a 100 × 100 matrix of measurements (2.7 mm × 2.7 mm) with an interslice center-to-center distance of 14.4 mm. The effects of global CBF fluctuations on each of the seven horizontal PET scans were minimized by normalizing the images such that the total brain blood flow was 50 ml per 100 g per min, as described elsewhere[11,13].

After performing the fixation-point control PET scan, a stimulus was presented on the computer screen, and the stimulus scan was taken. This sequence of control/stimulus scans was repeated four times, once for each of the four stimuli that were used. All stimuli consisted of a red and black checkerboard annulus with a central fixating point and varying eccentricities. The red and black checks alternated at a frequency of 10 Hz in order to maximize the induced CBF responses[8]. The stimuli were: a *macular* annulus, extending radially from 0.1° to 1.5°, with a radial check size of 0.5°; a *perimacular* annulus, ranging from 1.5° to 5.5°, with a radial check size of 1.0°; and *peripheral* hemi-annuli, extending from 5.5° to 15.5°, one *upper-field* and the other *lower-field*, with a radial check size of 2.0°.

MRI is a technique developed during the past decade to provide high-resolution anatomical images of living tissue by recording the mobility of protons. It provides a high level of contrast between gray and white matter. The MRI images used in this study, taken of the same subject at an earlier date, were obtained using a Diasonics 0.35 T scanner at the Huntington Medical Research Institutes, CA, USA. The images were obtained in the coronal and parasagittal planes using the spin-echo technique with a relaxation time, T_r, of 3 s and an echo time, T_e of 100 ms[14]. The parasagittal images were used for PET–MRI superpositioning, with an in-plane resolution of 0.95 mm and a slice thickness of 2.7 mm. After transfer to a Masscomp 5700 computer, each 256 × 256 pixel MRI image was expanded by linear interpolation in each of two orthogonal directions in the image plane to yield a 768 × 768 pixel image.

The alignment of the PET slices with respect to cranial features was determined from a lateral X-ray radiograph taken at the time of the PET scan with the subject's head held in a fixed position relative to the scanner. The radiograph in the upper-left panel of Fig. 1 shows the outline of the skull in the mid-sagittal plane with the centers of the seven horizontal PET slices. The radiograph was superimposed on the mid-sagittal MRI to align the PET and MRI coordinate systems. This resulted in the PET–MRI superpositions shown in Fig. 1, with the PET scans displayed as a transparent overlay in register with the MRI images. The scale used in the displays is illustrated at the top of Fig. 1, with MRI image intensity increasing to the right, and PET image intensity increasing to the bottom. Maximum PET response is in red to allow easier localization of the response focus. The upper-right panel of Fig. 1 depicts the cortical response to macular stimulation *without* subtraction of the fixation-point control scan. The bottom four panels of Fig. 1 illustrate the *net* responses to macular, perimacular, lower- and upper-field stimulation after the subtraction of the matching fixation-point control scans. They are all displayed using the same intensity scale. In the center-left panel, the macular stimulus produced a focal response at the posterior tip of the calcarine sulcus. The perimacular stimulus produced a response centered immediately anterior to the macular stimulus (center-right panel). The lower-field peripheral stimulus produced a response centered in the upper bank of the calcarine sulcus (lower-left panel), while the upper-field peripheral stimulus produced a response centered in the lower bank of the calcarine sulcus (lower-right panel). Changing the portion of the visual field being stimulated therefore shifts the PET response locales in visual cortex. The visuotopic organization thus revealed closely matches that inferred from visual-field defects produced by restricted visual-cortex lesions in humans[4–6].

Figure 2 shows the net responses to macular stimulation obtained from the extrastriate cortex and posterior thalamus in the same subject. The

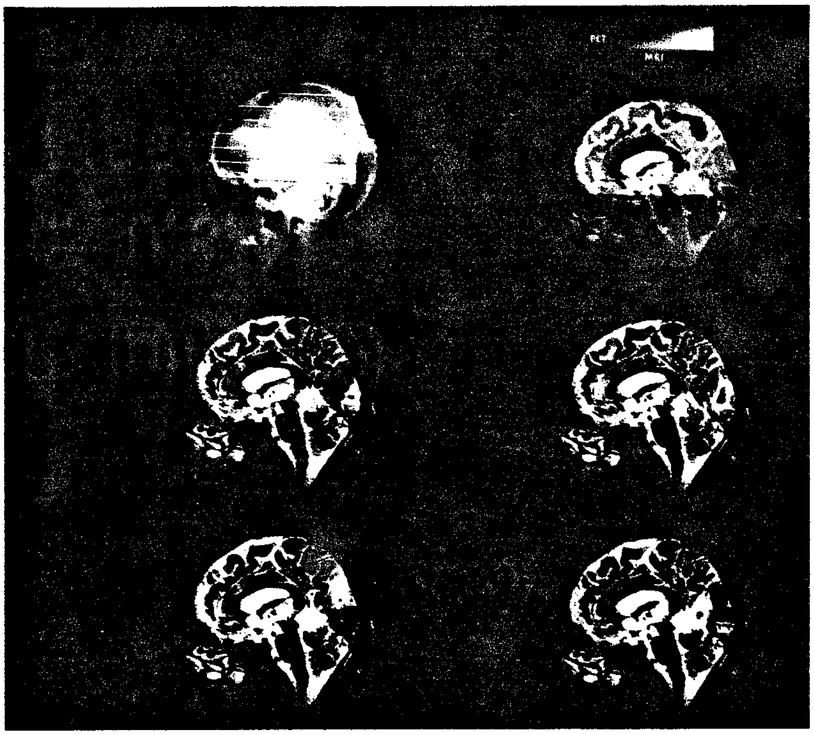


Fig. 1. *PET–MRI superimposed images from the same subject, showing cortical responses due to visual cortex stimulation.* Top left: *X-ray radiograph of the subject's head in the mid-sagittal plane with the centers of the seven horizontal PET planes delineated, plane 1 being the top-most plane and plane 7 the bottom-most plane. The remaining images are PET–MRI superimposed images localizing physiology on to anatomy. Shown to the left of each PET–MRI subtracted image is the stimulus used to activate the corresponding cortical structures. Shown at the top is the two-dimensional linear scale used in the display of the images: the horizontal axis is used to represent the MRI data, with MRI image intensity increasing horizontally to the right; the vertical axis is used to represent the PET data, with PET image intensity increasing vertically downwards. Thus, points in the superimposed images that have maximal PET activity and maximal MRI signal are represented in the bottom-right of the scale. The red in the PET scale allows easier localization of the peak PET responses in the images.* Top right: *cortical response to macular stimulation without subtraction of the control scan.* Center left: *subtracted response to macular stimulation following subtraction of the control scan.* Center right: *subtracted response to perimacular stimulation.* Bottom left: *subtracted response to peripheral lower-field hemi-annulus stimulation.* Bottom right: *subtracted response to peripheral upper-field hemi-annulus stimulation.*

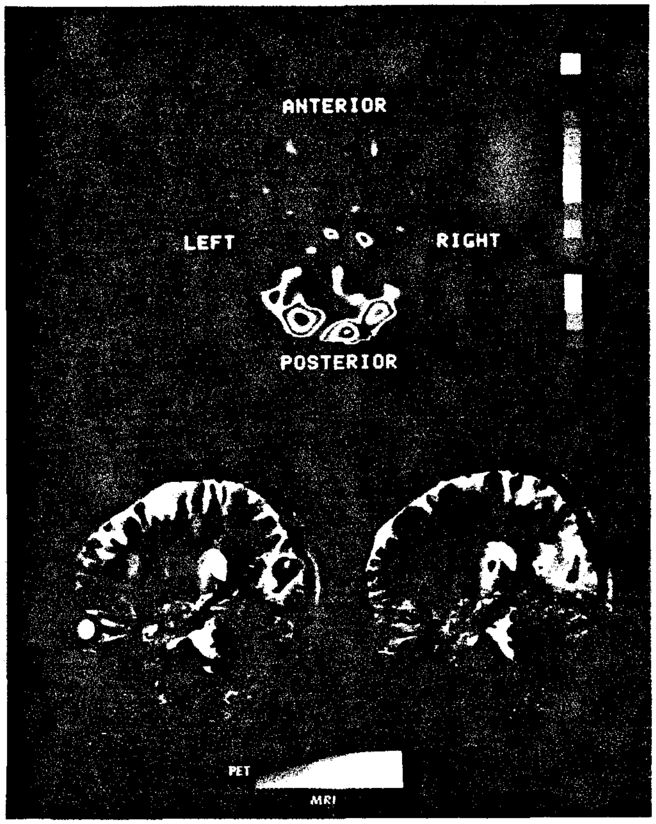


Fig. 2. *Subtracted responses to macular stimulation in extrastriate cortex and posterior thalamus in the same subject.* Top panel: *Conventional display of the subtracted response from horizontal PET plane 6.* Bottom panel: *parasagittal PET–MRI images illustrating subtracted cortical response at 2 cm from the midline. Right and left panels correspond to the right and left hemispheres of the brain, respectively. Note the activation of extrastriate visual cortex and the posterior region of the thalamus, which are more clearly observed using PET–MRI superpositioning (bottom) compared with a conventional PET display (top). The scale used in this display was different from that used in Fig. 1.*

upper portion is the response from horizontal PET plane six showing a peak centered on the midline, which intersects with the peak illustrated in the center-left panel of Fig. 1. The lower panels of Fig. 2 are parasagittal images, 2 cm to the left and right of the midline, which contain approximately symmetrical responses in the extrastriate visual cortex. These lateral extrastriate zones probably include the temporal-occipital-parietal pit (TOPP), which in other subjects was found to respond to fast-flickering and low-contrast, moving dot patterns and may correspond to area MT[15–17]. There are also responses in the posterior thalamus, which may arise in part from the lateral geniculate nucleus but appear to be centered more dorsally in the pulvinar complex. In monkeys, the pulvinar contains several maps of the visual field, in a manner analogous to the visual areas in extrastriate cortex[18,19].

PET–MRI superposition permits the observation of physiological responses displayed in register with brain structures from the same individual. This contrasts with the standard localization method used in PET studies in which averaged data from different subjects are referred to coordinates in a stereotactic atlas[13]. This method of referring PET data to a standard atlas is valuable but suffers from the difficulties presented by the substantial inter-subject variability among human subjects[20–22]. This problem of variability may be particularly acute in extrastriate cortex where the responses are lower in amplitude and the size of the constituent areas may be smaller than in primary visual cortex.

Selected references

1 Allman, J. M. and McGuinness, E. R. (1988) in *Comparative Primate Biology, Vol. 4: The Neurosciences* (Steklis, H. D. and Erwin, J., eds), pp. 279–326, Alan R. Liss
2 Van Essen, D. C. (1985) in *Cerebral Cortex* (Vol. 3) (Peters, A. and Jones, E. G., eds), pp. 259–329, Plenum Press
3 Zeki, S. M. (1978) *Nature* 274, 423–428
4 Inouye, T. (1909) *Die Sehstörungen bei Schussverletzungen der kortikalen Sehsphäre nach Beobachtungen an Versundeten der letzten Japanische Kriege* W. Engelmann
5 Holmes, G. and Lister, W. T. (1916) *Brain* 39, 34–73
6 Teuber, H., Battersby, W. and Bender, M. (1960) *Visual Field Deficits after Penetrating Missile Wounds* Harvard University Press
7 Brodersen, P. *et al.* (1973) *Arch. Neurol.* 28, 334–338
8 Fox, P. T. and Raichle, M. E. (1984) *J. Neurophysiol.* 51, 1109–1120
9 Paulson, O. B. and Newman, E. A. (1987) *Science* 237, 896–898
10 Fox, P. T. *et al.* (1986) *Nature* 323, 806–809
11 Fox, P. T., Miezin, F. M., Allman, J. M., Van Essen, D. C. and Raichle, M. E. (1987) *J. Neurosci.* 7, 913–922
12 Yamamoto, M., Ficke, D. C. and Ter-Pogossian, M. M. (1982) *IEEE Trans. Nucl. Sci.* 29, 529–533
13 Fox, P. T., Perlmutter, J. S. and Raichle, M. E. (1985) *J. Comp. Ass. Tomog.* 9, 141–153
14 Bradley, W. G., Newton, T. H. and Crooks, L. E. (1983) in *Modern Neuroradiology, Vol. 2: Advanced Imaging Techniques* (Newton, T. H. and Potts, D. G., eds), pp. 15–61, Claradel Press
15 Allman, J. (1988) in *Visual Processing of Form and Motion* (Rauschecker, J. and Mohn, G., eds), pp. 33–34, Tübingen
16 Miezin, F. M., Fox, P. T., Raichle, M. E. and Allman, J. M. (1987) *Soc. Neurosci. Abstr.* 13, 631
17 Miezin, F. M., Fox, P. T., Raichle, M. E. and Allman, J. M. (1988) *ARVO Abstr.* p. 326
18 Allman, J. M., Kaas, J. H., Lane, R. H. and Miezin, F. M. (1972) *Brain Res.* 40, 291–302
19 Bender, D. B. (1981) *J. Neurophysiol.* 46, 672–693
20 Talairach, J. and Szikla, G. (1967) *Atlas d'Anatomie Stereotaxique du Telencephale* Masson
21 Schaltenbrand, G. and Bailey, P. (1959) *Introduction to Stereotaxis with an Atlas of the Human Brain* Thieme
22 Stensaas, S. S., Eddington, D. K. and Dobelle, W. H. (1974) *J. Neurosurg.* 40, 747

Acknowledgements

We thank Marcus Raichle, Peter Fox, Fran Miezin, Joel Perlmutter and Tom Videen at Washington University School of Medicine in St Louis for their time and effort in performing the PET scans, and for generously offering technical advice and assistance, David Van Essen of the California Institute of Technology for his interest in this project and for use of his computer laboratory, and William Bradley of the Huntington Medical Research Institutes for providing the MRI images. This work has been supported by grants from NIH, the Gordon Trust, the Sloan Foundation, and the McDonnell Center for Higher Brain Research. Part of the equipment used in this project was acquired through support from Office of Naval Research contract number N0014-85K-0068 to D. Van Essen.

Spatial distribution of Purkinje cell lineages in cerebellar cortex suggests early developmental restrictions to neuroblast mixing

Karl Herrup[1,3]
George Carman[2]
James Bower[2]

[1] Dept. of Developmental Neurobiology
E.K. Shriver Center
200 Trapelo Road
Waltham, MA 02254

[2] Division of Biology
California Inst. of Technology
Pasadena, CA 91025

[3] To whom correspondence should be addressed

Abstract

Spatial relationships among Purkinje cell lineages have been examined in the cerebellar cortex of aggregation mouse chimeras. ß-glucuronidase was used as an independent cell marker to allow computer assisted microscopic reconstruction of large areas of Purkinje cell genotype distributions in specified regions of hemispheric folia crus II and the paramedian lobule. Application of a c^2 statistic to the data reveals a high degree of non-randomness in these distributions. The data are discussed in terms of the general questions of cell mixing and the role of cell lineage in the development of cortical networks in the mammalian CNS.

The question of where and how the complex structure of the brain is determined is one of the great unsolved mysteries in the life sciences. In recent years, developmental neurobiologists have approached the problem using model systems in creatures at all points along the phylogenetic scale. In general, work from invertebrate systems such as fruit fly and nematode, has led the way in uncovering some of the fundamental genetic bases of pattern formation in the CNS. These basic mechanisms, however have not as yet been translated in such a way that they have helped us to understand the developmental complexities of the larger brained animals in the vertebrate phyla. One particularly strong example of this is the mammalian cerebral cortex. Distinct organization features of cortex run in both the radial direction (along which cell types and properties vary giving rise to the system of the six main cortical layers) and in the circumferential- or tangential direction (along which the thickness and density of the individual layers varies giving rise to the different cytoarchitectonic areas). It is not yet known where or how this pattern is specified during development.

Examples from invertebrate systems suggest that early combinations of deterministic decisions and cell:cell and cell:chemical interactions lay down the basic plan of the CNS. From this has come the realization that cell lineage relationships are an important part of understanding the control of CNS development in any organism (Herrup,1986; Sanes, 1989; Cepko). Specifically if early differentiative decisions are made by altering a cell's developmental genetic makeup then these decisions will be shared by all of the mitotic descendants of that progenitor - even if those descendants are widely separated in location. (e.g. Wetts & Herrup, 1983). At the same time, even if cell lineage does not regulate cell fate, the location of the various lineages in the three dimensional space of the brain can serve as a valuable indicator of the presence of restrictions to cell mixing. It is this latter issue that forms the basis for this report.

The possible role of a cell's mitotic lineage in determining its final form and function has long intrigued developmental biologists (Wilson, 1898), and recent advances in techniques for histologically identifying clonally related cells (Weisblat *et al.*, 1978; Price *et al.*., 1987; Sanes *et al.*., 1986) have led to a renewed interest in the role of cell lineage in the development of the vertebrate central nervous system (Jacobson, 1985; Kimmel and Law, 1985; Kimmel and Warga, 1986; Sanes, 1989; Price *et al.*., 1987; Wetts and Fraser, 1988; Herrup and Crandall, 1989). One outcome of these studies has been to begin to sketch the spatial constraints that are placed on cells during their developmental migrations. For cerebral cortex, the results indicate that there is a substantial radial bias to the migration of clonally related neurons. This result was expected given the known radial orientation of glial cells believed to provide a substrate for neural migration (Rakic, 1978; Rakic, 1988). An unexpected result, however, was that there is also a significant amount of horizontal mixing. How the radial and horizontal components of genotype mixing in this system are related and what their consequences are for the development of cerebral cortex is a topic of current debate

In contrast to cerebrum, the cerebellar cortex is a remarkably uniform anatomical structure with a stereotyped repeating pattern of neuronal types and inter-neuronal connectivities. As a result, there is much less area to area variation in network structure. The only cortical projection neurons, the Purkinje cells, are largely uniform in their anatomical features, further reducing the complexities introduced by different subclasses of cerebral cortical cells each with different shapes and cortical connection patterns. In addition, detailed information is available about the two dimensional distributions of other features of cerebellar cortex including various enzymes and antigens (Wassef and Sotelo, 1984; Hawkes *et al.*., 1985; Chan-Palay *et al.*., 1981; Bower *et al.*,1989 in preparation), the spatial patterns of afferent projections (e.g., Armstrong, 1974; Brodal, 1940, Bower and Woolston, 1983; Welker, 1987), and the spatial distributions of physiological properties of neuronal responses (Sasaki, Bower and Llinas, 1989). As one of the ultimate objectives of cell lineage studies is to determine what influence lineage has on the development of the final form and function of neural systems, the opportunity to correlate lineage with anatomical, physiological, and biochemical features will presumably be important. The cerebellum offers one attractive model system in which to approach this question.

We have approached the question of the clonal origin of cells and the mixing of cell lineages during the establishment of cortical structures in cerebellum. We began with this model since the cerebellum represents a system in which only the horizontal dimension is relevant to questions of mixing. Further, relatively simple statistical techniques are accessible to quantify the significance of the two dimensional distributions of cell lineage. We illustrate here the construction of Purkinje cell linage maps through the use of mouse aggregation chimeras. The spatial distribution of the lineages we have determined reveal a high degree of non-randomness. The implications of this finding are discussed in terms of two models of cerebellar (cortical) development.

Materials and Methods

Animals

Inbred mouse strains were purchased from the Jackson Laboratories in Bar Harbor, ME. Random bred, CD-1 females were purchased from Charles River Lab, Kingston. All mice were maintained on a 14/10 hour light/dark cycle with food and water *ad libitum.* Animals used for this study were sacrificed between 3 and 12 months of age.

Chimeras were produced by standard protocols (Mintz, 1962; Mintz, 1965; Tarkowski, 1961; Mullen & Whitten, 1971). One embryo of each chimera was C57BL/6 in background; the other was C3H/HeJ. This combination insures that each successful chimera was a mosaic of cells with intrinsically high (C57BL/6 — Gus^{b}/Gus^{b}) or low (C3H/HeJ —Gus^{h}/Gus^{h} ß-glucuronidase activity.

Histology

Animals were sacrificed by transcardial perfusion under deep Avertin anesthesia. The fixative used was ice-cold 4% phosphate buffered paraformaldehyde. The protocols used have been described previously (Mullen, 1977). Immediately prior to wax (polyethylene glycol distearate 400) embedding, the brains were bissected on the midline. Serial sagittal sections were collected through each hemi-cerebellum. Beginning at the vermis/hemisphere junction (approximately 1 mm from the midline) every second, or in some cases every third, section was stained for ß-glucuronidase activity by the method described in Mullen (Mullen, 1977). The substrate used was naphthol-AS-BI-ß-D-glucuronide.

Lineage distributions across the two-dimensional Purkinje cell layer were reconstructed using computer assisted acquisition, display, and analyses of data.

Histological sections were scored using a Leitz Orthoplan microscope equipped with a Stahl 517MF micropositioning stage used to determine the x,y coordinates of each Purkinje cell soma. These data were recorded along with the genotype (Gush^{h} or Gus^{b}) of that cell. The z coordinate was calculated as the basis of the number of the µm section. since Purkinje cells have a diameter of approximately 20 µm in our material, sections were scored at intervals of 24 µm on average to assure that each Purkinje cell was recorded only once. Once recorded, these data were transferred to a Silicon Graphics Iris 2400 graphics workstation to align the various sections and reconstruct a full three-dimensional model of each folium. Lineage maps of the two-dimensional Purkinje cell layer were obtained by color-coding the cells according to genotype (Fig. 1D). Such maps were obtained for the dorsal regions of two mouse cerebellar folia, crus II (cII) and the paramedian lobule (pml). Eight lineage maps of cII and eleven maps of pml were obtained from a total of nine different $Gus^{h}/Gush^{b}$ chimeras. On average, the maps included 750 Purkinje cells and covered an area of 1000 µm x 700 µm. These lineage mapswere projected onto a plane osculating the dorsal surface of the folia (Fig. 1E). In order to render these results in a more easily visualizable form, a 6 x5 grid was superimposed on the resconstruction. For each grid square, the difference between the ratios of the observed to the expected number of cells was calculated.

Results

Twenty mouse chimeras were constructed using standard protocols for the aggregation of pre-implantation embryos at the eight cell stage of embryogenesis (Tarkowski, 1961; Mintz, 1962; Mintz, 1965). With this technique, the mouse that is eventually born to the host mother is a mix of the cells descended from each of the two original embryos.. If the genotypes of the embryos are chosen such that one embryo carries a marker genetic locus, then cells from the different genotypes can be distinguished in the adult chimeric mouse. In this study, each chimera was constructed from embryos that differed in the structural gene of the enzyme ß-glucuronidase (gene symbol: *Gus*). One allele of the ß-glucuronidase gene, Gus^h, causes a cell autonomous reduction in enzyme activity throughout the brain. The normal activity allele is denoted Gus^b. Chimeras were made from one C57BL/6 embryo (Gus^b/Gus^b) and one C3H/HeJ embryo (Gus^h/Gus^h). Following histological preparation (Mullen, 1977), Purkinje cells with Gus^b/Gus^b genotype, and hence normal glucuronidase activity appeared distinctly red (Fig. 1A) while cells with Gus^h/Gus^h genotype appeared unstained (Fig 1B). As reported previously, the genotype related staining differences are retained in the mosaic cerebellar cortex (Fig. 1C). A set of serial 10um sagittal sections was collected from half-cerebellar. For the analysis reported here the dorsal surface of the paramedian lobule and Crus I in the cerebellar hemispheres were reconstructed as described in the methods section briefly. A computer readable microscope stage was used to digitize the position and genotype of each Purkinje cell. Subsequently, sections were aligned (Fig. 1D) and rotated to produce a view perpendicular to a plane tangential to the cerebellar surface (Fig. 1E). This view of cerebellar cortex is, in essence a lineage map.

Inspection of the 19 reconstructed lineage maps reveals clear intermixing of the two lineages throughout the dorsal surface of the hemispheres. In no part of any map is one or another lineage found exclusively. However, examination of the reconstructions reveals striking regional variations in the relative occurrence of the two genotypes across the folia (Figure 1E). Such spatial variation would be expected if the clones (or polyclones) or ventricular zone cells remain at least partially segregated during the development. In order to test for the significance of this perceived mixing, a χ^2 analysis was computed for each lineage map as follows. For each map, the "global" ratios for the number of cells of each genotype to the total number of cells was determined. The lineage map was then subdivided into a 6 x 6 grid of quadrats (average size175 μm x 120 μm) and the number of cells of each of the two genotypes counted within each quadrat. The expected number of *Gush* cells was calculated for each quadrat by multiplying the global ratio forGus^h by the total number of cells in each quadrat. A similar calculation was performed for Gus^b cells. The numbers of cells of each genotype were then compared against the expected values by calculation of the χ^2 across the two genotypes and all quadrats. Quadrats for which the expected number of cells of either genotype was less than five were omitted from the analysis. Probability (p) values were determined for the χ^2 using df = (Number of quadrats -1) either by reference to standard statistical tables when df <= 30, or by conversion to the alpha point on the normal distribution when df > 30 (Selby, 1971). Since the size and alignment of the quadrats on the lineage map will in general affect the value of the χ^2 obtained, this analysis was performed for several different grid sizes (not shown). In addition, offsets for each of the maps in increments of one-third grid square in both X and Y direction were done.

The results of these analyses are summarized in Table 1. The value of χ^2 will underestimate the true spatial variability according to the mismatch between the location and scale of quadrats and any "genotypic domains". Therefore, for any given grid size (scale), we may select the location of the grid of quadrats which yields the largest χ^2 as our best estimate of the degree of variability for that scale. Accordingly, the probabilities that such distributions could arise by chance were computed based on the largest χ^2 obtained for various locations of a fixed grid size. Of the 19 lineage maps, all but 3 showed significant ($p< 0.05$) nonrandom variation in the distribution of Purkinje cell genotypes across the maps. Significant spatial variation was observed for all grid offsets tested for 12 of the lineage maps, while the remainder of the maps had significant variation for at least some grid sizes or offsets. In several cases, the probability of the genotypic distributions arising by chance was vanishingly small -- less than or equal to zero to four decimal places. As a demonstration that our analysis was sensitive to the spatial distribution of genotypes, and not to either the spatial locations of cells or to the genotypes of cells alone, we modified maps by randomization of either the positions or the genotypes of cells in one of the lineage maps previously found to be highly significant. As expected, no significant variation was found in either of these randomized maps at any grid size or offset tested.

In order to obtain a visualization of the spatial variation detected by the χ^2, we constructed two dimensional plots of the relative deviation of the observed from the expected number of cells for the two genotypes. For each quadrat, the difference between the ratios of the observed to the expected number of cells for the two genotypes was calculated. $d_{bh} = (n_{oh} - n_{eh})/n_{eh} - (n_{ob}-n_{eb})n_{eb}$.

$= (n_{oh}/n_{eh}) - (n_{ob}/n_{eb})$

where n_{oh} and n_{ob} are the observed numbers of cells for Gus^h/Gus^h and $Gus^b/Gus^{b,}$ respectively, and n_{eh} and n_{eb} are the numbers of cells expected under the null hypothesis of complete mixing for Gus^h/Gus^h and $Gus^b/Gus^{b,}$ respectively. This quantity is equivalent to the difference between the deviation χ, relative to the values expected under the null hypothesis, and was selected to provide a measure of spatial variation detected by the χ^2. Values of d_{bh} will be either negative, zero, or positive depending on whether the number of cells observed relative to the number expected for Gus^b/Gus^b was less than, equal to, or greater than that of Gus^h/Gus^h · These values are then bilinearly interpolated in the plane of the quadrats, and displayed using a color scale in which negative values are shown as shades of red, while positive values are shown as shades of green.

Since the value of d_{bh} will be either negative or positive according to the relative contributions of Gus^b/Gus^b or Gus^h/Gus^h to each quadrat, the red and green patches observed in these plots may be interpreted as *genotypic domains* in which one or the other genotype is present in numbers in excess of that expected on the basis of complete mixing of the two genotypes. Regions shaded red represent the genotypic domains of Gus^b/Gus^b, while shaded green represent the genotypic domains of Gus^h/Gus^h A plot of raw data in Fig. 1E is shown in Fig. 1F. In this case, each grid square contains an average of approximately 26 Purkinje cells.

Discussion

The results our study demonstrate that the spatial distribution of Purkinje cell lineages in mouse cerebellar cortex is highly non-random. Cell lineage was revealed by ß-glucuronidase staining of sections from Gus^h/Gus^h Gus^b/Gus^b chimeras followed by computer assisted reconstructions of approximately $1mm^2$ areas of the dorsal surfaces of two hemispheric folia from several different animals. Non-randomness was determined by application of a c^2 statistic to the reconstructions.

Statistical considerations

The detection of significant variation in the spatial distribution of the two genotypes in the Purkinje cell layer can be approached in several ways.Our choice of the χ^2 analysis was based on several factors. First, it is sufficient to reveal significant spatial variation in the genotypic distributions. Second, other authors have made use of methods similar to ours in that they were based on a measure which assumed the χ^2 distribution (e.g. Schmidt et al. 1985). Third, while the χ^2 statistic may be of limited power, this could only result in a failure to detect significant spatial variation in some cases where it actually existed (failure to reject the null hypothesis when false, or beta error), and does not detract from findings of significant spatial variation in those cases where it is found to exist. Fourth, since our lineage maps were based on incomplete reconstructions of the Purkinje cell layer, (some cells located primarily in unstained sections are likely to have been missed) it was not possible to make use of other methods, such as the Grieg-Smith ANOVA (Grieg-Smith,1952; Mead, 1974) or other nearest-neighbor methods (Ripley, 1981; Diggle, 1983) to search for significant variation at different spatial scales. All factors considered, we believe our application of the χ^2 offers a quantitative, objective evaluation of the spatial variation in the distribution of the two genotypes which provides a conservative measure of the probability that such variation occurred by chance.

In analyzing the data, calculations were repeated after the position of the grid relative to the data set was varied in both X and Y directions by 1/3 of a grid square (9 different locations). We believe that it is appropriate to select the location of the grid that yields the largest χ as the most appropriate descriptor of the data set. The reason for this is best appreciated if one considers performing such an exercise with a checkerboard composed of squares of monolithic Gus^h/Gus^h or Gus^b/Gus^b cells. If a grid is placed on this data set such that it is offset by half of a grid square in each direction, then the genotype ratio of each grid square would be equal, yet no one would argue that the overall spatial distribution of cells was random.

Relationship to previous studies in cerebellum

Previous authors (Mullen, 1978; Mullen, 1977; Oster-Granite and Gearhart, 1981) have examined the organization of the Purkinje cell lineages in cerebellar cortex. Most of these studies have focussed on whether or not there are large homogeneous patches of lineage-related Purkinje cells. They, as we, have found no evidence for this sort of lineage distribution across the two dimensions of the Purkinje cell layer. Oster-Granite and Gearhart (1981) have suggested, based on a linear analysis of $Gpi\text{-}1^a/Gpi\text{-}1^a \leftrightarrow Gpi\text{-}1^b/Gpi\text{-}1^b$ chimeras, that there is a tendency for clones of 4 to 8 cells to remain spatially coherent in a small clonal patch. Mullen (1978; 1977) has argued that there is a near random dispersion of the Purkinje cell lineages. The latter analysis was of both single sections from Gus^h/Gus^h Gus^b/Gus^b chimeras and qualitative examination of a reconstruction of a 1.1 mm^2 region from the vermis of a single *pcd/pcd* wild type chimera. No tendency for clonally related Purkinje cells to remain contiguous was reported.

We too find that there are no large clusters of clonally related Purkinje cells, but our analysis does not support the idea that cell lineage distributions are random. We have found that lineage relationships form a pattern in which significant shifts in genotype ratios can occur over distances of less than 100 μm. At the same time, areas of common genotype ratios often extend over 8 - 10 individual grids including hundreds of Purkinje cells. The single section analyses of the previous work might not have allowed the authors to discover this result. Further, the qualitative nature of the previously reported reconstructions might not be sensitive to this finding.

In his reconstructions of Purkinje cells lineages, Mullen (1978) did call attention to significant differences in the lineage ratios of Purkinje cells when comparisons were made between entirely different cerebellar folia. Analysis of the current data indicates that, of the nine chimeras in which both pml and cII were analyzed, three showed marked differences in the global ratio of Gus^h/Gus^h to Gus^b/Gus^b between the two folia. However, our data shows much more consistently significant variations in genotype ratios within restricted regions of the same folia. Again, the methods used previously would not have detected these local patterns.

Significance for cerebellar development

Our results confirm the fact that Purkinje cells are not grouped into genotypically pure clusters. As suggested by others this rules out the possibility that single progenitors give rise to progeny that are completely restricted in their mixing potential. That lineage ratios in different regions do not appear to shift in quantal increments further suggests that the progenitors for each region are not equivalent to the set of 8 -10 Purkinje cell progenitors established during early stages of CNS development (Herrup and Sunter, 1986; Wetts and Herrup, 1982). Instead they are most likely polyclones of the later descendants of the original progenitors.

The lineage maps that we have reconstructed from adult chimeras represent the end of a process involving differentiation, mitosis, migration and cortical expansion. This suggests that the factors involved in establishing each Purkinje cell's final location will evade simple explanation. These are essentially two models that we feel could best be used to explain our results. One involves the existence of developmental boundaries, the other is a no-boundary model. The no boundary model draws on principles currently used to describe the formation of the early universe. For example, one could speculate that the non-random mixing patterns we observe arise from the slight intermixing of progeny at each stage of mitosis. In this case, we could further assume that there is a defined physical radius of intermixing that is maintained throughout the development of the system. As the real areal extent of the Purkinje cells progenitor population became larger, the effective mixing across the population would accordingly decrease. Simulations of other systems under similar assumptions have shown that inhomogeneties can be traced back to the most initial conditions of the system.

On the other hand, specific mixing boundaries could be established either by establishing physical barriers or changing the properties of two groups of cells such that mixing is not possible. The process of segmentation in *Drosophila* offers one extreme example of this scheme (Lawrence, 1981; Crick and Lawrence, 1975). In the vertebrate, our own results and those of others make this version of the model implausible, but a less restrictive variation could be entertained. Figure 2 is an illustration of how the appearance of boundaries in development might work to produce the results that we have observed. The two "cells" in the lower part of the figure are meant to represent the Purkinje cell progenitors as they are selected in the early neuroepithelium(Herrup and Sunter, 1986; Wetts and Herrup, 1982; Herrup, 1986). Through cell division, probably accompanied by a great deal of cell mixing, these cells give rise to descendants that are the sole source of the Purkinje cell population. At some stage before mitosis of this population is complete, an independent process sets up restrictions to mixing that bound off small cohorts of cells (the middle tier of the figure). If these restrictions are maintained during the translocation of cells from the ventricular zone to the cerebellar cortical plate, then variations in the ratios of the founder populations will be retained and produce the non-random distributions that we observe in the adult cerebellar Purkinje cell layer (upper tier of the figure). Presumably the specific paths of cell migration and pattern of formation of the Purkinje cell monolayer will affect the final form of the lineage maps. For all these reasons, additional conclusions require a much more detailed consideration of the developmental problem. In our view, a complete understanding of this process will require, additionally, model based simulations. Data, such as those reported here, obtained by actual reconstructions will be crucial for this effort.

As discussed above, one of the strong motivations for studying cell lineage in the developing CNS is the possibility that the clonal relationships may reveal several organizing features of CNS development. In order to suggest the idea that variations in lineage distribution could have significance for the development of cerebellar cortex or any other structure, it will be necessary to demonstrate that regions defined by lineage patterns are co-extensive with other significant features of the structure such as anatomy, physiology or biochemistry. In this regard, comparisons among lineage maps of the same folia in different animals suggest certain inter-animal consistencies. For example, we observed that the lateral regions of the paramedian lobule in 7 of 11 cases included a large area of relatively uniform genotypic ratio (see asterisk in Fig. 1F). This raises the possibility that different areas of the cortex might have different developmental constraints and it is noteworthy that this same region is frequently the site of a large receptive field from cutaneous receptors in the lower lip of the mouse (Bower, unpublished). The cerebellum represents an excellent model system since the spatial mapping of many physiological and biochemical modalities is well worked out. Future studies will be aimed at establishing just such correlations.

Acknowledgments

The authors wish to express their gratitude to Dr. David Van Essen of the Division of Biology, at Caltech for his permission to use the computer assisted microscope reconstruction system described here.

Thanks also to Jim Kinerim and David Bilitch for much selfless programming assistance, and to Karen Sunter for her valuable technical assistance.

The work described in this report was funded by the NIH (NS 18381 and NS 20591 to KH) and the March of Dimes Birth Defects Foundation (Grant #1-763 to KH). Dr. Herrup also wishes to express his appreciation for the Cornelius Wiersma Visiting Professorship that funded his stay at Caltech and made the completion of these experiments possible.

Table I

Chi-square analysis of the Purkinje cell lineage data

Animal	Folium	fraction Gus^h/Gus^h global ratio	lowest p value for c^2	Significance*
2kh	cII	.757	≤.500*	0
2kh	pml	.781	≤.900*	1
2ks	pml	.055	≤.005	0
3kh	cII	.877	≤.010	2
3kh	pml	.828	≤.005	2
3pp	cII	.324	≤.100*	2
3pp	pml	.362	≤.750*	1
4kh	pml	.756	≤.900*	1
4kh	cII	.781	≤.250*	0
5ppl	pml	.341	≤.0050	2
5ppl	cII	.666	≤.005	2
5ppr	pml	.339	≤.002	3
6ppl	pml	.425	≤.030	3
8ppl	cII	.270	≤.001	1
8ppl	pml	.556	≤.001	3
8ppr	cII	.542	≤.001	3
8ppr	pml	.324	≤.010	3
380r	cII	.657	≤.500*	1
380r	pml	.515	≤.001	1

* Data distributions were tested at each of 9 positions of the grid relative to the reconstruction data. If a p-value less than 0.05 is taken as significant, then the findings can be summarized with the following scale:

0: Significant at no grid position
1: Significant at some grid positions
2: Significant at most grid positions
3: Highly significant at all grid positions

References

Armstrong, D.M. 1974. Topographical localization in the olivo-cerebellar projection: an electrophysiological study in the cat *J. Comp. Neurol.* 154:287-302.

Brodal, J. 1940.

Chan-Palay, V., Nilaver, G., Palay, S.L., Beinfeld, M.C., Zimmerman, E.A., Wu, J.U. and O'Donohue, T.L. 1981. Chemical heterogeneity in cerebellar Purkinje cells: existence and co-existence of glutamic acid decarboxylase-like and motilin-like immunoreactivities. *Proc. Nat. Acad. Sci.* 78:7787-7791.

Crick, F.H.C. and Lawrence, P.A. 1975. Compartments and polyclones in insect development *Science* 189:340-347.

Diggle, P.J. 1983. Statistical analysis of spatial point patterns. New York: Academic Press,

Goldowitz, D. 1987. Cell partitioning and mixing in the formation of the CNS: analysis of the cortical somatosensory barrels in chimeric mice *Devel Brain Res* 35:1-9.

Grieg-Smith, P. 1952. The use of random and contiguous quadrats in the study of the structure of plant communities. *Ann. Bot.* 16-293-316.

Hawkes, R., Colonneir, M. and Leclerc, N. 1985. Monoclonal antibodies reveal sagittal banding in the rodent cerebellar cortex *Brain Res.* 333:359-365.

Herrup, K. 1986. Cell lineage relationships in the development of the mammalian CNS. III. Role of cell lineage in regulation of Purkinje cell number *Devel. Biol.* 115:148-154.

Herrup, K. 1987. Roles of cell lineage in the developing mammalian brain. In: Neural Development. Hunt, R.K., ed. Academic Press, New York, pp. 65-97

Herrup, K. and Crandall, J.E. 1989. Cell lineage in the mouse somatosensory cortex: analysis of chimeras. *Soc. Neurosci. Abst.* 15:599.

Herrup, K. and Sunter, K. 1986. Lineage dependent and independent control of Purkinje cell number in the mammalian CNS: further quantitative studies of lurcher chimeric mice *Dev. Biol.* 117:417-427.

Jacobson, M. 1985. Clonal analysis and cell lineages of the vertebrate central nervous system *Ann Rev Neurosci* 8:71-102.

Kimmel, C. and Law, R. 1985. Cell lineage of zebrafish blastomeres. III. Clonal analyses of the blastula and gastrula stages *Devel Biol* 108:94-101.

Kimmel, C.B. and Warga, R.M. 1986. Tissue-specific cell lineages originate in the gastrula of the zebrafish *Science* 231:365-368.

Lawrence, P.A. 1981. The cellular basis fo segmentation in insects *Cell* 26:3-10.

Luskin, M., Pearlman, A. and Sanes, J. 1988. Cell lineage in the cerebral cortex of the mouse studied in vivo and in vitro with a recombinant retrovirus *Neuron* 1:635-647.

Mead, R. 1974. A test for spatial pattern at several scales using data from a grid of contiguous quadrats *Biometrics* 30:295-307

Mintz, B. 1962. Formation of genotypically mosaic mouse embryos *Am. Zool.* 2:432.

Mintz, B. 1965. Genetic mosaicsm in adult mice of quadriparental lineage. *Science* 148:1232-1233.

Mullen, R.J. 1977. Site of pcd gene action and Purkinje cell mosaicism in cerebella of chimeric mice *Nature* 270:245-247.

Mullen, R.J. 1978. Mosaicism in the central nervous system of mouse chimeras. The Clonal Basis of Development. In: 36th Symp. Soc. Dev. Biol. Subtelney, S. and Sussex, I., ed. Academic Press, New York, pp. 83-101

Oster-Granite, M.L. and Gearhart, J. 1981. Cell lineage analysis of cerebellar Purkinje cells in mouse chimeras *Devel. Biol.* 85:199-208.

Price, J., Turner, D. and Cepko, C. 1987. Lineage analysis in the vertebrate nervous system by retrovirus-mediated gene transfer *Proc Nat Acad Sci* 84:156-160.

Rakic, P. 1978. Neuronal migration and contact guidance in the primate telencephalon. *Postgrad. Med. J.* 54:25-40.

Rakic, P. 1988. Specification of cerebral cortical areas *Science* 241:170-176.

Ripley, B.D. 1981. Spatial Statistics New York: John Wiley and Sons

Sanes, J. 1989. Analysing cell lineage with a recombinant retrovirus *Trends Neurosci* 12:21-28.

Sanes, J., Rubenstein, J. and Nicolas, J.-F. 1986. Use of a recombinant retrovirus to study post-implantation cell lineage in mouse embryos *EMBO J* 5:3133-3142.

Schmidt, G.H. Wilkinson, M.M. and Ponder, B.J. 1985. Detection and characterization of spatial pattern in chimeric tissue. *J. Embryol. Exp. Morph.* 88:219-230

Schmidt, G.J., Wilkinson, M.M. and Ponder, B.A.J. 1986. Non-random spatial arrangement of clone sizes in chimaeric retinal pigment epithelium *J. Embryol. Exp. Morphol.* 91:197-208.

Selby, S. M., ed. CRC Standard Mathematical Tables, 19th edition, 1971, pp. 611-612, 577-584.

Sidman, R.L., Angevine, J.B. and Taber-Pierce, E.T. 1971. Atlas of the mouse brain and spinal cord. *Abbrev.*

Tarkowski, A.K. 1961. Mouse chimeras developed from fused eggs. *Nature* 230:333-334.

Walsh, C. and Cepko, C. 1988. Clonally related cortical cells show several migration patterns *Science* 241:1342-1345.

Wassef, M. and Sotelo, C. 1984. Asynchrony in the expression of guanosine 3':5' phosphate dependent protein kinase by clusters of Purkinje cells during the perinatal development of rat cerebellum *Neuroscience* 13:1217-1241.

Weisblat, D.A., Sawyer, R.T. and Stent, G.S. 1978. Cell lineage analysis by intracellular injection of a tracer enzyme. *Science* 209:1538-1541.

Wetts, R. and Fraser, S.E. 1988. Multipotent precursors can give rise to all major cell types of the frog retina *Science* 239:1142-1145.

Wetts, R. and Herrup, K. 1982. Cerebellar Purkinje cells are descended from a small number of progenitors committed during early development: quantitative analysis of lurcher chimeric mice *J. Neurosci.* 2:1494-1498.

Wilson 1898. Cell lineage and ancestral reminiscence. In: Biological Lectures from the Marine Biological Laboratories, Woods Hole, Mass. ed. Ginn and Company, pp. 21-42

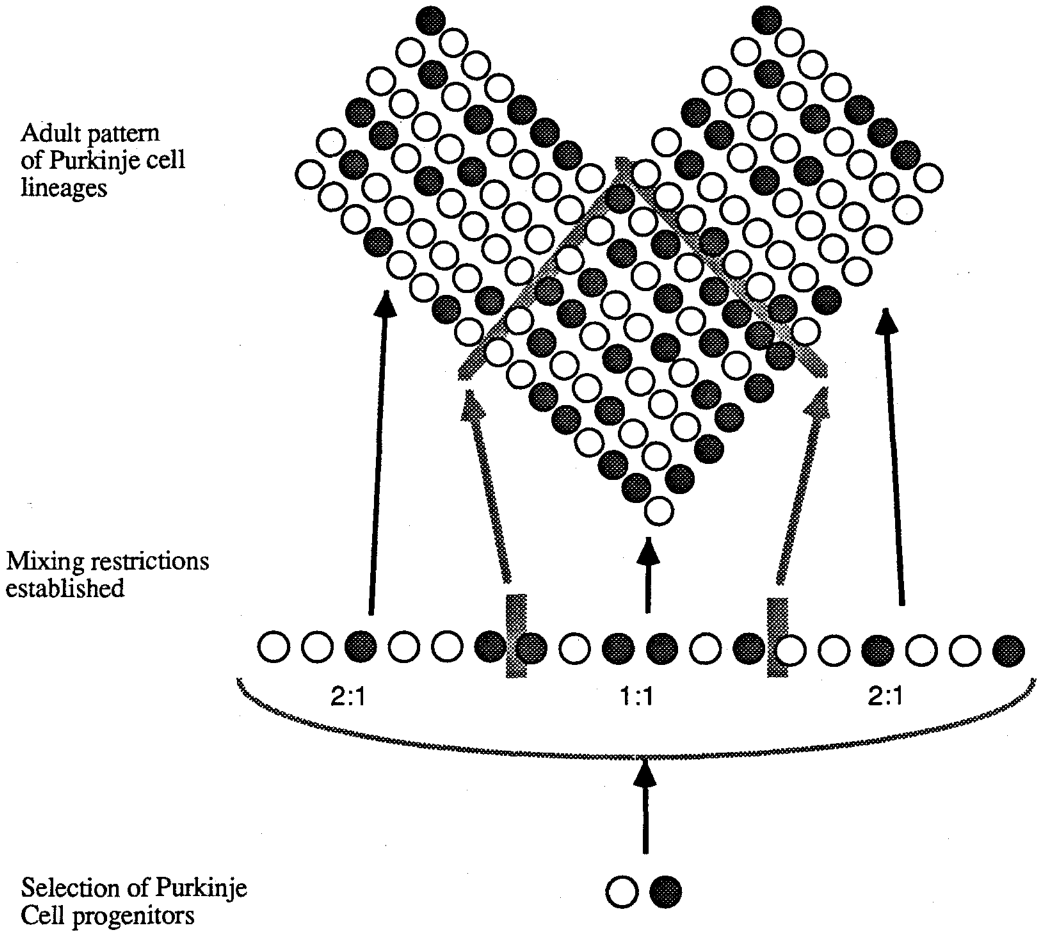
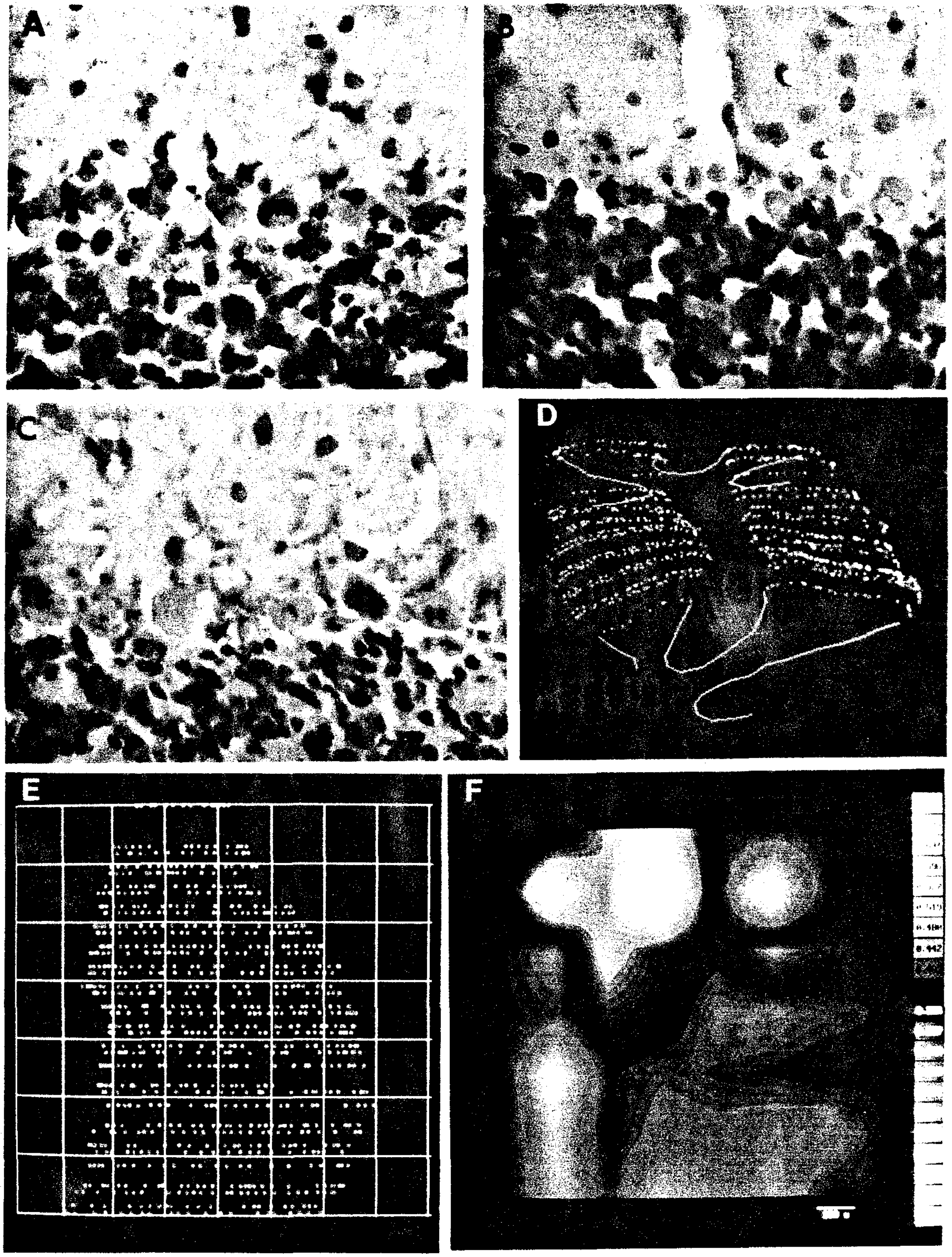

Adult pattern of Purkinje cell lineages
Mixing restrictions established
2:1
1:1
2:1
Selection of Purkinje Cell progenitors

Figure 2.

A schematic representation of one model for the developmental processes that could lead to the lineage patterns found among the cerebellar Purkinje cells. See text for details.


A
B
C
D
E
F
0.442