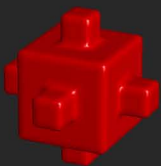
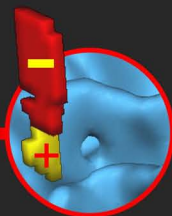
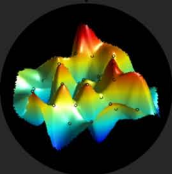


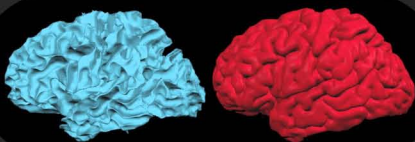
# Functional magnetic resonance **imaging** **object-vision** system



$$(p_2 - p_1) \Sigma_c^{-1} (p_2 - p_1)^T$$



$$\sum_{s,r} p(s,r) \cdot \log \frac{p(s,r)}{p(s) \cdot p(r)}$$





# Functional magnetic resonance imaging of the human object-vision system

---

methodological and empirical contributions

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Maastricht,  
op gezag van de Rector Magnificus, Prof.mr. G.P.M.F. Mols,  
volgens het besluit van het College van Decanen, in het openbaar te verdedigen  
op donderdag 14 oktober 2004 om 14.00 uur

door

Nikolaus Kriegeskorte

**Promotor**

Prof.dr. R. Goebel

**Copromotor**

Dr. E. Formisano

**Beoordelingscommissie**

Prof.dr. G. Kok, voorzitter

Dr. L. Blomert

Dr. B. Jansma

Prof. Dr. W. Hussy, *Universit t zu K ln* (K ln, Duitsland)

Prof. Dr. W. Singer, *Max-Planck-Institut f r Hirnforschung*  
(Frankfurt, Duitsland)

# Contents

|                                                                                    |     |
|------------------------------------------------------------------------------------|-----|
| <b>Introduction</b>                                                                | v   |
| <br>                                                                               |     |
| <b>Chapter 1</b><br>Topologically correct cortex segmentation                      | 1   |
| <br>                                                                               |     |
| <b>Chapter 2</b><br>Cortical object recognition from a visual motion flowfield     | 29  |
| <br>                                                                               |     |
| <b>Chapter 3</b><br>Information-based functional brain mapping                     | 57  |
| <br>                                                                               |     |
| <b>Chapter 4</b><br>Single-object-image response patterns in inferotemporal cortex | 89  |
| <br>                                                                               |     |
| <b>References</b>                                                                  | 107 |
| <br>                                                                               |     |
| <b>Appendix</b>                                                                    |     |
| Summary                                                                            | 117 |
| Samenvatting                                                                       | 123 |
| Acknowledgments                                                                    | 129 |
| Curriculum vitae                                                                   | 130 |
| Contact information                                                                | 130 |



# Introduction

Functional magnetic resonance imaging (fMRI), the empirical technique at the center of this thesis, is a noninvasive method of measuring brain activity in humans and animals. The whole brain or large portions of it can be measured simultaneously with tens of thousands of measurement channels arranged in a regular 3D grid. Each channel represents a cuboid volume element (voxel) of a width of 1-5 mm.

The high spatial resolution of fMRI represents an advantage over the other brain imaging techniques available for human studies. Alternative techniques include positron emission tomography (PET) and electro- and magnetoencephalography (EEG, MEG). In addition to its high spatial resolution, fMRI combines the advantages of depth-pervasive 3D measurement (which it shares with PET) and noninvasiveness (which it shares with EEG and MEG). In contrast to PET, which involves injection of radioactive isotopes, fMRI is not known to be associated with any health risk to the subjects.

Neuronal activity is measured indirectly in fMRI—via the vascular response. A localized increase of neuronal activity entails a localized vascular response, which channels more oxygenated blood into the activated region. The fMRI signal reflects this increase in blood-oxygen level and is therefore temporally delayed (peaking after about 6 s) and blurred with respect to the underlying neuronal activity. This limits the spatial and temporal detail, at which brain activity can be observed.

Because of its combination of depth-pervasive whole-brain coverage, good spatial resolution, and noninvasiveness, fMRI has revolutionized the study of human brain function. Compared to the classical technique of invasive electrophysiology, which is mainly used in animal research, fMRI has an obvious disadvantage: its much lower resolution in both space and time. However, its advantage consists in the spatial coverage (up to whole-brain coverage) and the tremendous number of channels that can be recorded simultaneously. fMRI is therefore becoming increasingly important also in studying the monkey brain.

## **Illuminating the black box**

The contrast mechanism of blood-oxygen-level-dependent fMRI (the dominant technique, which is also used in these studies) has only been discovered in 1990. In the short period of its existence, fMRI has played a central role in the formation of the transdisciplinary field of cognitive neuroscience. Because it provides a means to study human brain activity at the systems level without any known

health risks, it is naturally suited to illuminate the black box, whose contents have been described in the abstract through decades of research in cognitive psychology. An obvious application, then, is the localization of brain structures subserving the cognitive component functions postulated by cognitive psychology on the basis of purely behavioral evidence.

The field of human brain mapping attempts to characterize the global architecture of the human brain by assigning functional labels to macroscopic brain regions. This research paradigm goes back to phrenology, but it has empirically proven itself. Brain imaging has demonstrated that the global patterns of activity associated with a wide range of cognitive processes are consistent across individual brains. Like the organs of the body, different brain regions serve specific functions. The gross functional parcellation is largely uncontroversial for perceptual and motor subsystems.

### **Current challenges**

It has yet to be established empirically to what extent different parts of cortex conform to the attractive model of functional modules. Is all of cortex—like early visual cortex—subdivided into crisply demarcated macroscopic areas that exist in every individual brain and serve clearly defined functions that can be characterized by one-word functional labels? For frontal-lobe higher cognitive functions, for example, the spatial layout has been much less precisely defined than for the occipital-lobe visual functions. Despite vigorous research, it remains unclear to what level of spatial precision the functional layout of the frontal lobes is consistent across individuals.

Understanding brain function will take more than a global functional map. Current fMRI research goes beyond mere functional labeling of macroscopic brain regions based on group-averaged data. Functional regions are defined in individual subjects. Their activity levels during different cognitive processes are compared. The time course of the activity and functional interactions between regions are investigated. Based on anatomical magnetic-resonance-imaging data, the folded cortex of each subject is reconstructed as a two-dimensional manifold. Such individual cortical-surface representations provide constraints for advanced methods of statistical analysis of functional data and allow visualization of activity patterns on flattened cortical maps. These methodological developments allow current fMRI research to characterize more precisely the distributed computational processes occurring in the brain.

### **This thesis**

Motivated by questions about visual object recognition, this thesis combines methodological and empirical contributions. It presents two new methods: (1) a cortex-segmentation method that takes the known topology of the cortical sheet into account and (2) a statistical analysis method that utilizes the information contained in fine-grained fMRI activity patterns. These methods are applied in two empirical studies. The first study, taking the global view, addresses the interaction between the dorsal and the ventral visual stream—two subsystems often conceptualized as largely independent. The second study, zooming in on fine-grained activity patterns, investigates the distributed representation of object images in ventral temporal cortex. We apply method (2) to study how partic-

ular exemplars of an object category (e.g. particular faces) are represented in fine-grained response patterns that may be unique to each individual brain.

As a whole, the thesis demonstrates the advantages of developing theory, experiment, and statistical analysis methods in close interaction. This interaction is in evidence in the first half (Chapters 1 and 2), but even closer in the second (Chapters 3 and 4). In the second half, a theoretical prediction (the existence of macropopulation codes for object exemplars) motivates the development of a new analysis technique (information-based functional brain mapping). In conjunction with a novel experimental approach (the single-object-image approach), the analysis technique leads to the challenge of a widespread assumption (that the fusiform face region represents faces at the individual level) and suggests the existence of a functional region that has not previously been described (the anterior inferotemporal face-exemplar region).

## Overview

This thesis combines methodological developments (Chapters 1 and 3) and empirical investigations of human object vision (Chapters 2 and 4). The methods developed concern the analysis of anatomical (Chapter 1) and functional (Chapter 3) magnetic-resonance-imaging data and are employed in the empirical investigations. The chapters can be read in isolation as each is self-contained and corresponds to a paper either published (Chapters 1 and 2), submitted (Chapter 4), or in preparation (Chapter 3).

Chapter 1 presents a method revealing the spatial structure of the cortex in anatomical magnetic resonance volumes. An explicit representation of the cortical sheet of each subject allows brain activity to be visualized on a cortical map and is prerequisite to a number of advanced methods of statistical analysis of fMRI data. Our approach avoids the topological errors produced by established methods.

Chapter 2 presents an investigation of object recognition based on perception of a visual motion flowfield. Object recognition is known to involve the ventral stream of visual cortex. The dorsal stream, in contrast, processes information about object motion and location. In the natural world, objects are often in motion relative to the observer, so both systems are usually coactivated. In this study, we use structure-from-motion stimuli to address how the dorsal and the ventral visual subsystem interact in object recognition.

Chapter 3 presents information-based functional brain mapping, a novel statistical method, which finds brain regions whose fine-grained spatial response pattern contains information about the experimental condition. The information-based approach is in contrast to conventional activation-based brain mapping, which is best suited for finding regions whose spatially averaged activity is greater in one condition as compared to a control condition.

Chapter 4 explores the representation of visual objects in ventral temporal cortex. In contrast to previous studies, we do not average responses to different images of the same object category, but estimate the spatial response patterns elicited by four particular images (two faces and two houses). This allows us to address how exemplars within the same category are distinguished. The study is crucially dependent on the method of information-based brain mapping introduced in the previous chapter.

The Summary (in English and Dutch in the Appendix) provides an accessible guided tour of the thesis. It introduces the minimal background required to motivate each project and explains the relationships between the projects at a conceptual level. More technical detail is provided in each chapter's abstract.

# Chapter 1

## Topologically correct cortex segmentation

**Corresponding publication:** Kriegeskorte N and Goebel R. An efficient algorithm for topologically correct segmentation of the cortical sheet in anatomical MR volumes. *NeuroImage*, 14(2):329-46, 2001.

**Abstract:** Polygon-mesh representations of the cortex of individual subjects are of anatomical interest, aid visualization of functional imaging data, and provide important constraints for their statistical analysis. Due to noise and partial volume sampling, however, conventional segmentation methods rarely yield a voxel object whose outer boundary represents the folded cortical sheet without topological errors. These errors, called handles, have particularly deleterious effects when the polygon mesh constructed from the segmented voxel representation is inflated or flattened. So far handles had to be removed by cumbersome manual editing, or the computationally more expensive method of reconstruction by morphing had to be used, incorporating the a-priori constraint of simple topology into the polygon mesh model.

Here we describe a linear-time-complexity algorithm that automatically detects and removes handles in presegmentations of the cortex obtained by conventional methods. The algorithm's modifications reflect the true structure of the cortical sheet. The core component of our method is a region growing process that starts deep inside the white-matter voxel object, is prioritized by the distance-to-surface of the voxels considered for inclusion, and is selftouching-sensitive, i.e. voxels whose inclusion would add a handle are never included. The result is a binary voxel object identical to the initial object except for a number of small "cuts". Each cut is located in the thinnest part of a handle. By applying the same method to the inverse object, an alternative set of solutions is determined. Each cut of the inverse object constitutes a "filling" for one of the handles. For each handle separately, the solution more consistent with the intensities of the original anatomical MR scan (cut or filling) is chosen. The accuracy of the resulting polygon-mesh reconstructions has been validated by visual inspection, by quantitative comparison to an expert's manual corrections, and by crossvalidation between reconstructions from different scans of the same subject's cortex.

## Introduction

The human cerebral cortex is a convoluted sheet of varying thickness whose two-dimensional macrostructure early anatomists have recognized long before its layered-microstructure—its third dimension—was discovered. Because of its 2D macrostructure, the cortical sheet lends itself to representation in a single planar map for each hemisphere. Such maps were first produced manually from post-mortem brains by rather coarse techniques.

Today a representation of the cortex of a living subject can be unfolded, flattened, and used as a map by computational methods. To this end, the cortical sheet is reconstructed as a polygon mesh from anatomical magnetic resonance (MR) imaging data (Carman, et al., 1995). These techniques are useful not only to the neuroanatomist studying the physical structure of the cortex, but also to neuroscientists primarily interested in cortical function (Van Essen and Drury, 1997; Van Essen et al., 1998).

A polygon-mesh representation of the structure of a subject's cortex allows, for instance, the visualization of functional imaging data on a two-dimensional flatmap (Serenio et al., 1995). Taking advantage of the inherent structure of the cortex, surface-based representations are concise in that they assemble in a single image information otherwise spread over a series of slice images, whose detailed spatial relations are cumbersome to understand. Despite a certain degree of inevitable geometric distortion, surface-based representations give an idea of the intracortical—as opposed to the Euclidean—distance between locations.

But beyond mere visualization, an accurate polygon-mesh representation of an individual subject's uniquely folded cortex provides an important constraint for statistical analysis of functional data from that subject. On the simplest level, the analysis can be restricted to the region within the functional data set that represents the cortex (Goebel and Singer, 1999). More complex applications of surface representations include intraarea mapping, e.g. the computation retinotopic maps of early visual areas (Serenio et al., 1995; Linden et al., 1999; Formisano et al., 2003), and the use of anatomically informed basis functions (Kiebel et al., 2000) for computation of functional maps in general.

The computational methods used to obtain a polygon mesh that accurately represents the cortex can be grouped according to the type of representation they operate on. On one hand, there are segmentation methods operating on the anatomical MR volume in voxel space, on the other hand, there are iterative morphing methods operating on polygon-mesh representations.

Dale et al. (1999) take the approach of segmenting the white matter, thus recovering the inner boundary of the cortical gray matter in voxel space. The initial polygon-mesh representation of the cortex is computed by exact recreation of the boundary of the binary voxel object produced by segmentation. Slight smoothing of this initial polygon mesh yields a representation of the cortical sheet that is geometrically very accurate, but typically contains a number of topological errors, called handles.

Handles are toroidal structures of the polygon mesh. They are incompatible with what is known about cortical anatomy. The cortex is a folded sheet, i.e. the pial surface may touch itself across a sulcus but the gray matter is nowhere grown

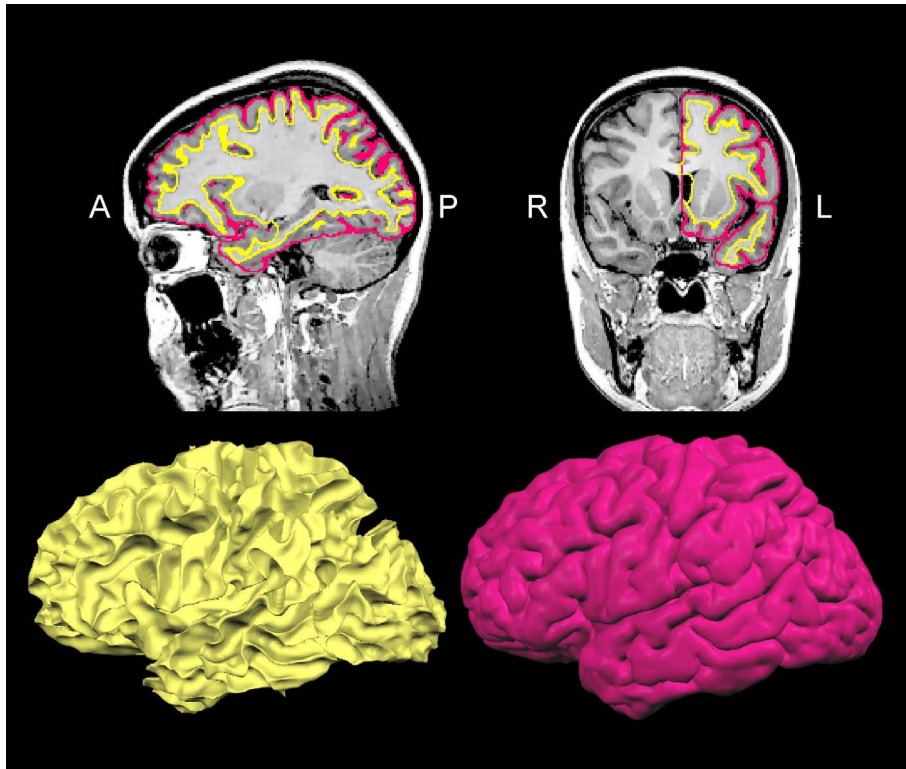


Figure 1.1: **Precise polygon-mesh representations of the inner and the outer boundary of the gray matter**

Our algorithm was used to obtain a topologically correct initial polygon-mesh representation of the cortical sheet. Polygon meshes (bottom) representing the inner (yellow) and outer (magenta) boundaries of the gray matter have been computed by morphing the initial reconstruction. Projections of the surfaces back into the original anatomical MR volume (top) demonstrate the accuracy and precision of the representation of the two gray matter boundaries.

together through the pia mater. The reconstructed boundary of the voxel object should therefore ideally be a simple polyhedron, i.e. it should be without handles. Handles result from noise and partial volume sampling of the anatomical MR volume. Because the cortex has a simple topology, every handle represents an error in the reconstruction of the cortical sheet. But beyond local inaccuracy of surface representation, handles have particularly deleterious effects—causing large geometric distortions—when the polygon mesh is inflated (Fischl et al., 1999) or flattened. In the approach of Dale et al. (1999), handles have to be removed by manual intervention, which typically requires about 30 minutes per cortical hemisphere.

MacDonald et al. (2000) take the contrasting approach of recovering the structure of the cortical sheet by iterative morphing of an initial standard polygon-mesh model. The model includes both the inner and the outer boundary of the gray matter. This top-down, model-based approach lends itself naturally to the incorporation of a-priori constraints, including simple topology of the surfaces. MacDonald et al. (2000) overcome the problem of geometric inaccuracy tradi-

tionally associated with iterative morphing methods: their reconstructions accurately represent the inner and the outer boundary of the gray matter. Whereas the resulting representation of the cortex is both topologically and geometrically correct, the extensive morphing of the initial standard model is extremely expensive computationally.

Here we describe an algorithm that detects and removes handles as part of the segmentation in voxel space. In combination with established methods, this algorithm allows topologically and geometrically accurate cortex reconstructions to be obtained at much lower computational costs.

First, we presegment the white matter automatically by methods similar to those described in Dale et al. (1999). Then we correct topological errors in the voxel representation using our novel algorithm whose core element is a distance-to-surface prioritized, selftouching-sensitive region growing process. This allows us to obtain a topologically correct initial polygon-mesh representation of an individual subject's uniquely folded cortex by a fast standard one-pass reconstruction method. Finally, the initial polygon-mesh reconstruction is subjected to morphing to obtain precise polygon-mesh representations of the inner and outer boundaries of the gray matter. In contrast to the approach of MacDonald et al. (2000), only very few iterations are required and simpler morphing methods suffice since the initial reconstruction already quite accurately represents the cortical sheet.

Our topology correction algorithm can be implemented such that its time complexity is linear in the number of input voxels. With anatomical MR datasets of 1 mm isotropic resolution, the complete process from an anatomical MR volume in Talairach space to polygon meshes representing the inner and the outer boundary of both hemispheres (Figure 1.1) takes less than 15 minutes on a current PC workstation (1000 MHz dual Pentium).

## Method

### Presegmentation

The topology correction algorithm described in this paper can be applied to any binary presegmentation of the white matter that contains topological defects. As most of the white matter presegmentation methods we use have previously been described (see citations below), we only give a brief description of our presegmentation procedure here.

White matter presegmentation is performed by applying a sequence of operations to a 3D T1-weighted anatomical data set. In our approach, the anatomical MR volume is first projected into Talairach space (Talairach and Tournoux, 1988; for our method see Linden et al., 1999). This is the only step of our procedure currently requiring user interaction. The automatic presegmentation process first applies standard masks in Talairach space to remove the cranium and the cerebellum. This step preserves the intensity information of the original anatomical volume in the segmented region including the cortex. The deletion of voxels representing the cranium and the cerebellum is not yet clean at this stage:

bits and pieces representing these structures remain in the volume. The mid-brain structures and the ventricles are solidly filled in a pseudocolor. Whereas the midbrain structures are filled using a standard Talairach mask, filling of the ventricles requires a more adaptive approach, if the representation of the cortical sheet in this region is to be preserved. The ventricles are therefore filled by a region growing process seeded at points automatically located by analysis of the intensity values in the ventricular region.

Potential spatial variation of the white matter signal intensity is corrected for by fitting Legendre polynomials to estimate a three-dimensional bias field (inhomogeneity correction, see Vaughan et al., 2001). In the intensity histogram computed over the remaining voxels within the Talairach proportional grid, the two peaks corresponding to white and gray matter are identified. The intensity difference between the two histogram peaks is used to control the inclusion range parameter of a sigma filter (Lee, 1983), which performs nonlinear, edge-preserving smoothing, reducing the noise within white and gray matter without smoothing the white-gray matter boundary. The intensity values at the two histogram peaks determine the intensity range, within which a region growing process is performed to segment the white matter. The region growing process, seeded automatically in the white matter of one hemisphere above the corpus callosum, cleanly segments the entire white matter discarding isolated pieces of cranium and cerebellum that have not previously been removed.

Segmenting along the boundary between white and gray matter rather than along that between gray matter and cerebrospinal fluid yields a more accurate initial voxel space representation of the structure of the cortical sheet (Dale et al., 1999). The binary object obtained by region growing is subjected to slight binary-voxel-object smoothing (cf. Dale et al., 1999), which implements the a-priori constraint of finite curvature of the cortical sheet. Finally, the two hemispheres are disconnected automatically. Below the corpus callosum, the boundary of the presegmented voxel object follows the solidly filled region masking the midbrain structures and the ventricles (henceforth referred to as the “sub-callosal mask”), resulting in a flat representation of the medial plane within and below the corpus callosum. Everywhere else the boundary of the presegmented voxel object follows the cortical sheet—including the deep sulcal structure—along the boundary between gray and white matter. All subsequent operations are invoked separately for the two presegmented hemispheres.

At this point, the topology correction algorithm is applied to the presegmented voxel representation of each hemisphere. The boundary of the voxel object representing the cortical sheet is reconstructed by a simple tessellation method, which exactly recreates the cubic voxel structure in the polygon-mesh representation. The resulting polygon mesh is morphed to represent either the inner or the outer boundary of the gray matter by shifting each vertex along its surface normal until its position coincides with the respective intensity contour in the anatomical MR scan. The final result is shown in Figure 1.1.

## Rationale of the algorithm

The core of the topology correction algorithm we propose is a region growing process (Figure 1.2, A) that starts deep inside the binary voxel object obtained

by the presegmentation as described above. As the region grows, it eventually comes to represent the object as a whole. During the growth process the region is never allowed to selftouch (i.e. to form rings, definition below), ensuring that its boundary at every point of the process is a simple polyhedron (Figure 1.2, B). The growing is seeded at a point within the object that is maximally distant from the object's surface. Among the fringe voxels poised to be included into the region, a voxel maximally distant from the object surface is chosen at each step. This ensures that the region grows into the thinner parts of the object, where all voxels are closer to the surface, last (Figure 1.2, C).

Let us assume for the moment that the handle in question resulted from erroneous inclusion of voxels into the object during the presegmentation. The handle is part of a closed ring, which may be cut at any location to render the resulting object ringless (i.e. its outer boundary a simple polyhedron). Where should the cut be placed? The algorithm implies a heuristic for identifying the erroneously included voxels according to which they form the thinnest part of the ring (see Formal definition of the algorithm and Discussion). This is in line with the idea that the modifications of the object should be kept to a minimum. The algorithm, however, does not strictly guarantee that the number of voxels changed is minimal (see Discussion). Since the region grows into the thinnest parts last, the region growing will terminate before the region selftouches in the thinnest part of the ring, excluding a small number of voxels (Figure 1.2, C).

The prioritization of the process has an intuitive physical analogy if we think of the object as a two-dimensional shape, its contour corresponding to the surface of a 3D object. Imagine a flat landscape with a pool, the shape of the 2D object. The two horizontal dimensions represent the object space. The third, vertical dimension is the depth of the pool (axis pointing downward), which represents the prioritization criterion distance-to-surface. While the landscape is flat outside the contour of the object, inward from the contour it declines at 45 degrees, forming the pool, whose depth at every point equals the distance to the contour in the 2D object (the distance-to-surface).

The prioritized region growing process is analogous to the pouring of a fluid into the pool from above the lowest point. The fluid will form a small pond, its surface rising in the pool until it reaches the level where the fluid can spill over into a second pond. This marks the beginning of a second phase, in which the fluid trickles down toward the lowest point of the second pond. The surface then rises in the second pond while it stays constant in the first. The fluid will wet one pond after another and at any point the surface will only rise in the pond where it is lowest. It will rise again in the first pond only after all other wet ponds have reached the surface level of the first pond. Eventually the fluid will fill the pool completely, reaching the level of the surrounding flat land.

The only aspect of the region growing process not covered by this analogy is the selftouching sensitivity: the region (the fluid) is never allowed to selftouch, that is its outer boundary may expand but not merge with itself to form rings.

The part of the algorithm described so far already constitutes a method of removing all handles: when the region growing terminates, the region is identical to the object except for a few missing voxels, which effectively cut all handles. Deletion of voxels seemed an appropriate correction method for the case of erroneously included voxels. Topological errors, however, can also result from

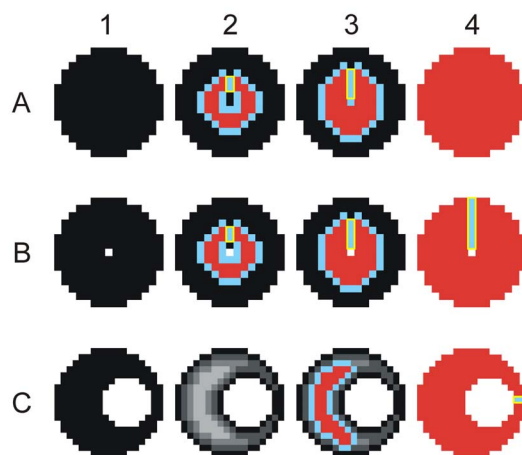


Figure 1.2: **Selftouching sensitivity and distance-to-surface prioritization**

For three different 2D objects (rows A, B, C), four snapshots of the process are shown in chronological order (columns 1-4). To stress the analogy to the 3D process, pixels are referred to as “voxels” and the outer contours of the objects as the “surface” in the following. Object voxels are black or gray. Region voxels are superimposed onto the object in red, fringe voxels in blue. Fringe voxels the region selftouches in, which are not eligible for inclusion into the region in the next step, are outlined in yellow.

(A) Simple disk (1): During the process (2, 3) the region may, by chance, come to selftouch temporarily in some fringe voxels (yellow rectangles). Note that the region in 3, does not selftouch in the fringe voxel immediately below the yellow rectangle. When this voxel is included into the region, the region no longer selftouches in the fringe voxel above the included voxel. Eventually the region comes to include all voxels (4)

(B) Ring: The object (1) differs from the previous one (A) only in that its central voxel is vacant, making it a ring. The region growing terminates (4) while there are still object voxels left that haven’t been included into the region. The excluded voxels (yellow rectangle) are all fringe voxels the region selftouches in. They represent an arbitrary cut through the ring.

(C) Ring of non-constant thickness: Here, the sequence of snapshots (1-4) demonstrates the distance-to-surface prioritization of the region growing. In this 2D example, the contours of the object (1) represent the surface. Each voxel’s distance-to-surface is indicated by its shade of gray (2): the brighter the voxel, the greater its distance-to-surface. The region growing starts in a voxel of maximal distance-to-surface (deep inside the object) and proceeds to voxels closer to the surface (3), until it reaches the thinnest part of the ring. There the region growing terminates (4), because there are no more fringe voxels the region does not selftouch in.

erroneous exclusion of voxels. Such errors can be thought of as “holes”, though from the vantage point of topology each hole implies handle and vice versa, rendering the two terms equivalent. If we assume the errors of the original segmentation to be random inversions of voxels, a method that can only delete voxels to render the object’s boundary a simple polyhedron seems incomplete.

Each topological error can be removed either by cutting the handle (deleting voxels) or by filling the hole (adding voxels). Assume we chose to remove an error by filling, which voxels should be added to the object? By the same logic applied above, it seems sensible to add a minimum number of voxels. Since each handle of the object corresponds to exactly one handle of the inverse object, we can apply the procedure described above to the inverse object. We will identify a set of voxels whose addition cuts the corresponding handle in the inverse object.

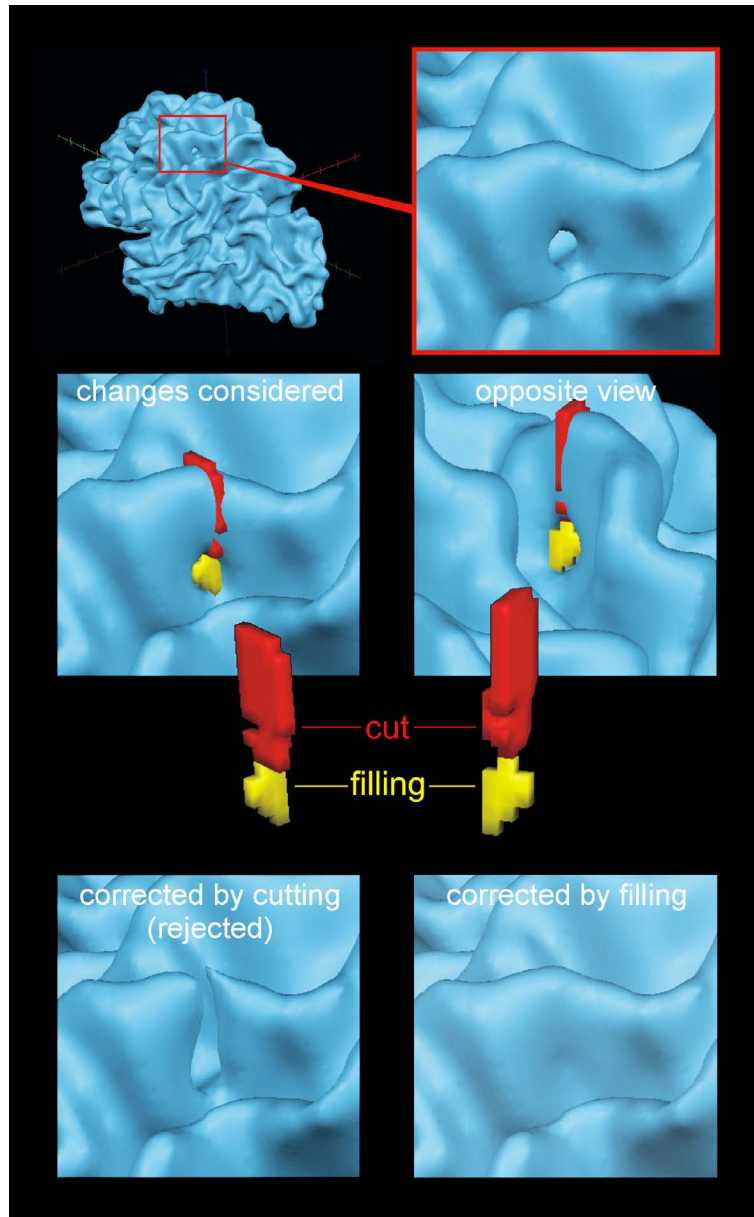


Figure 1.3: **Cutting the handle versus filling the hole**

Topological errors can result from erroneous inclusion or exclusion of voxels in the binary segmentation. The topological correction can therefore consist in removing voxels (cutting the handle, red) or adding voxels (filling the hole, yellow). The figure shows both solutions as computed by distance-to-surface prioritized, selftouching-sensitive region growing. The algorithm chooses the better solution for each topological defect separately by a heuristic (Figure 1.5). In this case, it has chosen to fill the hole.

Though the algorithm operates in voxel space, surface renderings have been used to visualize the intermediate and final results. Whereas for the cut (red) and the filling (yellow) the surfaces reflect the cubic shape of the voxels, the polygon meshes representing the pre- and postsegmentation (blue) have been smoothed to ease visual orientation.

If we add this set of voxels to the original object we have corrected the handle by filling it. The proposed algorithm computes both solutions and chooses, for each handle separately, the solution that does less damage (defined below) to the segmentation (Figure 1.3).

## Formal definition of the algorithm

The term “object” in the following refers to the binary presegmentation of a cortical hemisphere along the boundary between gray and white matter as described above.

### Distance-to-surface mapping

The distance-to-surface of a point within the object, ideally, is the length of the shortest line to the surface of the object. We estimate this value for every voxel by counting the number of times the object needs to be eroded until the voxel is removed (Figure 1.4, A). Let  $\text{Obj}$  be the set of voxels representing the object. For each object voxel  $v$ , the distance-to-surface function  $d(v)$  is computed as follows:

$$\begin{aligned}
 d(v) &= \begin{cases} -1 & \text{if } v \in \text{Obj} \\ 0 & \text{otherwise} \end{cases} \\
 cd &= 0 \\
 \text{while}(\exists \text{ voxel } u \in \text{Obj} \mid d(u) = -1) \\
 &\quad cd = cd + 1 \\
 &\quad d(v) = \begin{cases} cd & \text{if } d(v) = -1 \wedge (\exists \text{ voxel } w \mid \text{side-} \\ & \quad \text{adjacent}(v, w) \wedge d(w) = cd - 1) \\ d(v) & \text{otherwise} \end{cases} \\
 \text{endwhile,}
 \end{aligned}$$

where  $cd$  is the distance-to-surface currently being mapped and  $\text{side-adjacent}(v, w)$  means that voxels  $v$  and  $w$  share a side. By analogy  $\text{point-adjacent}(v, w)$  means that they share at least one point and  $\text{edge-adjacent}(v, w)$  means that they share at least one edge (Figure 1.4, B).

### Selftouching sensitive, distance to surface prioritized region growing

Let  $\text{Obj}$  be the set of voxels representing the object,  $R$  the set representing the growing region (initially empty), and  $F$  the set of fringe voxels poised to be included into the growing region (Figure 1.2). The fringe  $F$  is initially seeded with a single object voxel maximally distant from the surface. The region growing process representing the core of our algorithm can be expressed as follows:

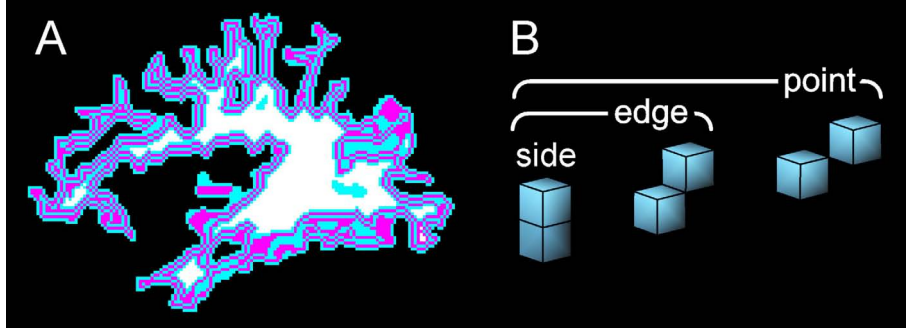


Figure 1.4: **Distance-to-surface prioritization and adjacency relations between two voxels**

A Distance-to-surface prioritization: Odd priority levels are shown in blue and even ones in magenta. Larger regions of one priority level occur where the 3D surface obliquely intersects the sagittal slice shown. The core of the object (white) does not need to be prioritized since its boundary already has a simple topology (see footnote 2, Implementation and time complexity). B Adjacency relations between two voxels: Two voxel are side-adjacent if they share a face, point-adjacent if they share at least one vertex, and edge-adjacent if they share at least one edge.

```

R = { }
F = {voxel u}, u ∈ Obj, d(u) = max
while(∃ voxel v ∈ F | d(v) = max in F ∧ ¬ selftouching(R, v))
    F = F \ {v}
    R = R ∪ {v}
    F = F ∪ {voxels w | side-adjacent(v, w) ∧ w ∈ Obj \ R}
endwhile.

```

Selftouching( $R, v$ ) means the region  $R$  growing inside the object  $Obj$  selftouches in a voxel  $v \in Obj$ . The function *selftouching*( $R, v$ ) is true if and only if  $v$  connects two region voxels point-adjacent to  $v$  that were not already connected within the neighborhood of  $v$ :

$$\begin{aligned}
 \text{selftouching}(R, v) \Leftrightarrow \\
 \exists (u, w) | u, w \in \text{neighborhood}(v) \cap R: \\
 \neg \exists \text{ point-adjacency path set } P \subseteq \text{neighborhood}(v) \cap R | u, w \in P,
 \end{aligned}$$

where *neighborhood*( $v$ ) is the set of all voxels point-adjacent to  $v$  excluding  $v$  itself and a point-adjacency path set is a set of voxels  $P = x_1, x_2, \dots, x_n$  such that  $x_i$  and  $x_{i+1}$  are point-adjacent for all  $i = 1 \dots n - 1$ . For verbal simplicity, a voxel the region selftouches in will be referred to as a selftouching voxel in the following.

Since selftouching fringe voxels are never included into the region, the region's outer boundary remains a simple polyhedron until the process terminates. If the genus of the object's boundary was greater than 0, i.e. if there were handles,

there will be a non-empty set  $C = Obj \ R$  of cut voxels.  $C$  is the union of disjoint sets  $C_1, C_2, \dots, C_k$ , each of which represents a single cut, i.e. a *contiguous* set of cut voxels:

$$\begin{aligned} \forall i = 1 \dots k: \\ \forall (v, w) | v, w \in C_i: \\ \exists \text{ point-adjacency path set } P \subseteq C_i | v, w \in P. \end{aligned}$$

### Region growing in the inverse object

The selftouching-sensitive, distance-to-surface prioritized region growing process described above is also applied to the inverse object  $Obj' = U \setminus Obj$ , where  $U$  is the set of all voxels. To ensure that the region can grow around the object, the object must not touch the outer boundary of the block of all voxels, i.e. the outer boundary of  $U$  must not contain points belonging to the object. Invoking the procedure described above substituting  $Obj'$  for  $Obj$ , yields a set of voxels not included when the region growing terminates. We will call it the inverse or negative cut set  $C'$ , which is decomposable into disjoint contiguous sets  $C'_1, C'_2, \dots, C'_l$ , each corresponding to a negative cut as described above for the positive cuts  $C_1, C_2, \dots, C_k$ .

### Filling or cutting: choosing the better solution for each handle

Simple topology could be achieved by performing either all positive or all negative cuts. Our algorithm, however, is locally adaptive in its choice of a combination of cuts that renders the topology simple. If we were forced to sketch the choice of cut combination in three sentences, they would read as follows: For each handle, there is a positive and a corresponding negative cut; either of them removes the handle. Corresponding cuts can easily be identified because they touch. For each handle independently, the cut doing less damage to the presegmentation is chosen.

Unfortunately, matters are more complicated as, first, more than two cuts can form a set whose elements interact in their effects on the topology and, second, the number of positive cuts can be different from the number of negative cuts, because the cuts are not ideal two-dimensional entities, but sets of voxels of non-zero volume.

Concerning the first point, sets of interacting cuts can be identified by their property of spatial linkage. For ideal two-dimensional cuts of zero volume, linkage between two cuts means that the cuts share a point. The elements of a set of interacting cuts can be thought of as the nodes of a graph whose edges represent linkage. The resulting graphs are trees. All cuts together form a forest. Cuts belonging to different trees of the forest have independent effects on topology, so each tree can be treated independently of the others. There is a simple rule for determining which combinations of cuts of a tree can be chosen, such that if, for each tree of the forest separately, one of these combinations is chosen, the genus of the object will be zero.

We do not use this rule, however, as, first, the fact that cuts are not ideal but have a non-zero volume further complicates matters and, second, in the domain of cortex segmentation from anatomical MR volumes, we have rarely encountered these more complex cases and, where we did encounter them, it was in the region of the subcallosal mask, where the boundary of the voxel object does not represent the cortical sheet. Instead, the present version of our algorithm chooses, for each set of interacting cuts separately, to make either all positive or all negative cuts. The loss of local adaptivity this entails is negligible, because almost all sets of interacting cuts contain exactly one positive and one negative cut.

First, we group all positive and negative cuts into sets of interacting cuts, which we call choices. A choice  $E$ , thus, is a set of cuts. For cuts as defined here (contiguous voxel sets of non-zero volume), the property of linkage, allowing us to identify the choices, is given if cuts are linked by a side-adjacency. Let  $S$  be the set of choices, initially containing, for each positive and each negative cut, a degenerate choice composed of only that cut. We can compute the choices by repeatedly merging sets of cuts linked by side-adjacency, as follows:

```

set of choices  $S$ 
  = {choice  $E$  |  $E = \{D\}$ ,  $D \in \{C_1, C_2, \dots, C_k\} \cup \{C'_1, C'_2, \dots, C'_l\}$ }
while ( $\exists$  voxels  $v, w$  |  $v \in \text{cut } A \in \text{choice } X \in S$ ,
       $w \in \text{cut } B \in \text{choice } Y \in S$ ,
       $X \neq Y$ , side-adjacent( $v, w$ ))
   $X = X \cup Y$ 
   $S = S \setminus \{Y\}$ 
endwhile.
```

As defined in the previous section,  $C_1, C_2, \dots, C_k$  is the set of positive and  $C'_1, C'_2, \dots, C'_l$  the set of negative cuts. When this procedure terminates,  $S$  is the set of all choices and each choice in  $S$  is a non-empty set of positive and negative cuts. Most choices will contain exactly one positive and exactly one negative cut, but the number of cuts in a choice set can occasionally be larger than two and the number of positive cuts need not be equal to the number of negative cuts.

To decide, for each choice separately, whether the positive or the negative cuts should be made, we estimate the damage done by each of the two solutions in terms of misclassification of voxels in the final segmentation and choose the solution doing less damage. The misclassification damage is estimated by the following heuristic: The damage done by a set of cuts is defined as the sum of the damages done by the cuts and the damage done by a cut is defined as the sum of the damages done by the voxels the cut comprises.

$$\begin{aligned}
\text{deletion damage}(\text{choice } E) &= \sum_{\text{cuts } D \in E} \sum_{D \in \{C_1, C_2, \dots, C_k\}} \text{damage}(\text{cut } D) \\
\text{addition damage}(\text{choice } E) &= \sum_{\text{cuts } D \in E} \sum_{D \in \{C'_1, C'_2, \dots, C'_l\}} \text{damage}(\text{cut } D) \\
\text{damage}(\text{cut } D) &= \sum_{\text{voxels } v \in D} \text{damage}(\text{voxel } v)
\end{aligned}$$

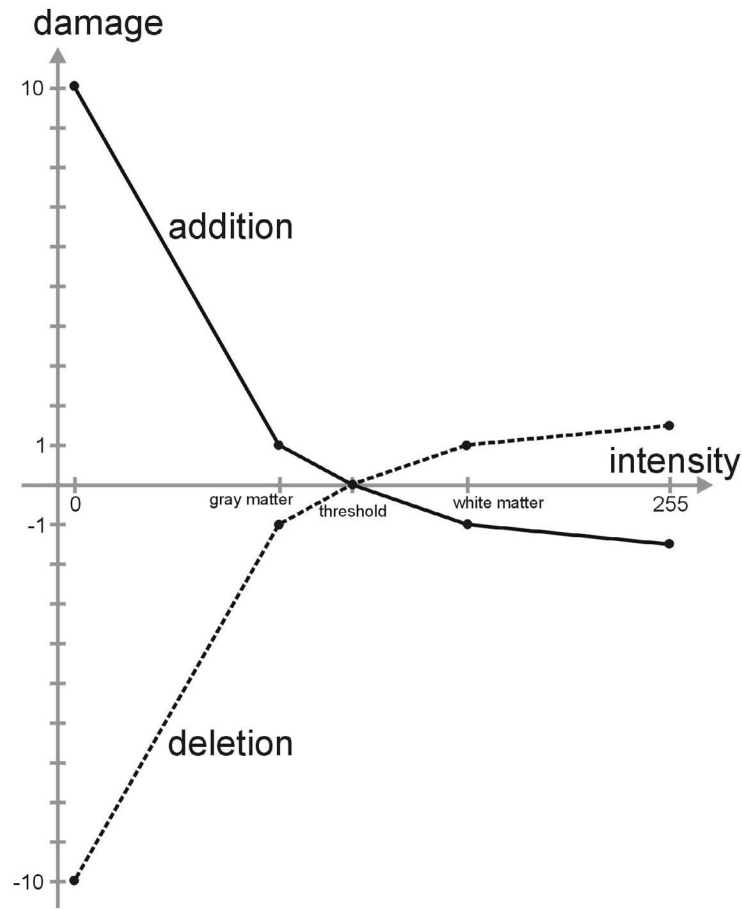


Figure 1.5: **Heuristic misclassification-damage function**

The function provides a heuristic estimate of the misclassification damage caused by inverting a voxel in the presegmentation. In order to choose between corresponding solutions (cutting and filling, Figure 1.3) for each locus of topological error, the damage each solution would do to the presegmentation is estimated by summing the misclassification damages of the inverted voxels.

The misclassification damage of a voxel depends on whether the voxel is to be deleted (positive cut) or added (negative cut), on the intensity of the voxel in question, the intensity of prototypical gray and white matter voxels in the original unsegmented MR volume (as identified before the presegmentation by histogram analysis), and on the threshold intensity used in the presegmentation.

Deletion (positive cutting) means that a voxel classified as white matter in the presegmentation is reclassified as an outside-white-matter voxel, and vice versa for addition (negative cutting). Ideally, voxels whose intensity is above threshold should be classified as white matter, and voxels whose intensity is below threshold should be classified as outside-white-matter. (This is not strictly the case after presegmentation, because in the presegmentation thresholding is followed by binary-voxel-object smoothing.)

The rationale of the heuristic used to estimate the misclassification damage caused by inverting a voxel (Figure 1.5) is as follows: If a voxel whose intensity is at threshold is inverted, no damage is done (damage value 0). If a voxel whose intensity is prototypical of white matter is misclassified as outside-white-matter, the damage is as great as that of misclassifying a voxel whose intensity is prototypical of gray matter as white matter. These outright misclassifications are arbitrarily assigned a damage value of 1.

If a voxel whose intensity is prototypical of white matter is correctly classified as white matter after inversion, the inversion clearly corrected an outright misclassification of the presegmentation process and is therefore assigned a damage value of -1. By the same logic, the classification of a voxel whose intensity is prototypical of gray matter as outside-white-matter is assigned a damage value of -1.

A voxel intensity close to 0 (black) indicates that the voxel represents a point in the brain that is not only outside of the white matter, but also outside of the gray matter and, thus, even further away from the gray-white matter boundary, which we aim to recover, than a prototypical gray matter voxel. The damage done by adding such a voxel should therefore be markedly greater than that of adding a prototypical gray matter voxel. In all empirical tests presented in this paper we assigned a damage value of 10 to the addition of a voxel of intensity 0. By symmetry, the reclassification of such a voxel as outside-white-matter (deletion), correcting a previous misclassification of the same severity, is assigned a damage value of -10.

In contrast to the case of black voxels, white voxels whose intensity exceeds that prototypical of white matter cannot be inferred to be further away from the gray-white matter boundary than prototypical white matter voxels. Such voxels are merely to be classified as white matter with greater confidence. Therefore the damage values assigned to reclassification of such voxels are only slightly greater in absolute value than those for reclassification of prototypical white matter voxels. The damage values assigned for reclassification of a voxel of intensity 255 are 1.4 for the case of deletion and -1.4 for the case of addition.

Between the points just motivated, the misclassification-damage functions for addition and deletion are linear as shown in Figure 1.5. In the present version, the total damage value of each voxel that is to be inverted is obtained by adding 1 to its misclassification-damage value, ensuring that among corresponding solutions (cutting versus filling) the smaller set of voxels is chosen if both sets cause similar amounts of misclassification damage.

## Empirical validation

Our algorithm is guaranteed to remove all handles, rendering the surface topology of the segmentation simple. Empirical validation is required to go beyond this analytical fact and answer the question whether the corrections reflect the true geometry of the subject's cortical sheet at the locations where the presegmentation contained topological errors. We validated the algorithm's behavior in three ways: first, by subjecting its topology correction to a handle-by-handle

and voxel-by-voxel comparison to a human expert’s correction performed independently, second, by inspecting the corrections of about a thousand topological errors in the context of the polygon-mesh surface representation as well as in the context of the voxel volume of the original anatomical MR scan, and finally by crossvalidation between segmentations based on different scans of the same subject’s brain.

## Comparison to expert performance

As our algorithm automatizes a task previously performed by an expert, the natural way to validate the algorithm is to compare its performance to the expert’s. The manual topology correction, as it was routinely performed in our lab before the algorithm was developed, takes about 30 minutes per subject, matching the duration mentioned in Dale et al. (1999).

Application of our algorithm as a test of topological integrity to presegmentations manually corrected previously, however, revealed that small handles had frequently been overlooked. For the purpose of this validation study, the expert (author RG) therefore took particular care in his manual correction of the presegmentations, frequently requiring more than a full hour per subject. The presegmentations corrected by the expert were tested for topological integrity before the comparison, allowing us to compare the corrections chosen by the expert and by the algorithm for every single topological error.

The expert and the algorithm independently proceeded from the same presegmentation of each scan, allowing a precise voxel-by-voxel comparison of the changes. Like the algorithm, the expert inspected the original grayscale MR scan to decide how to correct the topology in the presegmentation. Unlike the algorithm, the manual correction procedure also included repeated polygon-mesh reconstruction and partial inflation based on the topologically incorrect presegmentations to locate topological errors and assess the local 3D geometry.

This labor-intensive form of validation has been performed for 18 hemispheres with a total of 326 topological defects. The results are summarized in Table 1. The part of the medioventral region where the segmentation’s boundary does not represent the cortical sheet but the subcallosal mask used in the presegmentation (see Method) has been excluded from the comparison. The number of handles found in this region is given in Table 1.

The comparison of the changes was carried out at two levels of analysis. First, on the level of single voxels, we counted voxels inverted only by the algorithm, by both, algorithm and expert, and only by the expert. Second, on the more abstract level of topological errors, we counted the number of handles corrected “congruently” and the number of handles corrected “incongruently”. The corrections chosen by algorithm and expert to correct a particular handle were counted as congruent if and only if: 1) the same method (deletion or addition of voxels) was chosen by both and 2) the cut (or inverse cut) was placed at the same location along the ring. Two nonidentical cuts (contiguous sets of inverted voxels) were considered to be “at the same location along the ring” only if they overlapped, i.e. if their intersection was a nonempty set of voxels.

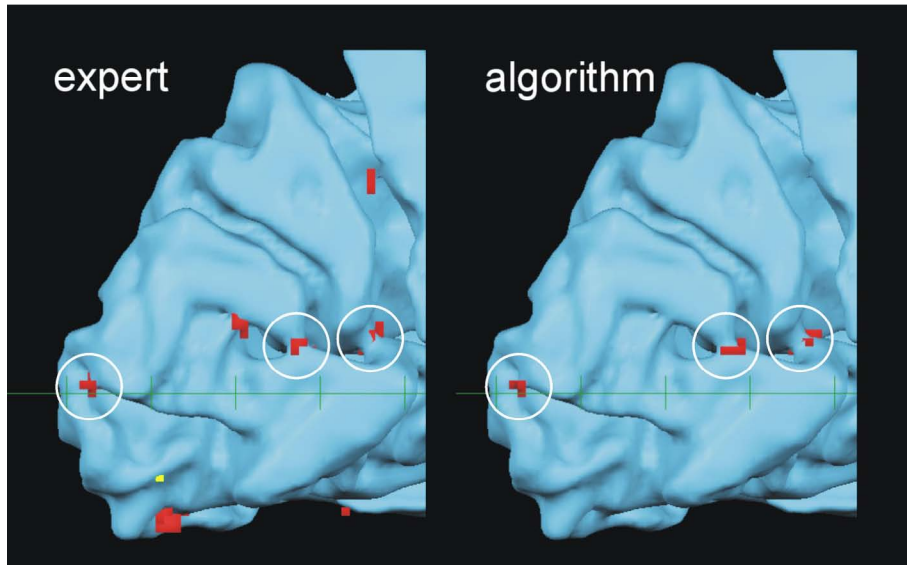


Figure 1.6: **Deletions by the expert and by the algorithm**

The figure shows how the expert and the algorithm corrected three topological errors (highlighted by white circles) located on the medial side of the occipital lobe of a left hemisphere. The presegmentation is shown as a smoothed surface (blue). Voxels deleted by the expert (left) and the algorithm (right) have been inserted in red. Note that the topology-correcting changes chosen by the expert and by the algorithm are extremely similar: both correctly identified the three errors and chose deletion as the method of correction in all three cases. Furthermore, the deletions chosen by the expert and by the algorithm are located at the same position within each toroidal structure. The expert made a number of additional changes (outside the white circles), a number of deletions (red), and one addition (yellow). Though they improve geometrical accuracy, these changes do not affect topology (see text).

The analysis on the level of topological errors shows that the algorithm chooses the same kind of correction (deletion or addition) and the same location for most handles (gray box in Table 1). Figure 1.7 shows several examples of how the algorithm and the expert corrected the same topological errors in the scan of subject CG. If we exclude the extreme case of subject TI, in which the presegmentation failed catastrophically due to an inhomogeneity artifact in the inferotemporal region causing the expert to invert 5518 voxels, the algorithm corrected 84% of the handles congruently. Close scrutiny was given to the 16% incongruent corrections chosen by the algorithm (b in Table 1). Each handle corrected incongruently was reexamined, revealing, first, that most of them occurred in regions where noise and complex geometry rendered the situation ambiguous even to the expert and, second, that where the same method (deletion or addition) of correction was chosen by expert and algorithm, the two sets of inverted voxels frequently touched without overlap. Though they were counted as incongruent, the latter changes are qualitatively identical and quantitatively so similar that the difference in the initial reconstruction they entail is likely to fade when the surfaces are morphed to precisely represent the inner or the outer boundary of the gray matter.

The conclusion that changes made only by the algorithm are few is also supported by the analysis on the level of single voxel inversions. The number of voxels inverted by the algorithm that have not been inverted by the expert is generally small (f in Table 1). Again excluding subject TI, the average number of voxels changed by the algorithm but not by the expert is 1.5 per topological error. This attests to the similarity of the topology-correcting changes chosen by expert and algorithm. The voxel-level analysis also shows that the algorithm’s corrections are extremely parsimonious. The algorithm inverted an average of only 4.3 voxels per topological error (subject TI excluded).

While parsimony, resulting from the thinnest-part heuristic (see Method and Discussion) implicit to our method, is an essential property of the algorithm’s behavior, this does not hold true for the expert. The expert makes many changes (contiguous sets of inverted voxels) that do not affect the topology (Figure 1.6: changes outside the white circles, c in Table 1) and where he corrects the topology he usually changes more voxels than the algorithm (Figure 1.7). The reasons for this are as follows. The 2D-slice representations of the presegmentation, in which the expert actually inverts the voxels, visually reveal a ring structure only if it happens to be oriented parallel to the plane of the slices. Conversely, ring-shaped configurations frequently appear in a 2D slice where there is no handle in the 3D object.

Thus, it is usually not obvious to the expert how a given topological defect can be removed changing a minimal number of voxels. Having located the defect in the polygon-mesh representation of the cortical sheet, the expert combines visual inspection of the original MR voxel intensities and knowledge about the spatial properties of the human cortex to improve the geometrical accuracy of the segmentation locally. Whether this removes the topological defect will become apparent only when topological integrity is checked in the next cycle of reconstruction and inflation. Since a geometrically accurate representation of the cortex is necessarily also topologically correct, improving local geometry is a good method of removing the topological defects.

Although the expert, thus, performed a more comprehensive task, essentially improving geometrical accuracy and removing topological errors as a side-effect, his behavior can serve as a reference in evaluating the algorithm, because we can show that the algorithm’s changes constitute a subset of the expert’s. The fact that changes made by the algorithm were, in general, also made by the expert represents evidence for correct behavior of the algorithm (a and e in Table 1). Since the algorithm is guaranteed to correct all topological errors, any additional changes made by the expert but not by the algorithm, though they may improve geometrical accuracy, are known to be unnecessary for topology correction and, thus, must not be considered in evaluating our algorithm’s performance. Only the few changes made by the algorithm, that the expert did not make (b and f in Table 1), indicate divergence between the topology-correcting changes effected by the algorithm and the expert.

In summary, the comparison revealed that the expert and the algorithm correct topological defects in very similar ways, except for the fact that the algorithm is consistently more parsimonious in its modifications than the expert.

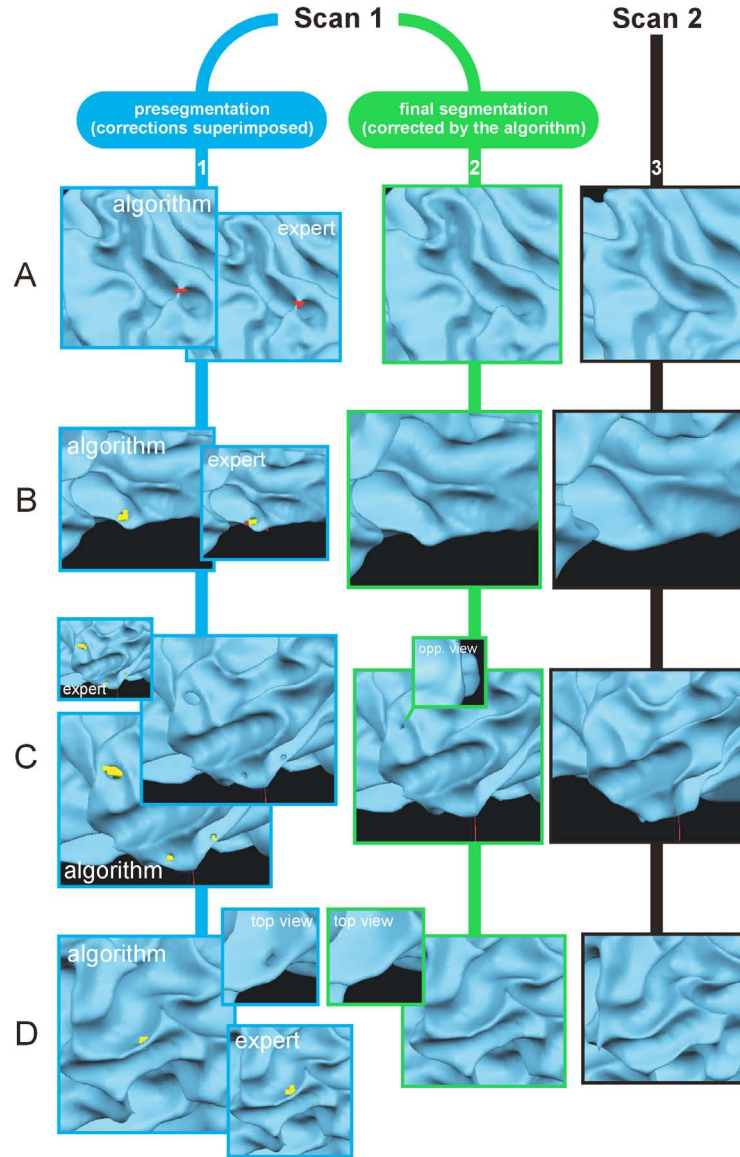


Figure 1.7: **Validation by comparison to a separate scan of the same subject (segmentation of scan 1 validated by scan 2)**

Smoothed reconstructions (blue) of the presegmentation (blue frames: 1), the final segmentation including the algorithm's corrections (green frames: 2), and a presegmentation of a different scan (scan 2) of the same subject (black frames: 3). Reconstruction has been performed completely independently for the two scans. Corrections by the algorithm and by the expert are inserted in column 1 (deleted voxels in red, added voxels in yellow). The figure shows medial views of the left parietal lobe (A), lateral views of the left occipital lobe (B), lateral views of the left temporal pole (C), and lateral and dorsal views of the right frontal lobe (D) of subject CG. Note the biggest of the three holes near the pole of the left temporal lobe (C1). There is a little dark spot in its place in the reconstruction of the corrected segmentation (C2, bottom). Examination of the opposite side of the hole (C2, top), however, shows that the topology is correct (see C3) and the dark spot is merely a small invagination of the surface. Axes, where visible, are Talairach axes.

Table 1: Comparison between the algorithm and a human expert

| subject,<br>hemisphere | automatic<br>segmentation                               |                                                      |                                                                                               | comparison between algorithm<br>and expert |                 |                                                                             |                                            |                                                        |                                               |     | scanner<br>(for models and<br>sequences,<br>see footnotes) |
|------------------------|---------------------------------------------------------|------------------------------------------------------|-----------------------------------------------------------------------------------------------|--------------------------------------------|-----------------|-----------------------------------------------------------------------------|--------------------------------------------|--------------------------------------------------------|-----------------------------------------------|-----|------------------------------------------------------------|
|                        | genus (number of handles)<br>of the presegmentation (g) | number of handles corrected<br>by deletion, addition | number of handles intersect-<br>ing the subcallosal mask (s),<br>excluded from the comparison | handle<br>corrections                      |                 | additional changes made by<br>the expert that do not affect<br>topology (c) | voxel inversions                           |                                                        |                                               |     |                                                            |
|                        |                                                         |                                                      |                                                                                               | congruent (a)                              | incongruent (b) |                                                                             | performed only by the<br><b>expert</b> (d) | performed by <b>both</b> ,<br>expert and algorithm (e) | performed only by the<br><b>algorithm</b> (f) |     |                                                            |
| VA <sup>1</sup>        | L                                                       | 10                                                   | 3, 7                                                                                          | 3                                          | 7               | 0                                                                           | 4                                          | 38                                                     | 39                                            | 0   | GE 1.5T <sup>2</sup>                                       |
|                        | R                                                       | 7                                                    | 0, 7                                                                                          | 2                                          | 3               | 2                                                                           | 2                                          | 33                                                     | 15                                            | 9   |                                                            |
| CG <sup>3</sup>        | L                                                       | 11                                                   | 4, 7                                                                                          | 0                                          | 9               | 2                                                                           | 9                                          | 99                                                     | 40                                            | 26  | Siemens 1.5T <sup>4</sup>                                  |
|                        | R                                                       | 14                                                   | 5, 9                                                                                          | 5                                          | 9               | 0                                                                           | 22                                         | 168                                                    | 32                                            | 25  |                                                            |
| JH                     | L                                                       | 7                                                    | 2, 5                                                                                          | 1                                          | 3               | 3                                                                           | 11                                         | 552                                                    | 27                                            | 16  | Siemens 1.5T <sup>4</sup>                                  |
|                        | R                                                       | 2                                                    | 1, 1                                                                                          | 0                                          | 1               | 1                                                                           | 7                                          | 47                                                     | 2                                             | 4   |                                                            |
| EF                     | L                                                       | 14                                                   | 2, 12                                                                                         | 1                                          | 12              | 1                                                                           | 64                                         | 1097                                                   | 40                                            | 23  | Siemens 1.5T <sup>4</sup>                                  |
|                        | R                                                       | 22                                                   | 8, 14                                                                                         | 5                                          | 13              | 4                                                                           | 36                                         | 721                                                    | 33                                            | 33  |                                                            |
| TI <sup>5</sup>        | L                                                       | 63                                                   | 31, 32                                                                                        | 7                                          | 36              | 20                                                                          | 58                                         | 4646                                                   | 222                                           | 170 | GE 3T <sup>6</sup>                                         |
|                        | R                                                       | 33                                                   | 8, 25                                                                                         | 3                                          | 23              | 7                                                                           | 44                                         | 593                                                    | 57                                            | 32  |                                                            |
| RG                     | L                                                       | 18                                                   | 3, 15                                                                                         | 2                                          | 12              | 4                                                                           | 16                                         | 93                                                     | 22                                            | 12  | Siemens 1.5T <sup>4</sup>                                  |
|                        | R                                                       | 19                                                   | 9, 10                                                                                         | 2                                          | 14              | 3                                                                           | 7                                          | 75                                                     | 20                                            | 7   |                                                            |
| SV                     | L                                                       | 22                                                   | 4, 18                                                                                         | 4                                          | 15              | 3                                                                           | 14                                         | 210                                                    | 27                                            | 39  | Siemens 1.5T <sup>4</sup>                                  |
|                        | R                                                       | 11                                                   | 0, 11                                                                                         | 2                                          | 8               | 1                                                                           | 12                                         | 50                                                     | 26                                            | 9   |                                                            |
| MC                     | L                                                       | 17                                                   | 0, 17                                                                                         | 1                                          | 13              | 3                                                                           | 9                                          | 98                                                     | 39                                            | 14  | Siemens 1.5T <sup>4</sup>                                  |
|                        | R                                                       | 20                                                   | 0, 20                                                                                         | 0                                          | 18              | 2                                                                           | 12                                         | 142                                                    | 49                                            | 31  |                                                            |
| FS                     | L                                                       | 24                                                   | 4, 20                                                                                         | 1                                          | 21              | 2                                                                           | 13                                         | 681                                                    | 95                                            | 37  | Siemens 1.5T <sup>4</sup>                                  |
|                        | R                                                       | 12                                                   | 4, 8                                                                                          | 1                                          | 10              | 1                                                                           | 0                                          | 293                                                    | 55                                            | 14  |                                                            |

numbers of contiguous  
voxel sets that have  
been inverted<sup>7</sup>

numbers of single  
voxel inversions<sup>8</sup>

<sup>1</sup> Courtesy of R. Malach, Weizmann Institute

<sup>2</sup> GE Signa Horizon LX 8.25, 1.5T, IR-prepared fast GRE T1-weighted sequence

<sup>3</sup> Scan 1 in the crossvalidation (see Figures 8 and 9)

<sup>4</sup> Siemens Magnetom Vision, 1.5T, T1FLASH

<sup>5</sup> Courtesy of B. Wandell, Stanford University

<sup>6</sup> GE Signa Horizon LXII, 3T, FGRE3D

<sup>7</sup> Each handle has either been excluded (those intersecting the subcallosal mask) or classified as congruently or incongruently changed, thus  $g=s+a+b$ .

<sup>8</sup> Outside the region defined by the subcallosal mask, the number of voxels inverted by the expert is  $d+e$  and the number of voxels inverted by the algorithm is  $e+f$ .

## Visual inspection of the algorithm's corrections

To be able to test our conclusions about the correctness of the algorithm's behavior on a larger sample of anatomical MR volumes, we next adopted a more efficient method of expert validation. We ran the complete segmentation and reconstruction procedure on a total of 30 anatomical MR volumes. The automatic correction of about a thousand topological errors was assessed, first, by comparing the polygon-mesh reconstructions of the presegmentation and the final topology-corrected segmentation of each of the 60 hemispheres and, second, by inspecting an additional voxel volume written for validation purposes by our implementation of the algorithm, in which every voxel inverted in the presegmentation as well as every voxel inversion considered but rejected (the inverse solutions) is color-coded in the context of the grayscale intensity values of the original MR scan. We found that, where corrections were counterintuitive, it was the method of correction (deletion or addition), rarely the placement of the cut along the ring that seemed incorrect.

In the quantitative comparison to expert performance described in the previous section, there had been topological errors, whose adequate correction the expert felt unable to judge. This implies that a portion of the congruent corrections chosen by the algorithm might be inadequate and a portion of the incongruent corrections might reflect the true structure of the cortical sheet. During visual inspection we therefore restricted classification of the algorithm's corrections to topological errors whose adequate correction was apparent, determining the cleaner measure of the proportion of modifications chosen by the algorithm that could unequivocally be classified as inadequate. The proportion of corrections judged as inadequate was merely 3%.

## Crossvalidation using duplicate scans of the same brains

The approach to validation described so far relied on expert judgement on what constitutes the true structure of the cortical sheet in regions where the presegmentation is topologically incorrect. To go beyond expert judgement, we performed a second anatomical MR scan for three of our subjects. Segmentation and polygon-mesh reconstruction was performed independently from the two scans of these subjects. Since the topological errors of the presegmentation are mainly caused by noise, they come to lie in different locations, allowing a crossvalidation between the two scans.

For subject CG the results are documented in Figures 1.7 and 1.8. Figure 1.7 shows the validation of the algorithm's segmentation of scan 1 by inspection of the presegmentation of scan 2. The reverse validation is shown in Figure 1.8. Note that the surfaces in these figures are initial reconstructions of the segmentation, not yet morphed to precisely represent the geometry of the inner or the outer boundary of the gray matter. They serve to show how the algorithm corrects the topology in voxel space. To ease visual orientation, the surfaces representing the pre- and final segmentations (blue) have been smoothed slightly, whereas the surfaces representing the changed sets of voxels (red for deletions, yellow for additions) are inserted in their original cubic voxel shape. Figure 1.1

shows the final precise representations of the inner and outer boundaries of the gray matter obtained by morphing these initial reconstructions.

The crossvalidation performed for 6 hemispheres confirmed our finding that inadequate solutions are rare. The three outright errors found all concerned the method of correction (deletion or addition), confirming our intuitive judgement during visual inspection. Examination of the original anatomical MR volume revealed spatial inhomogeneity of the anatomical images as the cause of the failure of the heuristic choice between the two solutions. For each of the three incorrect solutions, enforcing the inverse solution rendered the resulting reconstruction correct. Furthermore, for all other handles, which had been removed by the correct method, the location along the ring of the object (or inverse object) where the cut was placed, reflected the true anatomy very precisely (see Figures 1.7 and 1.8), lending further support to the thinnest-part heuristic implicit to the algorithm (see Method and Discussion).

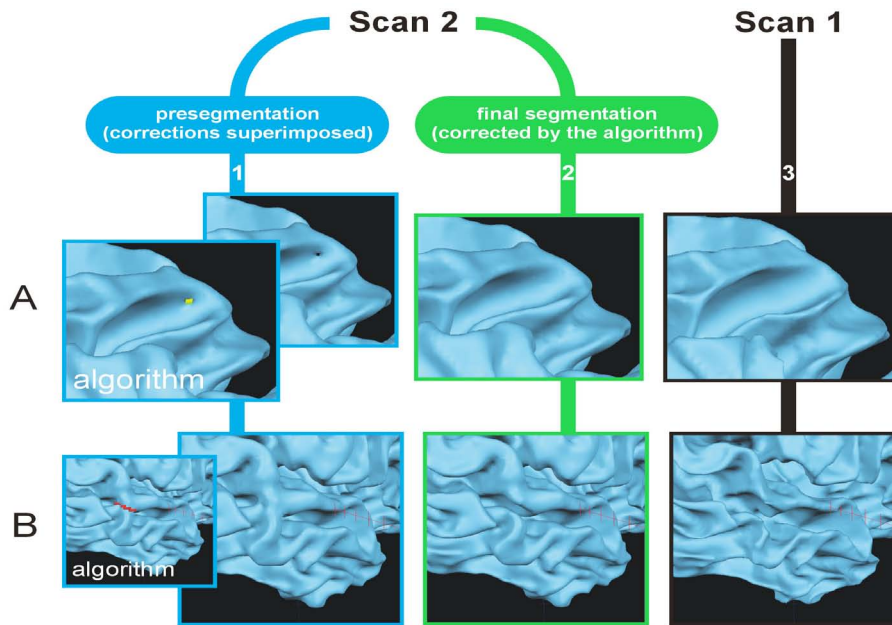


Figure 1.8: **Reverse validation (segmentation of scan 2 validated by scan 1)**

Here, the algorithm has been applied to scan 2 and the results are validated by examination of scan 1. The figure shows lateral views of the left parietal lobe (A) and of the right temporal lobe (B, topological error in the superior temporal sulcus). Axes, where visible, are Talairach axes.

## Linear-time-complexity implementation

In this section, we show that the time complexity of the core components of the algorithm (distance-to-surface mapping and selftouching-sensitive, distance-to-surface prioritized region growing) is linear in the number of voxels of the

input. Establishing the sets of interacting cuts, choosing which cuts to make, and making the changes is also possible in time linear in the number of voxels, though this is a less important issue because the number of handles, in the domain of cortex segmentation in anatomical MR volumes, is usually small compared to the number of input voxels, rendering the time required for these components of the algorithm negligible compared to that required by the core components.

Region growing in 3D voxel space has linear time complexity. It can be viewed as analogous to graph searching, the vertices representing the voxels and the edges the side-adjacency relation. Each vertex of a graph representing a voxel object in this way has at most 6 edges. Since the total number of edges in this special type of graph is, thus, proportional to the number of vertices, the searching can be performed in time linear in the number of vertices.

In our algorithm, the voxel object is, of course, represented by a 3D occupancy grid. Fringe voxels as well as object voxels already included in the growing region are marked in the grid. Fringe voxels are additionally stored in a linked list. Time complexity is linear because every object voxel is included only once, and all operations associated with the inclusion of a voxel, i.e. marking it as included and updating the fringe set, can be performed in constant time using the occupancy grid (for the object, the region, and the fringe) and the linked list representation of the fringe, which allows constant time access to the next fringe voxel to be included. Updating the fringe takes constant time because the number of object voxels to be considered has a constant upper bound of 6, the number of side-adjacent neighbors of the included voxel. To decide whether a voxel point-adjacent to the included voxel needs to be added to the fringe, all that is required is checking if it is an object voxel and if it is already present in the fringe set.

Region growing in 3D voxel space can be used not only for the selftouching-sensitive, prioritized region growing process, but also for the distance-to-surface mapping: The growing in each step of the erosion (see above, Formal definition of the algorithm) is constrained to voxels representing the outer layer of object voxels not yet mapped. Each region growing pass maps one layer of the object and marks it as mapped. Since each pass takes time linear in the number of voxels contained in the layer it maps, and the numbers of voxels of all the layers add up to the number of voxels of the object, the distance-to-surface mapping takes time linear in the number of object voxels.<sup>1</sup>

To find the thinnest part of each ring structure, two features need to be added to a 3D region growing process that merely floodfills the whole object: distance-to-surface prioritization and selftouching sensitivity.

Distance-to-surface prioritization can be added preserving linear time complexity by using a separate linked list or stack representation for fringe voxels of

---

<sup>1</sup>Performing the distance-to-surface mapping by region growing has an additional advantage: during each step of erosion it can be detected, whether the eroded object still contains rings. If it doesn't contain any more toroidal structures, the distance-to-surface mapping can terminate because how the region grows into the ringless core of the object in the main step of the algorithm is inconsequential for the result. If the eroded object still contains rings can be detected during the erosion by making the region growing marking the outer layer selftouching-sensitive. If the 2D outer layer selftouches, the 3D eroded object selftouches, i.e. contains rings. How selftouching sensitivity can be added to voxel space region growing keeping the time complexity linear is described below in this section.

each priority level and maintaining a pointer to the highest non-empty fringe subset. This allows a highest priority fringe voxel to be accessed in constant time at each point in the process.

Adding the feature of selftouching sensitivity in a way that keeps the time complexity linear is a little trickier. Whereas the distance-to-surface of a voxel remains constant throughout the process, whether the region selftouches in a voxel or not can change several times as the region grows. In a naive implementation using only the representations introduced so far, this can lead to extensive searching through the fringe set for a fringe voxel the region doesn't selftouch in, rendering the time complexity quadratic in the worst case. To avoid searching for an includable fringe voxel, a separate fringe set for voxels known to be selftouching can be used at each priority level in conjunction with a voxel position grid of pointers.

At each voxel position currently representing a fringe voxel known to be selftouching, the grid contains a pointer to the linked list element representing the fringe voxel. This is advantageous because the selftouching status of a voxel can only change when a point-adjacent voxel is included into the region.

Fringe voxels known to be selftouching (which are stored in a separate linked list) are not considered for inclusion into the region. When a fringe voxel not known to be selftouching is considered for inclusion, its selftouching status must be checked, which takes constant time because the checking operates only within the 26-voxel neighborhood of the voxel in question. If the region doesn't selftouch in the voxel, the voxel is included into the region. If it does selftouch in the voxel, the voxel is moved to the separate list containing only fringe voxels known to be selftouching, and a pointer to its list element is inserted into the 3D grid of pointers indicating that the voxel is known to be selftouching and where its list element is.

A fringe voxel known to be selftouching is moved to the corresponding list of fringe voxels not known to be selftouching when a fringe voxel in its point-adjacency neighborhood is included into the region because this and nothing else can change its selftouching status. To be able to move fringe voxels between the two subsets of fringe voxels maintained for each priority level in constant time, we use doubly linked lists.

This scheme preserves linear time complexity because each object voxel eventually included into the region is handled in constant time. An object voxel is handled for the first time when it is included into the fringe. It is initially added to the list of fringe voxels not known to be selftouching. While it is in the fringe, it may be moved to the list of fringe voxels known to be selftouching and back to the list of fringe voxels not known to be selftouching a number of times before finally being included into the growing region. Each such move is handled in constant time and the number of times the voxel is moved back and forth between the two fringe subsets cannot exceed 26, because a voxel's selftouching status can change only when a point-adjacent voxel is included into the region, which takes place only once for each object voxel.

Once all point-adjacent voxels are part of the growing region, the voxel can no longer be selftouching and is therefore included into the region the next time it is considered for inclusion. Actually the maximal number of times a fringe voxel

can go through the cycle (become selftouching and non-selftouching again) is much lower than 26, but any constant upper bound suffices to show linearity. Most voxels will be included into the region either when they are considered for inclusion for the first time or after very few moves.<sup>2</sup>

Selftouching-sensitive, distance-to-surface prioritized region growing, thus, can be performed in time linear in the number of object voxels, as can the distance-to-surface mapping. The same process applied to the inverse object is linear in the number of non-object voxels in the input, rendering the core of the whole process linear in the number of input voxels.

## Discussion

We have proposed an algorithm that makes changes in a binary voxel object to enforce simple topology of its boundary polyhedron. The intuitive rationale of the algorithm—minimizing the damage done to the presegmentation—makes our approach attractive for segmentation of any object whose topology is known to be simple. We have demonstrated and validated the algorithm’s operation for the particular domain of recovering the spatial structure of the human cerebral cortex from an individual subject’s anatomical MR scan. We have explained how the algorithm can be implemented such that its time complexity is linear in the number of input voxels. This section will critically discuss the thinnest-part heuristic implicit in the algorithm, the heuristic choice between filling and cutting, and possible improvements of our approach.

While our algorithm is guaranteed to output an object whose boundary is a simple polyhedron, its choice of topology correction relies on two heuristics: the thinnest-part heuristic, which is implicit to the idea at the core of the algorithm, and the more easily replaceable heuristic of choice between correction by addition and deletion. We will discuss each of them in turn.

### The thinnest-part heuristic

We have motivated the thinnest-part heuristic above by stating that it is more likely that a given ring structure resulted from erroneous addition or deletion of a small contiguous set of voxels than from addition or deletion of a larger one. Note, however, that the algorithm does not necessarily minimize the number of

---

<sup>2</sup>Note that a fringe voxel’s selftouching status may change back and forth while the voxel is in the list of voxels not known to be selftouching without this entailing any move between the two sets, because selftouching status is only checked when the voxel is considered for inclusion into the region. An alternative implementation maintaining a fringe subset of voxels known not to be selftouching in addition to the subset known to be selftouching would, by the same logic, have linear time complexity as well. Since such an algorithm, as opposed to ours, tracks all changes of selftouching status for all fringe voxels, however, it would be considerably slower.

voxels changed to cut a handle. Instead, the location of the cut along the ring is the location where the ring breaks first during repeated erosion.<sup>3</sup>

Minimal cutting would be computationally much more expensive than our method of selftouching-sensitive, distance-to-surface prioritized region growing. For the case of anatomical MR scans of the human cerebral cortex, we would not expect the results to differ much if the algorithm strictly minimized the number of cut voxels. Either approach is heuristical and, thus, has to be validated empirically for the domain it is to be applied in, which we have done for our approach as applied in the domain of anatomical MR scans of the human cortex. The thinnest-part heuristic emerges as very powerful in choosing the location of the cut along the ring.

## Heuristic choice between deletion and addition of voxels

An important property of our approach is the adaptive choice between two complementary solutions for each topological defect. Let's simplify a little and assume that there is a correct binary voxel representation of the object to be recovered whose genus is 0. After presegmentation the voxel object contains two types of error: voxels that are set but should be vacant (positive errors to be eliminated by positive cutting, i.e. by deletion) and voxels that are vacant but should be set (negative errors to be eliminated by negative cutting, i.e. by addition). The intensity threshold used in the presegmentation process influences the frequency of these complementary types of error. If the threshold is close to the intensity of the gray matter, there will be many positive errors and comparatively few negative ones. If the threshold is close to the intensity of the white matter, the reverse situation is to be expected. Positive cutting eliminates toroidal structures by removing voxels from the object (cutting handles). Negative cutting eliminates toroidal structures by adding voxels to the object (filling holes). The algorithm integrates both methods by choosing the better solution for each toroidal structure. It is, thus, adaptive in two important ways. First, it adapts to the local situation in the region of each toroidal structure. Second, it adapts to the threshold used globally in the presegmentation process. The threshold can thus be chosen independently of the topology correction step. This is an important property, because it allows other constraints to be taken into account in choosing the threshold.

The particular heuristic cost function by which the “better” solution is to be chosen is a separate issue. Here we chose the approach of minimizing the “damage” done to the presegmentation. The validation has shown that, though errors were rare, most of those that occurred concerned this heuristic choice, suggesting that a different cost function, e.g. based on local histogram analysis (to account

---

<sup>3</sup>The chosen cut  $C(h)$  is a set of voxels whose maximal distance-to-surface is minimal within the set  $PC$  of all possible cuts removing the ring structure:

$$\max \{d(v) | v \in C(h)\} = \min \{\max \{d(v) | v \in C'\} | C' \in PC\},$$

where  $d(v)$  is the distance-to-surface of voxel  $v$ . It is easy to imagine a ring structure for which the cut comprising the minimal number of voxels does not satisfy this criterion. If the cut of minimal size has the shape of a disk, for example, an alternative cut of a different shape can have more voxels that are all closer to the object surface than the maximal distance-to-surface within the disk-shaped minimal cut.

for spatial inhomogeneity of the signal) or on geometrical a-priori constraints might further improve the quality of our automatic topology correction.

## Future directions

Although the cortical surface reconstruction process as described in this paper is already very efficient and yields satisfactory results, we are following two lines of development.

The first concerns the nature of the other segmentation operations and the stage within the segmentation process at which the topology correction algorithm is invoked. Before topology correction in voxel space was available, it was necessary to reduce the number of topological defects by applying binary voxel object smoothing (as described in the section Method) in voxel space. For the purposes of this paper, the presegmentation has been performed in the same way, with the topology correction inserted as the final step of the segmentation, immediately preceding polygon-mesh reconstruction. Though smoothing is appropriate as it implements the a-priori constraint of finite curvature of the cortical sheet, it may be beneficial to reduce the amount of smoothing performed in voxel space. Furthermore, it may be preferable to perform the smoothing in a topology-preserving way after topology correction, as this would allow us to solve the problem of “needles”. Needles are protuberances of the object surface that have a 1D structure (a single voxel thick). Needles are not necessarily topological errors but clearly incompatible with what is known about cortex curvature. They occur due to noise in the original MR scan, and in the region of handles they sometimes resist smoothing if it is performed before the handle is removed, suggesting the iterative application to convergence of a topology-preserving binary voxel object smoothing operation after topology correction.

The second line of development concerns the selftouching-sensitive region-growing algorithm itself. We plan to develop the algorithm toward greater interactivity with the original anatomical MR scan: In the present version, the intensities of the original MR scan are used only in choosing, for each handle, between the previously determined complementary solutions of cutting and filling. Intensities could also be used to determine the locations of the cuts and fillings by including them in the prioritization function of the region growing process. The priority value of a voxel would reflect its classification confidence and be a function of both, the distance-to-surface (computed on the basis of a rough presegmentation) and the intensity of the voxel. Region growing would then be performed concurrently in the positive and in the negative object guided by this common priority metric, such that the two processes define the final segmentation interactively.

## Conclusion

We have defined and validated an algorithm that segments anatomical MR scans of the human brain to recover the spatial structure of the cortical sheet. Based on a selftouching-sensitive, distance-to-surface prioritized region growing process, the algorithm enforces the a-priori constraint of simple topology and

yields geometrically and topologically correct reconstructions of the cortical sheet. The algorithm is fast and, more importantly, has linear time complexity, making it suitable also for application to higher resolution or supersampled anatomical datasets.

Combined with segmentation and polygon-mesh morphing methods similar to those described by Dale et al. (1999) and MacDonald et al. (2000), respectively, our algorithm allows us to obtain, fully automatically, topologically and geometrically accurate polygon-mesh representations of the inner and outer boundaries of the cortical gray matter at a computational cost much lower than that of previous approaches. The complete process from an anatomical MR volume of 1 mm isotropic resolution in Talairach space to polygon meshes representing the inner and the outer boundary of both hemispheres takes less than 15 minutes on a current PC workstation.

## Acknowledgments

The authors wish to thank Miguel Castelo-Branco, Steffen Egner, and Elia Formisano (Universiteit Maastricht) for the insights and Roi Mukamel, Rafael Malach (Weizmann Institute of Science), and Brian Wandell (Stanford University) for the brains they contributed. This research was supported by the Max Planck Society and the Universiteit Maastricht.



## Chapter 2

# Cortical object recognition from a visual motion flowfield

**Corresponding publication:** Kriegeskorte N, Sorger B, Naumer M, Schwarzbach J, van den Boogert E, Hussy W, and Goebel R. Human cortical object recognition from a visual motion flowfield. *Journal of Neuroscience*, 23(4):1451-63, 2003.

**Abstract:** Moving dots can evoke a percept of a 3-D object’s spatial structure in the absence of other visual cues. This phenomenon, called structure from motion (SFM), suggests that the motion flowfield represented in the dorsal stream can form the basis of object recognition performed in the ventral stream. SFM processing is likely to contribute to object perception whenever there is relative motion between the observer and the object viewed.

Here we investigate the motion flowfield component of object recognition with functional magnetic resonance imaging. Our SFM stimuli encoded face surfaces and random 3-D control shapes with matched curvature properties. We used two different types of SFM stimulus: with the dots either fixed to the object’s surface or moving on it. Despite the radically different encoding of surface structure in the two types of SFM, both elicited strong surface percepts and involved the same network of cortical regions. From early visual areas, this network extends dorsally into the human motion complex and parietal regions, and ventrally into object-related cortex. The SFM stimuli elicited a face-selective response in the fusiform face area. The human motion complex appears to have a central role in SFM object recognition, not merely representing the motion flowfield but also the surface structure of the motion-defined object. The motion complex and a region in the intraparietal sulcus reflected the motion state of the SFM-implicit object, responding more strongly when the implicit object was in motion than when it was stationary.

## Introduction

The primate visual system recovers the 3-D surface structure of an object by combining information from a variety of visual cues, including contour, shading, binocular disparity and motion. Whenever an object moves relative to the observer, the visual motion flowfield is one of the information sources on whose basis the visual system determines the object's 3-D structure. The contribution of motion as a cue to structure has long been recognized (Wallach and O'Connell, 1953) and extensively been studied using moving dot stimuli constructed to minimize other cues (see Andersen and Bradley, 1998). The fact that primate observers can perceive the objects implicit to these structure-from-motion (SFM) stimuli suggests crosstalk between the dorsal and the ventral visual pathway (Ungerleider and Mishkin, 1982), two subsystems often assumed to be somewhat segregated. More specifically, it suggests that the motion flowfield represented in the dorsal stream can form the basis of object recognition performed in the ventral stream. What network of cortical regions subserves the extraction of complex object structure from a motion flowfield?

SFM perception is thought to involve an explicit representation of the motion flowfield (Treue et al., 1991). The prime candidate region for a motion flowfield representation is the human motion complex (hMT+, also called V5; Zeki et al., 1991; Tootell et al., 1995), the likely human homologue of a complex of motion-related regions in monkey cortex including the middle temporal area (MT) and its satellite regions, namely the medial superior temporal area (MST) and an area in the fundus of the superior temporal sulcus (FST). While many regions of primate visual cortex process motion information, the motion complex appears to have a special role in the representation of the motion flowfield (Maunsell and Van Essen, 1983a; Albright, 1984). It abstracts from other features of the visual input including orientation and color and integrates motion cues over small patches to resolve the local directional ambiguity that results from the aperture problem (Pack and Born, 2001). Lesions of MT entail SFM perception impairments in monkeys (Andersen and Bradley, 1998). Consistently, imaging studies have shown that hMT+ plays a role in SFM perception (Orban et al., 1999; Paradis et al., 2000). These studies, however, did not investigate SFM-based complex-surface perception and object recognition.

SFM object recognition is probably performed on the basis of the motion flowfield representation in hMT+. If the process involves a sequence of stages, the lateral occipital region (LO) and posterior aspect of the fusiform gyrus (pFs), subsumed under the name lateral occipital complex (LOC), might be the next step (Malach et al., 1995; for a review see Malach et al., 2002). LOC has been shown to be involved in many different types of object perception (e.g. Grill-Spector et al., 1999; Amedi et al., 2001) and is conjectured to represent perceived object shape (Kourtzi and Kanwisher, 2001). It is noteworthy that LOC and hMT+, though associated with ventral and dorsal stream, respectively, are very close together on the cortex. In anesthetized monkeys, the homologues of these regions and others have been found to be responsive to 3-D geometrical objects defined by various visual cues including motion (Serenio et al., 2002).

In this functional magnetic resonance imaging (fMRI) study, we investigate the human cortical network of regions subserving SFM complex-surface perception and SFM object recognition. As implicit objects, we use face surfaces defined by

visual motion. This choice of stimulus allows us to trace SFM object recognition all the way from the motion flowfield presented to a category-specific response in the ventral stream’s fusiform face area (FFA).

Face-selective neuronal responses have long been known in monkey electrophysiology (e.g. Perrett et al., 1982). Face-selective human cortical activity has been described more recently (Allison et al., 1994; Puce et al., 1995). The term FFA refers to a face-selective region found in the human fusiform gyrus anterior to pFs. This region is defined purely functionally, i.e. not by cytoarchitectonics but by its selective response to faces, which has been demonstrated for many different forms of visual representation using fMRI (e.g. Kanwisher et al. 1997; Kanwisher et al. 1998; Kanwisher et al. 1999).

## Materials and methods

### General experimental rationale

#### Classical versus on-surface SFM

Building on our group’s initial finding that random-dot SFM stimuli of rotating human heads, can activate FFA (Goebel et al., 2000. Soc. Neurosc. Abstr.), this study explores SFM object recognition for two radically different forms of SFM encoding.

We use the conventional SFM encoding, in which the dots are fixed to the object surface (Figure 2.1A), as well as a novel SFM encoding, in which the dots move on the object’s surface (Figure 2.1B). Do high-level object-selective regions including FFA respond to motion-defined faces presented in either type of SFM encoding? What respective networks of regions perform the presumably very different computations required for extraction of object structure in the different encodings?

#### Faces versus curvature-matched random shapes

We use motion-defined face surfaces to be able to elicit a category-specific response in FFA (Figure 2.1C, left). As SFM control stimuli, we use random surfaces created to have similar curvature properties as the face surfaces (Figure 2.1C, right; for details, see below: Stimuli, Encoded surfaces: faces and random shapes). This allows us to show that the low-level curvature properties of the face surfaces cannot explain the response of FFA to the SFM face stimuli. Everyday (e.g. man-made) objects or simple geometrical objects (e.g. a cube) are not well-suited as controls because their curvature properties represent a confound (for instance, sharp edges frequently occur in such objects but tend to be absent in faces).

In addition to the random-shape SFM controls, we use moving and static non-SFM random-dot control stimuli closely matched in terms of low-level properties (details below).

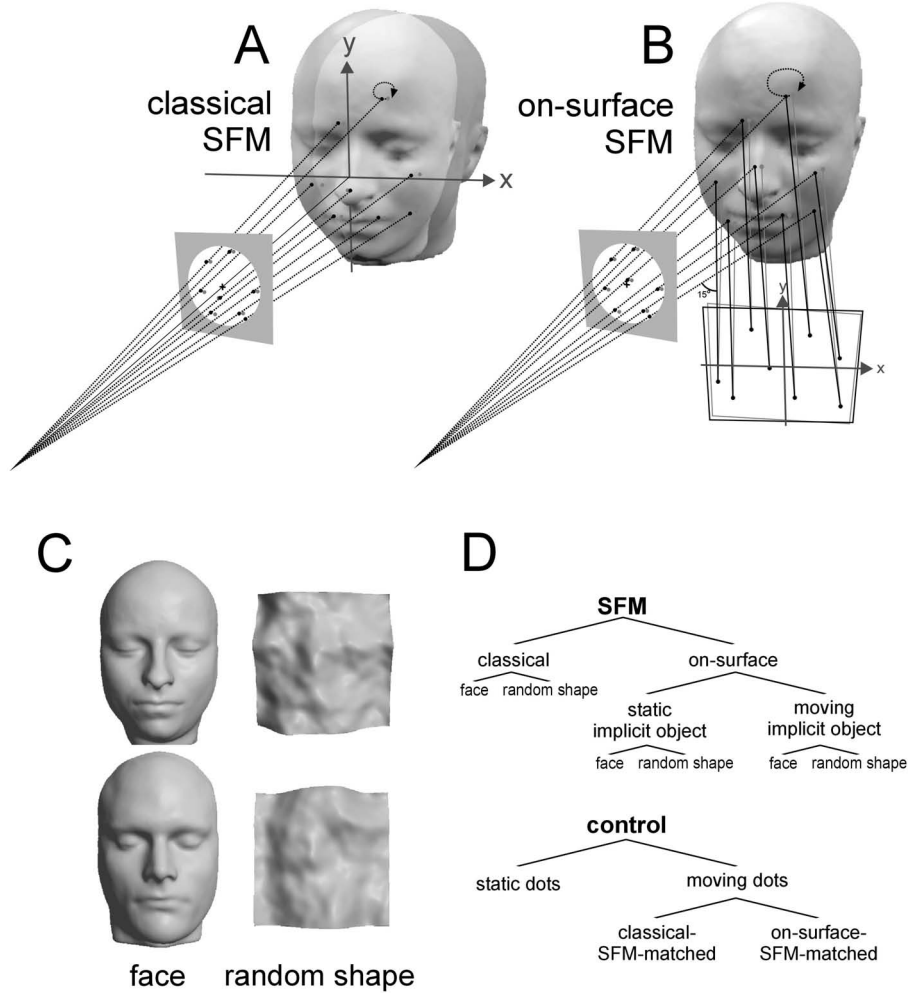


Figure 2.1: **Construction of classical SFM and on-surface SFM stimuli**

This figure contrasts the construction of the two types of SFM encoding used. (A) In our version of the classical SFM stimulus, dot locations are randomly selected on an object's surface. These fixed locations are then polar-projected onto an image plane as the object moves in 3-D. Each dot, thus, has a fixed position relative to the object implicit to the stimulus. (B) In our novel on-surface SFM stimulus, the dots move on the object's surface as though they were projections of parallel laser beams randomly arranged in an array, which moves rigidly. The motion of the laser array (rectangle in B) as well as the motion of the implicit object (A) was a rotatory harmonic oscillation around each of the orthogonal  $x$ - and  $y$ -axes with a 90 deg phase shift. In a separate condition, both laser array and implicit object underwent this type of motion (moving-implicit-object on-surface SFM, not shown). In the actual stimuli, the background and circular aperture were black and the dots were white. The figure does not represent the quantities (number of dots, relative positions of the elements) correctly. (C) Polygon-mesh surfaces used as SFM implicit object (here shown as shape-from-shading stimuli). Each random shape was produced from the face shown next to it by scrambling the phases in the depth maps' Fourier transforms. (D) Taxonomy of the moving-dot conditions used in the SFM experiment (for details, see Stimuli and Experimental designs and tasks under Methods and materials).

### Face-nonface categorization task

Subjects fixated a central cross and performed a face-nonface categorization task, providing a behavioral control of the object recognition process. Since face and random-shape stimuli have similar curvature properties, this detection task cannot be performed on the basis of a few local measurements at isolated positions in the visual field. It requires a complex and spatially more or less continuous surface representation.

Had everyday or simple geometrical objects been used as controls, salient local features of the motion flowfield (e.g. those occurring at sharp edges of a 3-D object) would have allowed subjects to classify the object as a nonface without having formed continuous surface percept.

### Circular aperture eliminates outer object boundary

We use a circular aperture to eliminate the outer boundary of the implicit objects in all stimuli (Figure 2.1, A and B). The outer object boundary is often visible in SFM stimuli defining complex object surfaces as the contour between dot-filled and empty regions. Here, we eliminate this non-motion cue to surface structure by superimposing a circular aperture to all stimuli, which completely and continually occludes the outer boundary of all faces and random shapes. The result is a circular region completely filled with dots in each frame. This approach allows us to show that the activation of FFA by SFM stimuli is not dependent on the outer-boundary cue.

### Fixation-friendly SFM stimuli: objects wobble about the fixation point

The mode of motion most frequently chosen in SFM studies is rotation (e.g. Bradley et al., 1998). However, a rotating SFM stimulus evokes much weaker percepts when subjects are asked to fixate than when they are allowed to let their eyes naturally follow the stimulus. This is plausible for two reasons. Firstly, smooth pursuit of a point on the object surface minimizes the retinal velocity in the region of the fovea and thereby enhances sensitivity to the subtle velocity vector differences that encode the surface structure. Secondly, perceiving the object naturally triggers smooth pursuit, so to fixate may require active suppression of the object percept.

To resolve the tension between the conflicting instructions to fixate and to perceive the object, we attached the fixation dot to a fixed position on the object's front surface. This point on the object's surface is the center of its motion and thus does not move in 3-D space.

### Separate localization of cortical key regions

The key regions of interest hMT+, LOC and FFA were localized in each individual subject separately by contrasting (see above: fMRI measurements and analysis) appropriate stimulus conditions. For hMT+ localization, we contrasted the moving and static random-dot control conditions of the main experiment.

## Stimuli

### Common features of the experimental and control stimuli

Classical SFM, novel on-surface SFM (Figure 2.1, A and B, respectively), as well as all random-dot control stimuli shared the following features: Displays consisted of about 1000 dots within a circular aperture. The dots (single pixels) were square and of a width of approximately 0.06 deg visual angle. Each stimulus contained a fixation cross at the center of the aperture, which served to prevent the cue of the outer object contour (e.g. the head silhouette) from contributing to the surface percept in SFM object conditions. The aperture had a size of approximately 15 deg visual angle.

For classical and on-surface SFM experimental stimuli, the 3-D locations of the dots on the object surface were presented under perspective projection. The aperture completely occluded the outer object contour in each frame. The number of 1000 dots is approximate because in SFM conditions object structure and motion lead to frame-to-frame fluctuations of the number of dots as dots move in and out of the region occluded by the aperture.

### Classical SFM stimulus

A set of positions was chosen pseudorandomly from a uniform distribution over the surface area. The dot trajectories were computed by projection onto the image plane of this set of positions as the object moves in 3-D. Whereas the SFM object is often modeled as transparent with the dots continuously visible even when they move to the back surface, the object in this study was modeled as opaque. Dots on the self-occluded portion of the surface were not shown.

For dots fixed on the object surface to evoke a surface percept, the object needs to be in motion. As the mode of object motion, we chose wobbling about fixation because—as explained above—this resolves the tension between the conflicting instructions to fixate and to perceive the object. The object wobbles about a point on its front surface, which coincides with the fixation point. The wobbling motion resembles the precession of the spin axis of a proton in a magnetic field. More precisely, our wobbling consists in two simultaneous rotatory oscillations—around frontoparallel horizontal and vertical axes through the fixation point (see Figure 2.1A). The two rotatory oscillations are harmonic, have a 90 deg phase shift relative to each other and an amplitude of 10 deg.

The resulting stimuli can be statically fixated without deterioration of the object percept. Our method has the additional positive effect of keeping the view angle approximately constant despite the continual motion of the object.

### Pilot exploration of the perceptual effects of different types of on-surface SFM

On-surface motion can take many forms, not all of which evoke a strong percept of the 3-D surface. We informally tested a number of on-surface motion stimuli encoding faces that did not evoke surface percepts. In these stimuli, each dot moved at a constant on-surface velocity and followed an independent on-surface

trajectory changing direction only as the surface dictated. We varied the following aspects of the stimulus: (1) the dots were initially randomly distributed on the surface or on the image plane, (2) the dots moved in random directions or all in the same direction with their trajectories constrained to be straight on the image plane (leaving only image-plane velocity as a cue to surface orientation), and (3) orthographic or perspective projection was used. Though not all possible combinations were tested, these informal psychophysical experiments suggested that on-surface motion stimuli do not evoke surface percepts when the dots are set on independent trajectories on the surface. A different type of on-surface motion appeared to be required.

### **The “parallel lasers” on-surface SFM stimulus used in this study**

A rigid array of “parallel lasers” was constructed by pseudorandomly choosing positions within a rectangle. The lasers were modeled as projecting dots onto the surface orthogonally from their fixed position within the rectangular array. The array was modeled to wobble as described above for the object motion of our classical SFM stimulus (harmonic oscillatory rotation around two orthogonal axes with a 90 deg phase shift), causing the projections to describe cyclic trajectories on the surface. Motion is essential to the 3-D surface percepts evoked by these displays: when the dynamic stimuli are halted, the percept immediately disintegrates. The on-surface SFM stimuli were closely matched to the classical SFM stimuli. Nevertheless, their motion flowfields have different velocity vector distributions.

There were two on-surface SFM conditions: In the first, the implicit object the dots are projected on (face or random shape) was stationary. In the second, the implicit object—like the laser array—precessed (see Figure 2.1B).

### **Encoded surfaces: faces and random shapes**

The two face surfaces used (Figure 2.1C, left) were obtained by reconstructing the outer boundary of whole-head T1-weighted anatomical magnetic resonance (MR) scans as polygon meshes. Our SFM techniques (see above) ensured that the faces were always presented approximately in frontal view. The two matched random 3-D shapes (Figure 2.1C, right) were created to have similar curvature properties as the faces by randomizing the phases in the Fourier transform of the faces’ depth maps, while preserving the amplitudes at each combination of frequency and orientation. To prevent the depth discontinuities between opposite edges of the depth maps from adding artifacts to the power spectrum, the faces’ depth maps were periodicized prior to phase randomization by Gaussian smoothing applied selectively at the edges with toroidal wrap-around.

### **Random-dot control stimuli**

Moving-dot control stimuli were constructed by shifting each dot’s motion trajectory vertically and horizontally by a random amount with toroidal wrap-around, effectively relocating every dot randomly in a square region. This spatial scrambling obliterated the surface encoding (subjects did not perceive surfaces

in these stimuli), while preserving the shape of each single dot's trajectory as well as the temporal phase relation of the dots' trajectories. The resulting display was restricted to the same circular aperture used in the experimental stimuli. Due to selective occlusion by the aperture, the trajectories visible were not exactly identical to those visible in the original stimuli, but they were closely matched. Two such motion control stimuli were constructed, one from an SFM face stimulus, the other from an on-surface SFM face stimulus. Since the two faces and the random 3-D control shapes used had qualitatively and quantitatively similar curvature properties, these two motion control stimuli have similar motion trajectory content as the classical and on-surface SFM experimental stimuli. Static-dot control stimuli were obtained by taking random single frames from the SFM face stimuli.

### **Photos used for FFA and LOC localization**

The locations of brain regions involved in object and face perception were determined in separate experiments with photo stimuli, very similar to those described by Malach et al. (1995) and Kanwisher et al. (1997). For consistency with our SFM stimuli, we slightly varied the established localization procedures: Photos of objects, scrambled objects and faces were presented within a circular aperture of the same size as in the SFM stimuli. The photos were 252-by-252-pixel grayscale images.

In the LOC localization experiment, object photos (obtained from various sources and including different object categories) and scrambled versions of the same photos were presented. The scrambling of each photo was performed by tessellating the image into little squares of  $10 \times 10$  pixels (resulting in  $25 \times 25$  tiles), selecting the subset of tiles falling within the circular aperture and randomly rearranging those tiles.

In the FFA localization experiments, face photos and the same object photos used in the LOC localization experiment were presented. As in the SFM face stimuli, the faces were presented in frontal view and the outer contour (including hair) was hidden by the aperture. The face photos did not contain accessories such as glasses or jewelry and were of approximately the same size (about 15 deg visual angle) as the SFM face stimuli.

### **Retinotopy mapping**

Visual area borders were determined using a conventional polar mapping technique (Serenio et al., 1995; Goebel et al., 1998). Rotating checkerboard wedges were 22.5 deg in polar angle width, and spanned 0.5 to 20 deg in eccentricity. A fixation point was shown at the center of the screen. Wedges reversed in contrast 8.3 times/second and rotated counterclockwise starting at the upper vertical meridian. Each participant completed ten cycles (64 seconds per cycle) that lasted 10 minutes 40 seconds with an additional 20 seconds of fixation at the beginning and end of each run.

### Stimulus presentation in the scanner

The stimulus image signal was generated by a PC at a frame rate of 60Hz. The image was projected onto a frosted screen located at the end of the scanner bore (at the side of the subject's head) with a Sony VPL-PX21 LCD projector equipped with a special lens. The subject viewed the stimuli via a mirror mounted to the head coil at an angle of about 45 deg. In SFM and localization experiments, the stimuli had a size of approximately 15 deg visual angle.

## Experimental designs and tasks

### Design of SFM experiment

The experiment comprised nine random-dot stimulus conditions (taxonomy in Figure 2.1D: classical SFM: moving faces, classical SFM: moving random shapes, on-surface SFM: static faces, on-surface SFM: moving faces, on-surface SFM: static random shapes, on-surface SFM: moving random shapes, moving-dot control matched to classical SFM, moving-dot control matched to on-surface SFM, static-dot control). Each condition appeared twice in each run, except for the two moving-dot control conditions, each of which appeared only once in each run. There were, thus,  $7 \times 2 + 2 \times 1 = 16$  stimulation periods separated by  $16 + 1 = 17$  fixation periods. As each period had a duration of 16s, an experimental run lasted 8min and 48s. The condition sequence was pseudorandom but symmetrical. Each of the seven subjects underwent four runs of the SFM experiment.

### Task in SFM experiment

Subjects were familiarized with the stimuli prior to the fMRI experiment. They were instructed to continually fixate a central cross visible throughout the experiment and to classify each stimulus presented as either "face" or "non-face" as soon as they could by pressing one of two buttons (two-alternative forced choice). Because of a technical problem (broken light fiber), only responses indicating a face percept were recorded. As none of the subjects pressed the face button in any of the non-face conditions and all of the subjects pressed it in every single face condition, we can nevertheless conclude that all stimuli were classified correctly.

### LOC and FFA localization experiments

In both LOC and FFA localization experiments, a block design alternating stimulus and fixation periods was used. Each run consisted of six 30s stimulus blocks and seven 20s blocks of fixation (resulting in 5min and 20s measurement time per run). In each block, 45 different photos were presented foveally at a rate of one every 670 ms. The stimulus blocks alternated between the two different conditions. Subjects were instructed to view passively but attentively. All seven subjects underwent LOC and FFA localization experiments. For LOC localization, each subject underwent two runs. For FFA localization, five of the subjects underwent two runs, two subjects one run.

## Subjects

Seven subjects aged between 21 and 34 participated in the study (average age: 25.3 years). They had normal (4 subjects) or corrected-to-normal (3 subjects) vision. Four of them were female, three male. Six of them were right-handed, one left-handed. Potential subjects received information about magnetic resonance imaging and a questionnaire allowing us to exclude those to whom the experiment would have entailed a health risk. All subjects gave their informed consent by signing a form. The form as well as the experimental techniques used in this study were approved by the ethical committee CWOM of the Academisch Ziekenhuis (university hospital) associated with the Katholieke Universiteit Nijmegen (The Netherlands).

## Functional and anatomical magnetic resonance imaging

### Functional measurements in the SFM and localization experiments

We measured 20 transversal slices at 1.5T (Siemens Magnetom Sonata) using a single-shot gradient-echo echo-planar-imaging sequence. The pulse-sequence parameters were as follows: in-plane resolution:  $3.125 \times 3.125 \text{ mm}^2$ , slice thickness: 5mm in the SFM experiment and 4mm in the localization experiments, gap: 0mm, slice acquisition order: interleaved, field of view (FoV):  $200 \times 200 \text{ mm}^2$ , acquisition matrix:  $64 \times 64$ , time to repeat (TR): 2000ms, time to echo (TE): 60ms, flip angle (FA): 90 deg. A functional run lasted 5min and 20s in the localization experiments and 8min and 48s in the SFM experiment.

### Functional measurements in the retinotopy mapping experiment

25 axial slices with  $3 \times 3 \times 3 \text{ mm}^3$  isotropic voxels were acquired on a 3T scanner (Siemens Medical Solutions, Erlangen, Germany). TR: 2000ms, TE: 35ms, FA: 70 deg. One scan lasted 11min and 20s, yielding 340 volumes.

### Anatomical measurements

Each subject underwent a high-resolution T1-weighted anatomical scan at 1.5T (Siemens Magnetom Sonata), which was either a 3-D MPRAGE lasting 8min and 34s (192 slices, slice thickness: 1mm, TR: 2000ms, TE: 3.93ms, FA: 15 deg, FoV:  $250 \times 250 \text{ mm}^2$ , matrix:  $256 \times 256$ ) or a 3D-T1-FLASH lasting 16min and 5s (200 slices, slice thickness: 1mm, TR: 30ms, TE: 5ms, FA: 40deg, FoV:  $256 \times 256 \text{ mm}^2$ , matrix:  $256 \times 256$ ).

## Statistical analysis

### Preprocessing

Prior to statistical inference, the fMRI data sets were subjected to a series of preprocessing operations: 1) Slice-scan-time correction was performed by resampling the time courses with linear interpolation such that all voxels in a given

volume represent the signal at the same point in time. 2) Small head movements were automatically detected and corrected by utilizing the anatomical contrast present in functional MR images. The Levenberg-Marquardt algorithm was used to determine translation and rotation parameters (6 parameters) that minimize the sum of squares of the voxelwise intensity differences between each volume and the first volume of the run. Each volume was then resampled in 3-D space according to the optimal parameters using trilinear interpolation. 3) Temporal high-pass filtering was performed to remove temporal drifts of a frequency below 3 cycles per run (3/528s). 4) The functional volumes were projected into Talairach space using the scanner's position parameters, which relate the functional slices to an anatomical volume measured in the same session for each subject. 5) Only for the group analysis, each functional volume was smoothed by spatial convolution with a Gaussian kernel of a full width at half maximum of 4mm. The BrainVoyager 2000 (version 4.8) software package was used for all stages of the analysis (preprocessing, multiple linear regression, reconstruction of the cortical sheet and visualization of functional maps).

### Multiple linear regression at every voxel

Single-subject and Talairach-space group ( $n=7$ ) analyses were performed by multiple linear regression of the response time course at each voxel using nine predictors corresponding to the nine experimental conditions (see Figure 2.1D and Experimental design and task). The predictor time courses were computed using a linear model of the hemodynamic response (Boynton et al., 1996) and assuming an immediate rectangular neural response during each condition of visual stimulation.

To reveal the SFM-object-recognition network (Figure 2.2), we performed an extra-sum-of-squares F-test at each voxel for all six SFM conditions together. To contrast conditions of the main as well as the localizer experiments (Figures 2.2, 2.4 and 2.5), we computed t-statistics at each voxel on the basis of the  $b$ -weights ( $\beta$  estimates). In the figure legends, the thresholds used are described by their  $p$ -values (Bonferroni-corrected for multiple comparisons).

### Response profile analysis for individually defined key regions

For each subject, the key regions hMT+, LOC and FFA were individually localized by appropriate contrast analyses as described in the previous section. For every key region of every subject, the spatially averaged time course was subjected to multiple linear regression analysis using predictor time courses computed from the stimulation protocol on the basis of a linear model (Boynton et al., 1996) of the hemodynamic response. The key-region time courses were standardized, so each  $b$ -weight, reflects one condition's BOLD response amplitude relative to the variability of the signal. The  $b$ -weights obtained for the individually localized regions were averaged across subjects and their standard errors were appropriately adjusted (Figure 2.3). This approach is to be preferred over averaging effect estimates in %-signal change because the latter can vary widely between subjects and it is not clear whether this reflects interindividual variation of the effects in terms of neural activity. Possibly spurious effect differences

can, thus, lead to an average response profile for a small group that is dominated by one or two subjects and not qualitatively representative of the group.

### **Retinotopy mapping**

BOLD time series were analyzed separately for each hemisphere. A rectangular function reflecting when a stimulus entered the contralateral visual field (6s on period) was convolved with a hemodynamic impulse response function. The resulting hemodynamic-response-predictor time course was correlated with each voxel time course at 14 lags. Lags ranged from 0-14 TRs (i.e. 0-28s). Voxels were color-coded according to the lag that produced the highest correlation exceeding a threshold of  $r > 0.275$ . These lag correlation maps were projected onto a flattened representation of the cortical white-gray matter boundary. Borders between early visual areas were defined by phase reversals in the retinotopy map (Serenio et al., 1995).

## **Results**

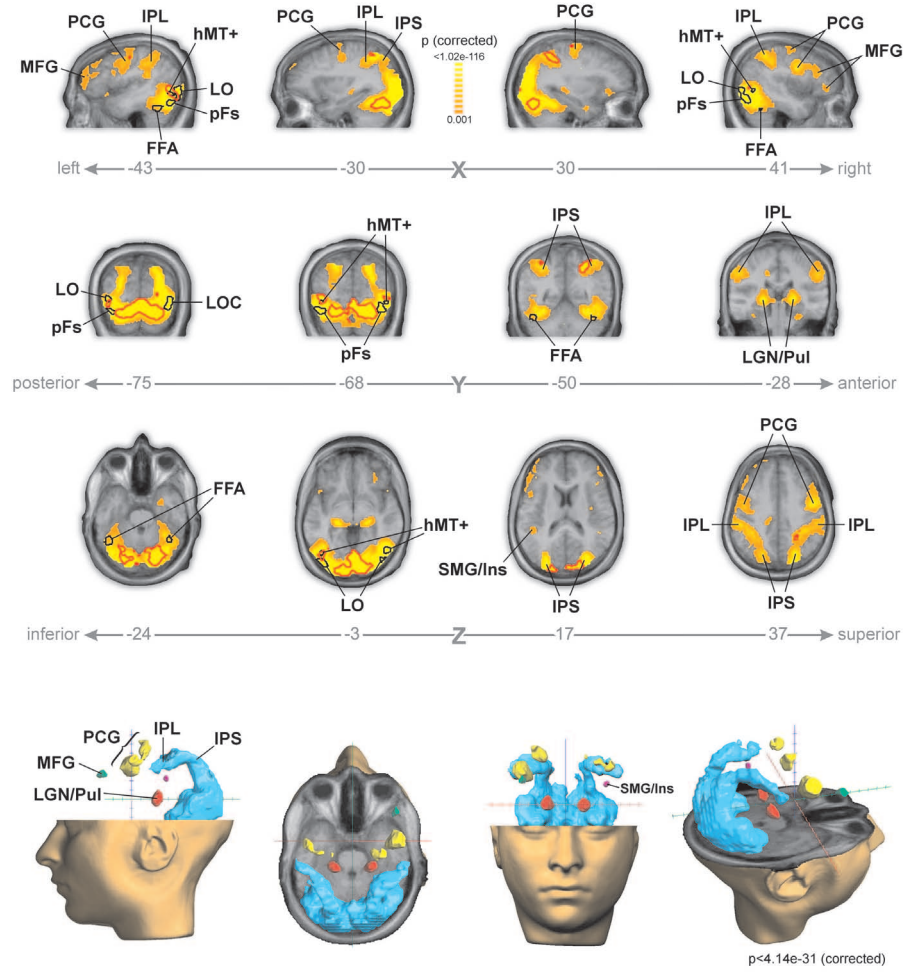
### **Strong surface percepts evoked by two radically different types of SFM encoding**

An SFM stimulus can evoke a strong percept of a complex 3-D object, such as a face, even when the outer boundary is completely eliminated by superimposing a circular aperture, as was the case in all our stimuli. SFM stimuli can be constructed in many different ways. The classical method widespread in the literature is to select a number of locations on an object's surface and to project these fixed locations as dots onto an image plane over a sequence of frames, across which the object moves continuously. The resulting moving-dot displays evoke strong percepts of both the object's structure and its motion. We will refer to this type of SFM as classical SFM (Figure 2.1A).

In a series of psychophysical pilot experiments, we found that 3-D surface structure can also be perceived in moving dot stimuli constructed in a radically different way (Figure 2.1B). In contrast to classical SFM, in which the surface-defining moving dots are fixed to the object surface, our novel structure from on-surface motion stimulus (on-surface SFM for brevity) consists of dots moving on the object's surface (for details, see Materials and methods). In a natural environment such encodings of surface structure in a motion flowfield arise when light-shadow contours or fluids move across an object. Like classical SFM stimuli, on-surface SFM stimuli can evoke strong surface percepts even in the absence of the outer-boundary cue.

### **A common network of regions subserving classical and on-surface SFM perception**

Despite the radically different encoding of the surface structure, the fMRI results show that both classical and on-surface SFM perception involve the same



**Figure 2.2: SFM-object-recognition network (group results)**

Brain regions active during SFM object recognition (Talairach-space group analysis). FFA: fusiform face area, hMT+: human motion complex, Ins: insula, IPL: inferior parietal lobule, IPS: intraparietal sulcus, LGN: lateral geniculate nucleus, LO: lateral occipital object-related region, MFG: midfrontal gyrus, PCG: precentral gyrus, pFs: posterior fusiform object-related region, Pul: pulvinar, SMG: supramarginal gyrus.

(Top) Orange to yellow regions are significantly active during SFM object recognition (as compared to fixation periods during which only a small central cross was visible; extra-sum-of-squares F-test for all classical and on-surface SFM predictors;  $p < 0.001$ , corrected). Regions outlined in red are significantly more active during on-surface than during classical SFM conditions (t-test,  $p < 0.005$ ; for details, see Materials and methods). Regions significantly more active during classical than during on-surface SFM were not found (reverse contrast, same threshold). Outlined in black are the key regions hMT+, LOC (lateral occipital complex, i.e. LO and pFs), and FFA as defined by separate localizer contrasts using appropriate stimuli including photos (see Stimuli in section Materials and methods). The thresholds all satisfy  $p < 0.05$  (corrected) but have been increased for hMT+ ( $p < 0.005$ , corrected) and LO/pFs ( $p < 0.0001$ , corrected) to select a t-map contour enclosing a plausible volume for each region. The numbers on the gray axes specify the Talairach location of the slices shown. (Bottom) Glass-brain representations of the SFM-object-recognition network. Regions inside the rendered activation surfaces (red, blue, yellow, green) are highly significantly active during SFM object recognition (group results;  $p < 4.14e - 31$ , corrected). The large contiguous cortical expanse shown in blue is symmetrical with respect to the medial plane and includes early visual areas as well as the key regions hMT+, LOC, and FFA.

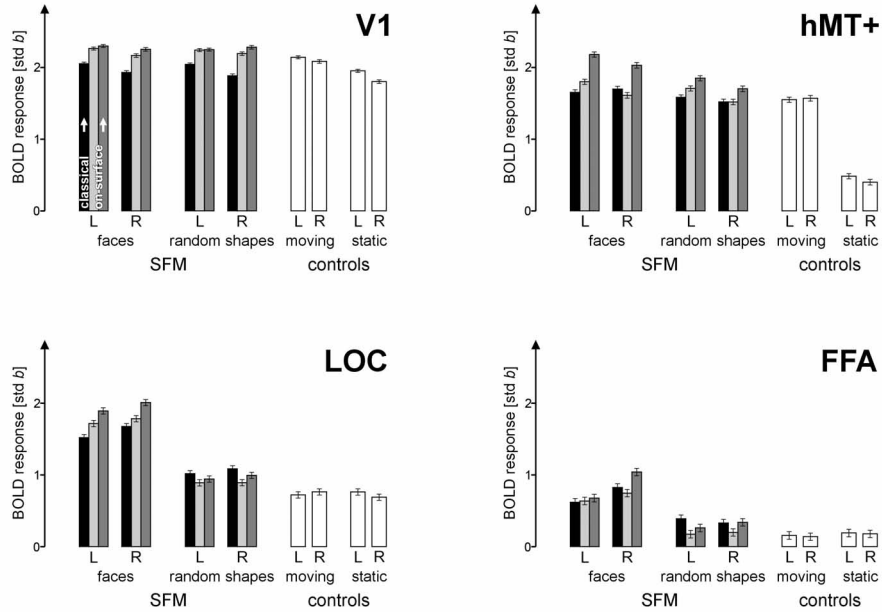


Figure 2.3: **Key-region response profiles (group results)**

Responses to SFM object and control stimuli in regions of interest as reflected in the linear-regression standardized  $b$ -weights ( $\beta$  estimates) averaged across subjects. Group averaging is based not on Talairach correspondence, but on individual localization of the key regions in each subject. V1 has been localized anatomically, hMT+, LOC and FFA functionally (see Materials and methods). For each subject and region, the  $b$ -weight entering into the average has been obtained by multiple-regression analysis of the spatially averaged time course. Error bars indicate the standard error of the average  $b$ -weight. “L” and “R” indicate left and right hemisphere responses, respectively. Black bars represent responses to classical SFM stimuli, gray bars responses to on-surface SFM stimuli. Light gray bars represent stationary-, dark gray bars moving-implicit-object on-surface SFM responses. White outlined bars represent responses to control stimuli as labeled. (For details, see Statistical analysis in section Materials and methods.)

network of regions (Figure 2.2). The network covers a large contiguous expanse of visual cortex (blue activation surface in Figure 2.2, bottom), which is strikingly symmetrical and extends from early visual areas dorsally into hMT+, the intraparietal sulcus (IPS) and other parts of parietal cortex and ventrally into LOC and more anterior ventral temporal cortex including FFA.

Additional cortical sites active during SFM object recognition were in the pre-central gyrus (PCG) and midfrontal gyrus (MFG). Subcortically, there was strong bilateral thalamic activity probably including the lateral geniculate nucleus as well as pulvinar.

### Early visual areas

The SFM network probably includes all retinotopic visual cortex. This conclusion is suggested by the Talairach-space extend of the activated region in the group analysis (Figure 2.2). It was confirmed in subject JS by retinotopy mapping (Sereni et al., 1995), which allows precise determination of the boundaries

of early visual areas (Figure 2.5, C and D). Notably, the SFM network also includes motion-sensitive areas V3a (Tootell et al., 1997) and V3b/KO (Orban et al., 1995; Van Oostende et al., 1997; Smith et al. 1998). For the purposes of this study, we subsume V3b/KO under V3a (Figure 2.5, C and D) because it is as of yet unresolved whether V3b/KO is a separate area or a subset of V3a representing the fovea (Singh et al., 2000) and because we have not performed experiments to distinguish between the two. V3a’s response profile will be described below in the context of that of key region hMT+.

### Midfrontal and precentral gyrus

In addition, a large bilateral region in the (PCG) and bilateral regions of the midfrontal gyrus (MFG) were consistently activated during SFM object recognition. These regions may contribute to the working-memory, attention and fixation components of the task.

The activated PCG region (Figure 2.2, top row and yellow blobs in bottom row) probably includes the frontal eye field (FEF). This conclusion is based on Talairach-space location and on an analysis of the response profile and time course. In the Talairach-space group analysis the coordinates of the left and the right PCG regions in question were -49, -1, 42 and 49, 0, 37 (centers of gravity), respectively. These regions’ activity was sustained for the full 16s of each condition period (Talairach-space group event-related average, not shown). Though motor cortex is close by in PCG, such sustained activity is unlikely to be button-press-related because only a single button press occurred toward the beginning each 16s stimulus period. Transient activity, whose latency was consistent with the average button-press reaction time of our subjects, was found more posteriorly in PCG (not shown). Because of its transient nature, this button-press-related motor activity did not appear as part of the SFM network but was only detected at a lower threshold using the same multiple-regression model.

That this bilateral PCG region’s response was sustained suggests that it is the FEF because the FEF is known to be active during fixation (Petit et al., 1999a), which was an important component of the task in this study. More specifically, the FEF may have contributed to the suppression of eye movements triggered by the moving displays. As the FEF contributes to both fixation and control of dynamic eye movements including smooth pursuit (Petit et al. 1999b), its activity does not allow any strong conclusions about subjects’ eye-movement behavior. We are confident that our subjects managed to fixate because our SFM stimuli were especially designed to be fixation-friendly—a property we tested elaborately outside the scanner. Moreover, the FEF response we found differed neither between static and moving random-dot conditions, nor between static- and moving-implicit-object SFM conditions (not shown). This suggests that our fixation-friendly SFM stimuli allowed fixation not only outside but also inside the scanner (see Materials and methods, General experimental rationale, Fixation-friendly SFM stimuli: objects wobble about the fixation point). FEF did respond slightly more strongly during SFM face conditions, which might be due to the suppression of stereotypical face scanning patterns required by the instruction to fixate.

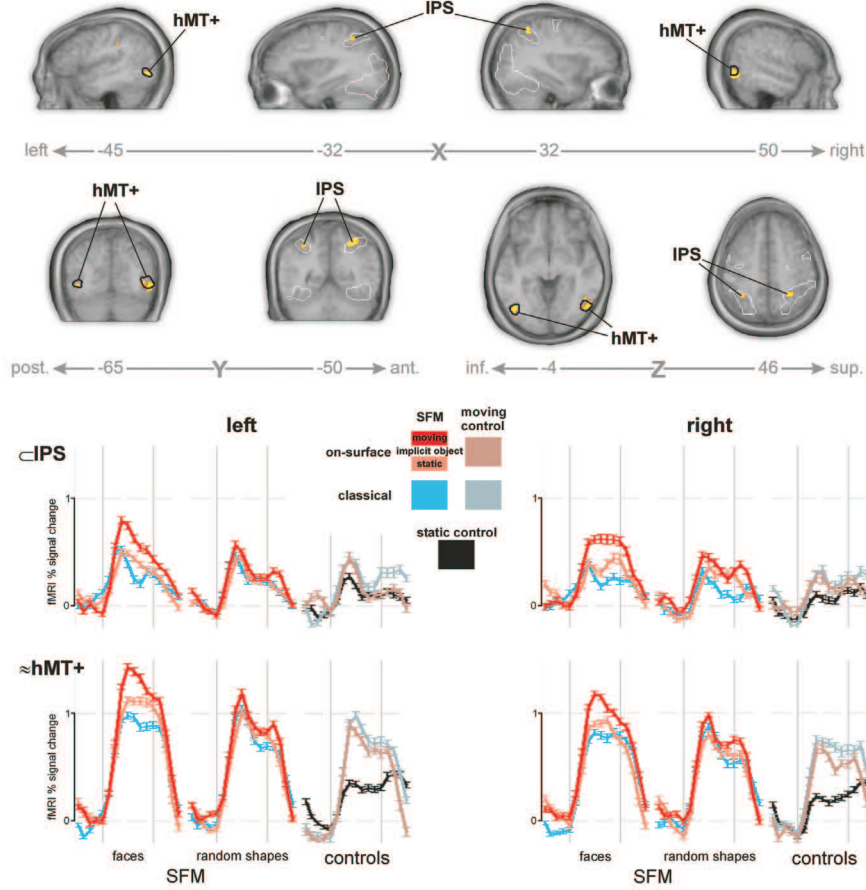


Figure 2.4: **Regions reflecting motion of the SFM-implicit object (group analysis)**

In on-surface SFM, the dots move on the object's surface, whereas the implicit object itself can be either stationary or moving. (Top) To detect regions reflecting motion of the implicit object independent of the surface-defining retinal motion, we contrasted moving- and stationary-implicit-object on-surface SFM conditions, which have very similar retinal motion flowfields. This contrast revealed a region in the intraparietal sulcus (IPS) and hMT+ ( $p < 0.005$ , corrected). Single-subject coordinates of the IPS region are given in Table 3. HMT+ as determined by a separate localizer contrast is shown outlined in black (same as in Figure 2.2). Regions more active during stationary- than during moving-implicit-object on-surface SFM were not found (reverse contrast, same threshold). As a spatial reference, the general SFM-object-recognition network shown in Figure 2.2 is outlined in white (threshold as in the bottom row of Figure 2.2). Group analysis is based on Talairach-space correspondence. (Bottom) Group event-related average time courses for all conditions spatially averaged across the regions as shown in the top panel. Error bars indicate the standard error of the mean. The color coding is defined in the visual legend. For statistical details, see Statistical analysis in section Materials and methods.

### **Stronger activity during on-surface than during classical SFM**

During on-surface SFM perception, activity was found to be slightly but significantly greater than during classical SFM in early visual areas, ventral stream regions including part of LOC but not FFA and dorsal stream regions including a part of hMT+ and a bilateral region in the IPS (see red outlines in Figure 2.2 and black versus gray bars in Figure 2.3). There were no regions significantly more strongly active during classical than during on-surface SFM perception. As low-level properties of the motion flowfield differed between classical and on-surface SFM conditions (e.g. greater mean velocity during on-surface SFM), it is questionable whether these effects are entirely caused by the different SFM encoding. However, low-level properties cannot entirely explain these effects as no significant differences were found between the two motion control conditions matched to classical and on-surface SFM in V1, V3a and hMT+ (Talairach-space group analysis).

### **Stronger activity during SFM-face than during SFM-random-shape perception**

When the implicit object was a face instead of a random 3-D shape of similar surface curvature all ventral stream regions responded more vigorously and FFA became active. During on-surface SFM with a moving implicit object this face effect was even evident in hMT+.

### **Response profiles of early visual areas and individually localized key regions**

Motivated by the sequential model outlined in the introduction, we localized the key regions hMT+, LOC and FFA using appropriate stimuli as described in the literature (random-dot stimuli for hMT+ and photos for LOC and FFA, see Materials and methods). Figure 2.2 (black outlines) shows the locations of these regions as determined by group analysis in Talairach space.

Determining the key regions in each individual subject showed that there is considerable variability in Talairach-space location across subjects (Table 1). We therefore defined each key region separately for each subject by the appropriate single-subject contrast analysis. The spatially averaged time courses reflecting the key regions' behavior during the SFM experiment were first analyzed for each subject individually. These individual analyses have been integrated in Figure 2.3, which shows the response selectivity of each key region, averaged across subjects (see Materials and methods).

### **Early visual areas**

V1 and adjacent early visual areas responded to all random-dot displays including static dots about equally strongly. Motion-sensitive area V3a responded significantly to static random dot displays but markedly more strongly to all moving dot conditions. (V3a was localized by retinotopy mapping in one subject

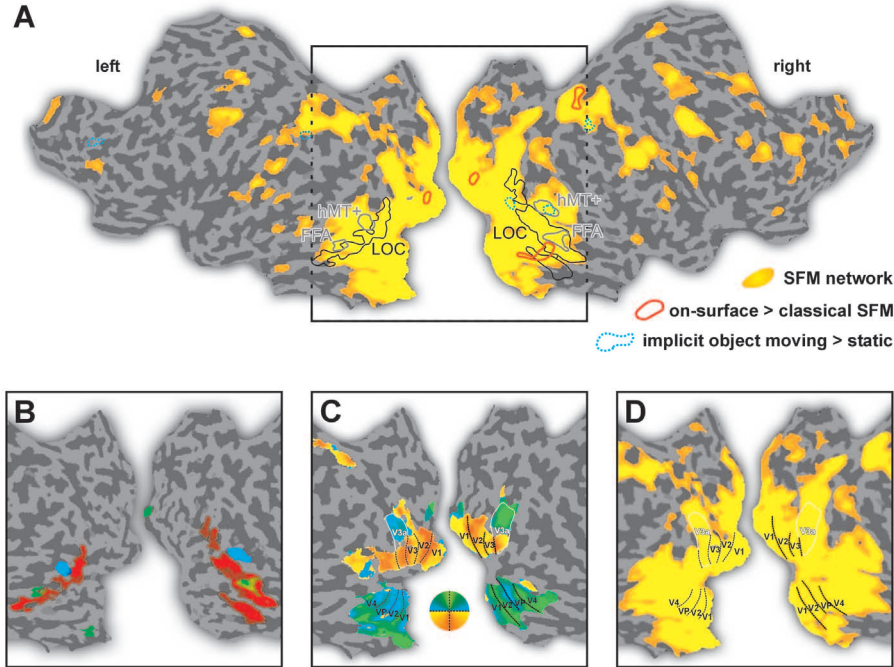
(Figure 2.5). In the Talairach-space group analysis, an isolated, highly motion-sensitive and appropriately located region was assumed to be V3a.) Like other early visual areas, V3a responded more strongly to on-surface than to classical SFM stimuli. This effect may partly be due to low-level properties of the motion flowfield which differed between classical and on-surface SFM. In V3a, in fact, the on-surface-SFM-matched motion control also elicited slightly stronger activity than the classical-SFM-matched motion control, though this effect was not significant. SFM stimuli, on average, did not drive V3a more strongly than motion controls. Taken together, these results suggest that V3a is not particularly sensitive to the surface structure implicit to SFM stimuli.

### Human motion complex

HMT+, as localized individually in each subject, showed an even more pronounced motion selectivity than V3a, responding very strongly to all moving-dot stimuli, while static dots evoked only a very weak response. Within the moving-dot conditions, the main effect is that of implicit-object motion: hMT+ responded much more strongly when the implicit object in on-surface SFM was in motion than when it was stationary (dark versus light gray bars in Figure 2.3). This effect was more pronounced for faces than for random shapes. In order to assess the effect of the SFM encoding (classical versus on-surface), conditions with equal implicit-object motion should be compared. In classical SFM the implicit object is necessarily in motion, and this condition elicits markedly weaker activity than on-surface SFM with a moving implicit object. Interestingly, however, the on-surface-SFM-matched motion control drove hMT+ slightly less than the classical-SFM-matched one (nonsignificant difference, not shown). This suggests that the stronger activity during on-surface than during classical SFM is really caused by the difference in SFM encoding and not by low-level differences of the respective motion flowfields. Consistent with the Talairach-space group analysis (Figure 2.2), the individually localized left hMT+ responded more strongly to on-surface than to classical SFM stimuli (Figure 2.3)—even when the implicit object in on-surface SFM was stationary. In summary, hMT+ is sensitive not merely to visual motion as a low-level property of the retinal input, but also to implicit-object motion, SFM encoding and shape of the implicit object. These findings suggest that hMT+, in contrast to earlier stages including V3a, has a central role in the explication of the properties of the implicit object.

### Lateral occipital complex

LOC responded with roughly equal moderate activity to both moving and static random-dot control stimuli. This response may have been elicited by the circular aperture that was present in all stimuli used in this study. The response was significantly larger for SFM stimuli encoding random shapes and much larger for SFM stimuli encoding faces. The response difference between SFM faces and SFM random shapes was much larger than that between SFM random shapes and control stimuli. LOC's response to our SFM stimuli, thus, was object-selective, but also—and more pronouncedly—face-selective. This is consistent



**Figure 2.5: SFM-object-recognition network (single-subject results)**

Results for subject JS presented on flatmaps of the cortical hemispheres. Dark and light gray regions roughly correspond to sulci and gyri, respectively (dark gray indicates concave, light gray convex shape of the cortical surface in its original folded state). Thresholded statistical maps are superimposed in color. Panel (A) closely parallels Figure 2.2 (group results). Orange to yellow regions were significantly active during SFM object recognition ( $p < 0.05$ , corrected). For abbreviations, see Figure 2.2. Regions outlined in red were significantly more active during on-surface than during classical SFM perception (t-test,  $p < 0.05$ ). Regions significantly more active during classical than during on-surface SFM were not found (reverse contrast, same threshold). Dotted blue outlines mark regions significantly more active when the implicit object was moving than when it was stationary in on-surface SFM perception. Outlined in gray and black are the key regions hMT+, LOC and FFA (refer to labels) as defined by separate localization experiments. Panel (B) shows the results of the localization experiments separately. The red map shows where object photos elicit a significantly stronger response than scrambled object photos. The green map shows where face photos elicit a significantly stronger response than object photos. The blue map shows where moving dots elicit a significantly stronger response than static dots. (For details, see Stimuli and Experimental designs and tasks in Materials and methods.) Panel (C) shows the result of a separate retinotopy-mapping experiment. At each location on the cortical surface map, the color represents the visual angle, at which a wedge stimulus maximally drives that location. The color disk shows how the colors on the map relate to visual field angle. The boundaries of early visual areas have been drawn manually based on the statistical map. (For details on stimuli, measurements and analysis, see Retinotopy mapping paragraphs in the respective subsections of Materials and methods). Area V3a as marked here may include area V3b/KO. In panel (D) the early visual areas' boundaries have been superimposed to the statistical map showing the SFM-object-recognition network. Significance thresholds all satisfy  $p < 0.05$  (corrected). (For statistical details, see Statistical analysis in Materials and methods.)

with its role as described in the literature (Malach, 1995; Kourtzi and Kanwisher, 2001) in that random shapes are not natural objects. Furthermore, even in comparison to natural objects (houses), faces have previously been found to elicit a stronger LOC response (e.g. Levy et al., 2001).

### **Fusiform face area**

FFA displayed an even more clearly face-selective response profile. FFA responses to SFM random shapes were almost at baseline level, close to those elicited by moving and static random-dot control stimuli. The right FFA responded slightly more strongly to SFM faces than the left.

The SFM encoding effect already mentioned (red outlines in Figure 2.2) is also reflected in the key regions' response profiles: On-surface SFM tended to evoke slightly stronger activity than classical SFM, except in FFA and for random shapes in LOC.

### **Location of peak SFM face-selectivity in relation to FFA**

FFA responded selectively to faces defined by SFM. If the representation in this part of the ventral stream is cue-invariant and FFA is the sole face-selective region, then SFM face selectivity should peak at the same point as photo face selectivity: at the center of FFA.

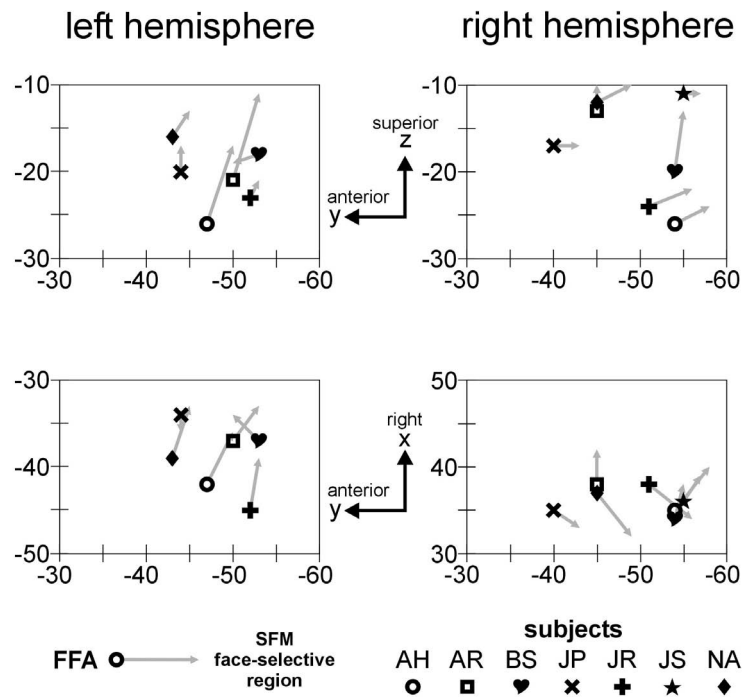
To test this prediction, we mapped the contrast between SFM face and object conditions (pooling classical and on-surface SFM conditions) and determined the peak of SFM face selectivity. The results are shown in Figure 2.6 (Talairach coordinates in Table 2). The peak SFM face selectivity roughly coincided with FFA in only one subject. Euclidean distance between FFA and peak SFM face selectivity ranged between 3 and 12 mm and the shifts appear somewhat consistent across subjects: The SFM face-selectivity peak is superior, posterior and medial to FFA. A regular kind of head movement (e.g. sinking deeper into the padding) relative to the scanner bore between measurement runs can neither explain the symmetry of the shifts with respect to the medial plane nor the consistency of the shifts across subjects because the order of localization and main experiments varied across subjects. Furthermore motion correction was performed for the functional volumes and the alignment between functional and anatomical volumes was visually validated.

### **The effect of the motion of the implicit object**

In contrast to classical SFM, where the implicit object has to be in motion for the stimulus to evoke a 3-D surface percept, the implicit object in on-surface SFM can be either stationary or moving. The state of motion of the implicit object in on-surface SFM can be varied with minimal effects on low-level properties of the motion flowfield encoding the surface—allowing identification of higher-order regions involved in the representation of object motion. To find out whether

there are regions whose response depends on the motion of the implicit object, we contrasted on-surface SFM conditions with the implicit object in motion or stationary.

A bilateral IPS region responding more strongly to on-surface SFM stimuli when the implicit objects moved was found in the group analysis (Figure 2.4, top) and in the individual analyses of most subjects (Table 3). This region responded only weakly to on-surface SFM stimuli encoding the same implicit objects not moving, despite the fact that the low-level properties of the dot motion are almost identical in the two types of stimulus. The only other region consistently responsive to implicit-object motion was hMT+. As described above, hMT+ responded strongly to all moving dot stimuli. However, there was an increase in activity whenever not just the dots, but also the object they encoded moved (Figure 2.4). This implicit-object-motion effect is markedly stronger in both the IPS region and hMT+ for faces than for random shapes (Figure 2.4, bottom). There was no region responding more strongly to stationary than to moving implicit objects in on-surface SFM.



**Figure 2.6: Talairach locations of FFA and SFM face-selective regions in each subject**

FFA as localized with photo stimuli is not identical to the peak face-selective region determined with SFM stimuli. Shift vectors point from the photo face-selectivity peak (i.e. FFA) to the face-selectivity peak obtained by contrasting SFM faces and random shapes. The shift vectors have been projected onto the sagittal plane (top row) and the transversal plane (bottom row). Axes represent Talairach-space coordinates.

## Discussion

### Structure from motion can activate high level object selective regions

SFM stimuli can engage object-selective ventral stream regions including LOC and FFA. SFM stimuli of face surfaces can elicit a response in FFA even when a circular aperture hides the outer contour of the head. This lends support to the view that high-level object-selective responses can be elicited by motion as the sole cue to structure. It further supports the idea of FFA as a region always active when a face is subjectively perceived. The activation of FFA by our stimuli cannot be explained in terms of the curvature properties of the surfaces because random control surfaces of similar curvature properties did not drive FFA more strongly than moving or static random-dot displays.

This finding is consistent with the electrophysiology and imaging literature indicating that inferior temporal cortex in monkeys (e.g. Tanaka, 2000) and its putative human counterpart, ventral temporal cortex (e.g. Haxby et al., 2001), contains complex object representations, that are somewhat cue-invariant (Stry et al., 1993, Amedi et al., 2001). It is in contrast to the report by Sereno et al. (2002) that inferior temporal regions TE and TEO do not respond to 3-D shapes defined by various visual cues including motion. However, Sereno et al. (2002) used simpler and less behaviorally relevant shapes and they studied anesthetized monkeys, not alert humans.

### SFM faces elicit stronger responses than SFM random shapes

During SFM face perception, activity was greater than during SFM random-shape perception—not only in FFA, but throughout the ventral stream and even in hMT+ under certain conditions. This effect can be attributed to the fact that in natural vision faces are more frequent and more behaviorally relevant than random shapes.

The random shapes we used had curvature properties matched to the faces, so the faces were more natural in global shape but not in local shape features. This suggests a bottom-up explanation in terms of shape templates forming a basis system optimized for the representation of complex natural shapes including faces. For the ventral stream regions, this bottom-up explanation is in line with the dominant view in the literature. Given our results it appears compelling because the effect of surface type (face or random shape) was very pronounced in the ventral stream.

However, a top-down mechanism with differential effects during face and random-surface conditions is also plausible—especially as the task was face-nonface categorization. For on-surface SFM conditions, subjects reported that faces appeared well-defined and rigid, whereas the random-shape condition elicited percepts somewhat ambiguous with respect to surface structure and sometimes nonrigid. Below we argue that prior knowledge about the shape of

faces may be used to disambiguate surface representations in hMT+ through feedback.

Future SFM-object-recognition studies may obtain complementary information by using natural objects instead of curvature-matched random shapes as controls—though this will almost inevitably make curvature content a confound.

## On-surface SFM elicits a stronger response than classical SFM

On-surface SFM was generally associated with slightly greater activity than classical SFM (red outlines in Figure 2.2 show where this effect is significant). Low-level differences of the motion flowfield may contribute to but cannot completely explain this effect (see Results).

At the most general level, this SFM encoding effect appears to be related to mental effort. Although all subjects performed the face-nonface categorization with 100% accuracy during the imaging sessions, subjects reported that they found on-surface SFM perception more difficult. The human visual system may be less well adapted to the challenge of computing structure from on-surface motion (e.g. water or light-shadow contours moving across a surface) is a rare phenomenon, whereas classical SFM processing contributes to perception whenever there is relative motion between the observer and the object viewed. If bottom-up computation of surface structure is challenging, the process may depend more strongly on feedback disambiguation. This would explain why on-surface SFM random shapes were sometimes perceived as nonrigid and ambiguous in terms of shape.

The notion of a mental-effort effect is vague in that it merely suggests that the same type of processing has to be carried out more vigorously. Computing surface structure from on-surface motion, however, is likely to require not just greater computational vigor but altogether different computations—which may be less efficiently organized in the visual system.

**Table 1: Average Talairach locations of the individually localized key regions**

| Region | Left        |             |             | Right      |             |             |
|--------|-------------|-------------|-------------|------------|-------------|-------------|
|        | $x \pm SD$  | $y \pm SD$  | $z \pm SD$  | $x \pm SD$ | $y \pm SD$  | $z \pm SD$  |
| hMT+   | $-43 \pm 5$ | $-68 \pm 4$ | $-3 \pm 2$  | $46 \pm 3$ | $-62 \pm 3$ | $0 \pm 3$   |
| LOC    | $-40 \pm 6$ | $-71 \pm 5$ | $-11 \pm 3$ | $40 \pm 3$ | $-71 \pm 3$ | $-8 \pm 6$  |
| FFA    | $-39 \pm 4$ | $-49 \pm 4$ | $-20 \pm 3$ | $36 \pm 2$ | $-49 \pm 6$ | $-18 \pm 6$ |

Talairach coordinates are averages of individually determined locations. The individual locations (specified in Table 2 for LOC and FFA) are centroids of a contiguous set of suprathreshold voxels. At the threshold used to define the regions the statistical map as a whole has a false-positive risk of  $p < 0.05$  (one-sided t-test, Bonferroni-corrected).

### **An intraparietal sulcus region may represent object motion abstracting from the retinal motion flowfield**

We found a region in the IPS that directly reflects implicit-object motion during on-surface SFM perception (Figure 2.4). This region responds most strongly to moving-implicit-object on-surface SFM, less to static-implicit-object on-surface SFM, and still less to classical SFM.

Braddick et al. (2000) find a similarly located IPS region active during coherent motion stimulation. Orban et al. (1999) find IPS regions active during SFM moving-object perception. While these findings are consistent with ours, they do not explain why a moving implicit object elicits stronger activity than a static one. This comparison is only possible with on-surface SFM stimuli, because in classical SFM the implicit object is necessarily in motion. The fact that retinal motion flowfields in the on-surface SFM conditions contrasted are extremely similar in terms of low-level properties suggests that implicit-object motion is really the crucial factor.

A related possibility that deserves to be considered is that the IPS region serves a function which depends on implicit-object motion. An example of such a function is attentive tracking. Culham et al. (1998) have suggested an attentive-tracking function for an IPS region of similar Talairach coordinates. Corbetta et al. (1995; 1998) describe similarly located IPS regions involved in attention and eye movements. Petit et al. (1999b) describe an IPS region involved in the control of smooth-pursuit eye movements. They identify this region as belonging to the parietal eye field. Since we did not perform eye tracking inside the scanner, we cannot exclude the possibility that implicit-object motion elicited low-amplitude smooth-pursuit eye movements—despite the fixation-friendly design of our SFM stimuli. Note, however, that the smooth-pursuit-related IPS region described by Petit et al. (1999b) is clearly removed from ours in Talairach space (14mm and 23mm for left and right subregions, respectively; these are significant distances given intersubject variability). Though our IPS region appears to be separate, implicit-object motion processing is likely to be closely coupled to smooth-pursuit-related processing—regardless of whether eye movements actually occur or not.

### **HMT+: motion flowfield and surface depth map**

The function most frequently attributed to hMT+ is the representation the visual motion flowfield. This view is supported by numerous studies in monkeys (e.g. Maunsell and Van Essen, 1983a; Pack and Born, 2001) as well as humans (Tootell et al., 1995; Goebel et al., 1998; for a review see Culham et al., 2001). Our results support the notion that hMT+, in addition to representing the motion flowfield, explicates more abstract information on object motion and shape.

Motion flowfield information is explicitly represented even in lower visual areas including V3a, and putative V3b/KO. The latter region has been shown to be sensitive not only to first-order but also to different types of second-order motion (Orban et al., 1995; Van Oostende et al., 1997; Smith et al. 1998). Here we subsumed V3b/KO under V3a as it may be a part of it (Singh et al., 2000) and we do not have the data to distinguish the two.

**Table 2: Individual Talairach locations of LOC and FFA in relation to peak face- and object-selective regions localized with SFM stimuli**

| Left hemisphere                          |     |     |     |         |                        |     |     |     |      | Right hemisphere                         |      |     |     |         |                        |     |     |      |  |
|------------------------------------------|-----|-----|-----|---------|------------------------|-----|-----|-----|------|------------------------------------------|------|-----|-----|---------|------------------------|-----|-----|------|--|
| Object-selective region (SFM experiment) |     |     |     |         | LOC (Photo experiment) |     |     |     |      | Object-selective region (SFM experiment) |      |     |     |         | LOC (Photo experiment) |     |     |      |  |
| x                                        | y   | z   | rad | ←-dist→ | x                      | y   | z   | rad |      | x                                        | y    | z   | rad | ←-dist→ | x                      | y   | z   | rad  |  |
| A.H.                                     | -44 | -73 | -11 | 5.5     | 3.2                    | -41 | -73 | -12 | 6.3  | 47                                       | -73  | -17 | 4.7 | 6.2     | 42                     | -70 | -15 | 10.2 |  |
| A.R.                                     | -40 | -75 | 3   | 10.1    | 9.9                    | -44 | -74 | -6  | 7.0  | 32                                       | -70  | 2   | 9.0 | 7.2     | 38                     | -74 | 2   | 7.8  |  |
| B.S.                                     | -41 | -76 | -6  | 13.8    | 5.1                    | -42 | -72 | -9  | 9.5  | 38                                       | -62  | -12 | 9.4 | 7.1     | 37                     | -69 | -11 | 8.6  |  |
| J.P.                                     | -35 | -80 | -10 | 6.6     | 6.0                    | -31 | -78 | -14 | 5.3  | 39                                       | -75  | -12 | 8.3 | 6.6     | 44                     | -72 | -9  | 6.4  |  |
| J.R.                                     | -48 | -67 | -10 | 7.0     | 5.2                    | -47 | -68 | -15 | 5.4  | 43                                       | -64  | -15 | 6.9 | 14.0    | 37                     | -76 | -11 | 9.7  |  |
| J.S.                                     | -34 | -70 | -14 | 2.8     | 5.9                    | -39 | -71 | -11 | 12.2 | 38                                       | -61  | -14 | 4.7 | 15.9    | 41                     | -68 | 0   | 10.3 |  |
| N.A.                                     | -45 | -69 | -14 | 1.4     | 14.2                   | -33 | -62 | -11 | 5.0  |                                          | n.l. |     |     |         | 39                     | -69 | -11 | 1.4  |  |

| Face-selective region (SFM experiment) |     |      |     |         |                        |     |     |     |     | Face-selective region (SFM experiment) |     |     |     |         |                        |     |     |     |  |
|----------------------------------------|-----|------|-----|---------|------------------------|-----|-----|-----|-----|----------------------------------------|-----|-----|-----|---------|------------------------|-----|-----|-----|--|
| FFA (Photo experiment)                 |     |      |     |         | FFA (Photo experiment) |     |     |     |     | FFA (Photo experiment)                 |     |     |     |         | FFA (Photo experiment) |     |     |     |  |
| x                                      | y   | z    | rad | ←-dist→ | x                      | y   | z   | rad |     | x                                      | y   | z   | rad | ←-dist→ | x                      | y   | z   | rad |  |
| A.H.                                   | -36 | -50  | -17 | 4.8     | 11.2                   | -42 | -47 | -26 | 3.0 | 40                                     | -58 | -24 | 6.8 | 6.7     | 35                     | -54 | -26 | 1.4 |  |
| A.R.                                   | -33 | -63  | -11 | 4.2     | 11.2                   | -37 | -50 | -21 | 6.2 | 42                                     | -45 | -10 | 2.8 | 5.0     | 38                     | -45 | -13 | 6.2 |  |
| B.S.                                   | -34 | -50  | -19 | 5.0     | 4.3                    | -37 | -53 | -18 | 3.5 | 38                                     | -55 | -13 | 8.1 | 8.1     | 34                     | -54 | -20 | 5.9 |  |
| J.P.                                   | -36 | -44  | -17 | 4.7     | 3.8                    | -34 | -44 | -20 | 4.8 | 33                                     | -43 | -17 | 4.4 | [3.6]   | [35                    | -40 | 17] | 0   |  |
| J.R.                                   | -39 | -53  | -21 | 7.2     | 6.4                    | -45 | -52 | -23 | 4.3 | 34                                     | -56 | -22 | 4.6 | 6.7     | 38                     | -51 | -24 | 5.5 |  |
| J.S.                                   |     | n.l. |     |         |                        | -39 | -54 | -19 | 1.2 | 39                                     | -57 | -11 | 6.4 | 3.6     | 36                     | -55 | -11 | 2.8 |  |
| N.A.                                   | -33 | -46  | -13 | 2.0     | 7.0                    | -39 | -43 | -16 | 2.4 | 32                                     | -49 | -10 | 2.5 | 6.7     | 37                     | -45 | -12 | 3.0 |  |

Talairach coordinates specify centroids of a contiguous set of suprathreshold voxels. At the threshold used to define the regions, the statistical map as a whole has a false-positive risk of  $p < 0.05$  (one-sided t-test, Bonferroni-corrected). The Euclidean distance (dist, in mm) between corresponding selectivity peaks found in the photo localization experiments and in the SFM experiment is specified. To give an idea of the size of the regions in relation to the distances between corresponding peak selectivities, the radius (rad, in mm) of a sphere of the same volume as the region of interest is specified. (The volume of the region of interest, thus, is  $v=3/4 \pi r^3$ .) The distance and volume measurements were performed individually for each subject, but in Talairach space. Though no spatial smoothing was performed in the individual analysis, activity volumetric measures like these are rough estimates: for any given  $p$ -threshold, they are somewhat dependent on the effect size and the amount of data acquired. Brackets ([.]) indicate values that have been obtained by lowering the threshold below the 5-% significance level. The abbreviation “n.l.” (not localizable) indicates that the region could not be localized (effect absent or insignificant).

Our results show that activity in hMT+ but not V3a reflects what might be thought of as third-order motion—the motion of the SFM-implicit object—as well as object shape information.

Our findings are consistent with early electrophysiological results in the monkey. Monkey MT cells have been shown to carry information not merely about motion direction and velocity, but also about depth. About two thirds of the cells exhibit binocular-disparity-dependent responses (Maunsell and Van Essen, 1983b), almost half of them may be selective for motion-defined surface orientation (Xiao et al., 1997), and there is evidence that these cells’ activity determines the visual depth percept (DeAngelis et al., 1999). Bradley et al. (1998) presented a transparent rotating random-dot cylinder SFM stimulus to monkeys. They have shown how MT cells with near and far depth-selectivity reflect the monkey’s interpretation of this ambiguous stimulus. If hMT+ cells, like monkey MT cells, represent depth and surface orientation, then hMT+ contributes to the representation of object shape. The idea that hMT+ represents depth structure in SFM perception is supported by the findings of Orban et al. (1999).

**Table 3: Individual and average Talairach locations of the implicit object motion IPS region**

| Subject | Left            |          |          |          |  | Right           |          |          |          |
|---------|-----------------|----------|----------|----------|--|-----------------|----------|----------|----------|
|         | <i>t</i>        | <i>x</i> | <i>y</i> | <i>z</i> |  | <i>t</i>        | <i>x</i> | <i>y</i> | <i>z</i> |
| A.H.    | Not localizable |          |          |          |  | Not localizable |          |          |          |
| A.R.    | 5.2             | −33      | −48      | 48       |  | 4.4             | [29      | −50      | 49]      |
| B.S.    | 2.8             | [−23     | −57      | 50]      |  | 3.6             | [33      | −46      | 54]      |
| J.P.    | Not localizable |          |          |          |  | Not localizable |          |          |          |
| J.R.    | 4.4             | [−23     | −57      | 46]      |  | Not localizable |          |          |          |
| J.S.    | 5.2             | −36      | −45      | 46       |  | 5.2             | 39       | −42      | 46       |
| N.A.    | 2.8             | [−23     | −52      | 53]      |  | 3.2             | [27      | −49      | 53]      |
| Group   | 5.2             | −30      | −48      | 48       |  | 5.2             | 30       | −48      | 50       |

Talairach coordinates (*x*, *y*, *z*) of centroids of a contiguous set of suprathreshold voxels. The threshold *t*-value used to define the region is specified together with the *p*-value indicating significance of the map as a whole (one-sided *t*-test, Bonferroni-corrected for multiple comparisons). Brackets ([.]) indicate values that have been obtained by lowering the threshold below the 5-% significance level (corrected). The threshold used in these cases was  $p < 0.003$  (uncorrected). The words 'not localizable' indicate that the region could not be localized and may not be present.

Taken together, the evidence suggests a dynamic model of structure from motion, in which hMT+ initially represents merely the motion flowfield, with near and far cells equally contributing to the representation but not yet reflecting the structure of the surface. Recurrent processing within hMT+ (cf. Andersen and Bradley, 1998), in interaction with early retinotopic areas and ventral-stream regions embodying the constraints of prior shape knowledge of natural objects, may then lead to the formation of a surface representation—with depth at each location coarse-coded in the activity pattern across hMT+ cells of varying depth selectivity.

Like the IPS region discussed above, hMT+ responded slightly more strongly when not merely the dots but also the object they encoded moved. A possible explanation for this finding is feedback from the IPS region. This leaves unexplained how hMT+ can host the implicit-object motion representation in addition to the motion-flowfield and surface representations. A more parsimonious explanation follows from the model suggested above. If hMT+ represents surface structure as a depth map, a moving implicit object will require constant updating of the depth values. At any given location, the coarse-coded depth representation changes across time as the implicit object moves. Thus, a larger population of cells (not just cells of one depth selectivity per location but many) will come to be excited over a complete cycle of implicit-object motion. Each subpopulation will be active only for a shorter period of time, thus reducing adaptation effects and increasing overall activity.

In moving-implicit-object on-surface SFM conditions, hMT+ responded more strongly to faces than to random shapes. Is the representational machinery in hMT+ specialized for the structure of natural objects (cf. Kourtzi et al., 2002)? The evidence for this hypothesis is not compelling. Since the effect is weak and restricted to moving-implicit-object on-surface SFM, it is more parsimoniously

explained by feedback from ventral-stream regions contributing prior knowledge to the hMT+ depth-map representation. Of all SFM conditions, moving-implicit-object on-surface SFM poses the greatest challenge to the visual system and is, thus, likely to be most dependent on feedback disambiguation of surface structure, which can take effect only in the face conditions. The findings of Kourtzi et al. (2002) are also consistent with the depth-map interpretation, as the intact-versus-scrambled object activity difference they found in hMT+ depended on the degree to which depth cues (including binocular disparity and shading) were present in the stimuli.

### **LOC: shape representation specialized for natural objects**

LOC responded more strongly to SFM faces than to SFM random shapes. This is consistent with its role as described in the literature (Malach, 1995; Kourtzi and Kanwisher, 2001) in that random shapes are not natural objects. Furthermore, even in comparison to natural objects (houses), faces have previously been found to elicit a stronger LOC response (e.g. Levy et al., 2001).

### **FFA: cue-sensitive face representation**

FFA as localized with photo stimuli shows a clear face-selective response to SFM stimuli. This shows that there is some degree of cue invariance to the representation in FFA. However, the peak of selectivity for SFM faces is considerably shifted with respect to the peak of selectivity for face photos in all subjects but one (JP). These shifts are 3 to 12 mm large and their directions in Talairach space are somewhat consistent across subjects (Table 2, Figure 2.6). This suggests that the representation in this part of the ventral stream does not completely abstract from the visual cue defining the object. Either FFA has a more highly face-selective neighbor under SFM conditions or its internal response pattern reflects the object-defining cue. More generally, FFA may not be an area in the sense of a patch of cortex with a boundary clearly and consistently defined by both function and cytoarchitectonics as are the boundaries of early visual areas. Instead, FFA may be a fuzzy region, in which face-selectivity peaks for different defining cues form a cluster (cf. Haxby et al., 2001). The present results, in our opinion, are more readily accommodated under this latter view.

It must be noted that these conclusions are tentative as the baseline conditions were not well-matched between photo and SFM experiments. The control objects were everyday objects in the photo experiment and random shapes in the SFM experiment. Future studies could employ SFM versions of natural objects or shape-from-shading representations of faces and random shapes (cf. Figure 2.1C) to further explore these issues.

## Acknowledgments

This research was funded by the Universiteit Maastricht (Maastricht, The Netherlands) and the Donders Centre for Cognitive Neuroimaging (Nijmegen, The Netherlands). The authors thank Elia Formisano and two anonymous reviewers for their helpful comments on a draft of this paper.

## Chapter 3

# Information-based functional brain mapping

**Corresponding publication:** Kriegeskorte N, Bandettini P, and Goebel R. Information-based functional brain mapping. *In preparation*.

**Abstract:** We propose a novel method of functional brain mapping based on functional magnetic resonance imaging data. The currently dominant brain mapping approach is *activation-based* in that it localizes regions that are activated as a whole in one condition as compared to a control condition. Our approach, by contrast, is *information-based* in that it localizes regions whose intrinsic spatial response pattern contains information about the experimental condition, thus addressing a more general question. The information in local contiguous response patterns is mapped out by scanning the imaged volume with a spherical “searchlight”, whose contents are analyzed multivariately at each location in the brain.

The idea of distributed representations at the spatial scale accessible to fMRI has recently become important in the field of neuroimaging. Our method allows discovery and localization of such representations. While we maintain the important idea of *mapping* the brain for functional regions at a coarse spatial scale, at which interindividual consistency is to be expected, we can detect individually unique fine-grained representations by exploiting high-spatial-frequency information, which is provided by current fMRI (1-2 mm voxel width) but filtered out by conventional analysis in the interest of noise reduction and statistical integration across subjects.

## Introduction

Distributed activity patterns are believed to represent complex perceptual, cognitive and motor information at the microscopic scale of individual neurons. Recent studies (Edelman et al., 1998; Ishai et al. 1999, Haxby et al., 2001; Cox and Savoy, 2003) suggest that distributed codes play a major representational role at the macroscopic scale accessible to fMRI as well. We will refer to such codes as *macropopulation codes*—in distinction to population codes at the level of single neurons: *micropopulation codes*. The cited studies have applied multivariate analysis techniques to analyze predefined regions of interest for distributed information.

Here we consider how to *discover* regions containing a particular type of information, when their location is not known a priori. We will argue that localization of macropopulation codes is not in general possible with classical brain-mapping methods, but requires multivariate measures to be mapped throughout the measured volume. We present locally multivariate methods for 1) localization of regions whose spatial activity patterns reflect differences between the experimental conditions and 2) statistical analysis and visualization of multivariate effects in a given region of interest, including estimation of lower bounds on the information content (in bits) of a region’s multivariate fMRI response.

## Classical activation-based brain mapping

The classical approach to brain mapping based on neuroimaging data aims to localize cognitive functions in the brain by finding activated regions. We refer to this approach as *activation-based*. In the prototypical case, a comparison is performed between two experimental conditions, which differ only by the cognitive component function that is to be localized. Voxelwise subtraction or univariate multiple linear regression is used to find all voxels where activity was greater when the cognitive component in question was present. For example, one might localize visual regions by finding voxels more active while the subject viewed object photos than while the subject viewed a black screen.

To give a more specific example, one might localize the fusiform face region by contrasting the viewing of faces and the viewing of other objects (see e.g. Kanwisher, 1997). In the classical activation-based approach, the effects are standardly assumed to be spatially smooth, leading to activated regions that extend across multiple voxels. This assumption allows spatial smoothing of the data to be used for noise reduction and plays a key role in statistical inference in several established methods including the widespread statistical parametric mapping (Friston et al., 1994; Friston et al., 1995; see also: Poline et al., 1997). The classical methods are sensitive to extended contiguous regions of voxels with homogeneous selectivity (i.e. regions of voxels whose activity during condition A is greater than during condition B, but not vice versa).

If a region contains a macropopulation code, however, it cannot be assumed that one condition (e.g. perception of a face) will elicit a greater response at every voxel of the region than another condition (e.g. perception of a house or a different face). The main functional advantage of population codes is the representational power afforded by the combinatorics of single-unit activities.

One would therefore expect the response patterns to be complex, populating more of the multivariate space than the linear subspace of the regional average. In order to localize macropopulation codes, it is thus necessary to drop the assumption of spatial smoothness of effects. In other words, the focus of the analysis needs to include what is usually considered spatial noise: the high-spatial-frequency content of the signal.

## Information-based brain mapping

Here we argue for an alternative approach to brain mapping that addresses a more general question than the classical activation-based approach: Where in the brain does the local spatial response pattern differentiate the experimental conditions? A region more strongly active during one condition than during the other certainly differentiates the conditions. But so may a region whose spatial-mean activity is the same in both conditions. For example, different voxels of the region may have opposite selectivities (some responding more to condition A, others more to condition B) such that the spatial-mean effect is insignificant. Whether it is the spatial-mean or the fine-grained pattern of the regional activity that differentiates the conditions or both, the region's response pattern contains *information* about the condition.

We refer to any method that localizes such response-pattern information as *information-based* functional brain mapping. Information-based brain mapping is a generalization of activation-based brain mapping, because it is sensitive to both activated regions (of homogeneous selectivity) and regions whose spatial-mean activity stays constant, but whose activity pattern reflects the experimental condition. Whereas activation-based mapping consists in *univariate* analysis performed for each voxel, information-based mapping involves *multivariate* analysis of response patterns.

In order to maintain the important classical idea of mapping the brain in terms of macroscopic regions comprised of multiple voxels, we target *focally* distributed representations, i.e. representations in contiguous regions, as opposed to noncontiguous widely distributed representations. This restriction is theoretically as well as practically motivated. In terms of neuroscientific theory, it is in line with the cytoarchitectonically defined concept of cortical area and with the more general idea of macroscopic functional regions, which lies at the heart of brain mapping and has proven to be a very successful research paradigm. Practically, the restriction to contiguous regions makes the statistical computations tractable with the currently available computer hardware.

To give a concrete example, one might contrast activity during perception of face images and perception of house images to find regions differentiating the two categories. Or one might target a more specific level of representation by contrasting responses evoked by two particular face images (see Chapter 4). However, information-based brain mapping is not restricted to vision or perception. It can be used to localize focally distributed representations in all domains including motor control and higher cognitive function.

The core of this chapter is divided into two sections. In the Motivation section, we introduce the idea of a spherical multivariate searchlight, which is moved

throughout the volume to continuously map multivariate effects. We apply multivariate and information theoretic techniques to synthetic and real fMRI data to demonstrate that this locally multivariate statistical mapping is more sensitive to focally distributed effects than the univariate approach currently dominant in fMRI analysis. To make this point, we rely on simple multivariate statistics.

In the Method section, we formally define a more advanced implementation of information-based functional brain mapping. Our method utilizes the assumption of Gaussian noise, which is widely accepted as reasonable in univariate fMRI analysis. However, in contrast to the classical activation-based approach, we jointly analyze responses within the spherical searchlight at each location and model the local spatial noise covariance as a multivariate normal distribution. Taking local noise covariance into account makes our parametric information-based brain mapping method more powerful than the nonparametric methods used to motivate the locally multivariate approach. The parametric method of information-based brain mapping defined in this section has been applied in the context of an object-recognition fMRI experiment. This application is described in Chapter 4.

## Motivation: Why is a new approach needed?

### Focally distributed effects elude activation-based univariate brain mapping

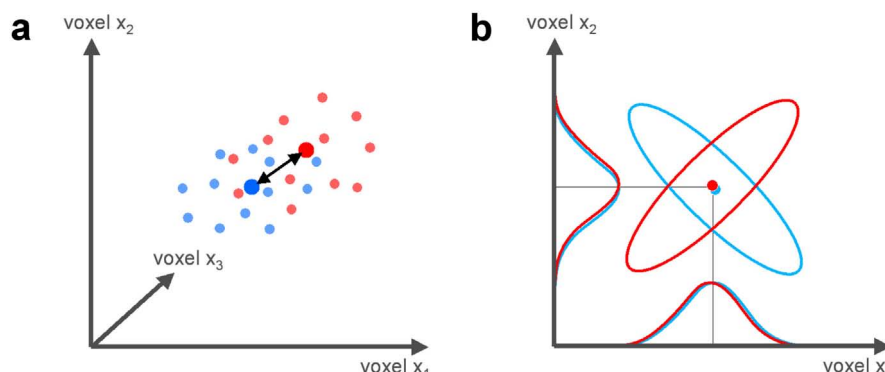
#### Multivariate phenomena can be completely invisible to univariate analysis

Consider a single region of the brain of a single subject. The region may correspond to 30 voxels in an fMRI data set. For simplicity, let us consider the case of two conditions of a slow-event-related experiment. We can estimate the spatial response pattern elicited by each trial of the two conditions in our region of interest.

Figure 3.1a shows spatial response patterns visualized as points in multivariate space. Figure 3.1b illustrates that the reduction of complexity implicit to considering only the univariate projections of a multivariate data set can completely obscure the difference between the two conditions, which consists in the distribution of responses in multivariate space here. In the multivariate dataset the means are identical but the distributions (bivariate Gaussian) have different orientations. In each univariate projection, in contrast, the means as well as the distributions are identical.

The example of Figure 3.1b illustrates the inevitable loss of information associated with reducing a multivariate dataset to its univariate projections. However, it is of limited relevance under homoscedasticity, i.e. when the distributions are the same across all conditions. The assumption of homoscedasticity is widely accepted for fMRI data and we rely on it in the methods to be presented here.

When the distributions are the same for all conditions, differences between conditions can lie only in the condition means. A multivariate condition mean is



**Figure 3.1: A multivariate phenomenon is more than the sum of its univariate components**

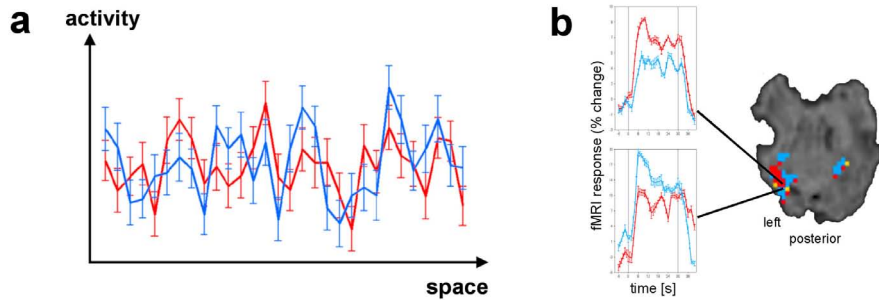
(a) A spatial response pattern of a given brain region can be thought of as a point in a multidimensional space spanned by the voxels' activities. In this invented illustration, the spatial activity pattern during each trial of condition A is shown as a small red dot (blue for condition B). The larger dots are averages across all trials of a given condition.

(b) Multivariate phenomena can be completely invisible to univariate analysis. Each ellipse corresponds to an iso-probability-density contour of a bivariate Gaussian distribution. The bell-shaped curves adjacent to the axes show the projections of the bivariate probability density functions onto the two dimensions of the space. They are univariate Gaussian probability density functions. The two distributions (red and blue) differ in multivariate space, but are identical in each of their univariate projections. (The bell-shaped curves are meant to coincide, but have been shifted slightly with respect to each other, so both are discernable.)

an activity pattern corresponding to a point in multivariate space (large dots in Figure 3.1a). When multivariate condition means differ, this necessarily implies differences in the univariate projections. In the absence of noise, therefore, considering only univariate projections would not obscure any effects. But although multivariate effects, under homoscedasticity, will necessarily leave a trace in the univariate projections, they can be much harder to detect in the presence of noise—as the following paragraphs will show.

### **Weak-at-every-voxel distributed effects can elude voxelwise univariate testing**

Figure 3.2a shows two simulated spatial activity patterns (e.g. ventral-temporal-cortex responses to two face images). For simplicity, space is represented here as a single dimension. The difference between the conditions is not significant (overlapping error bars) at *any* single voxel. In a classical voxelwise univariate multiple linear regression analysis, therefore, no effect would be detected. (This is the case even though correction for multiple comparisons, which would require a more conservative threshold, has been omitted.) A multivariate analysis of variance (MANOVA), however, reveals that the two patterns of activity are in fact significantly different ( $p < 0.05$ ). The reason, in this example, is that the univariate effects are close to significant at many voxels making the multivariate difference unlikely under the null hypothesis.



**Figure 3.2: Distributed effects elude classical activation-based mapping when they are weak at every voxel and inhomogeneous**

(a) Mean activity (vertical axis) of 30 simulated voxels (horizontal axis) during each of two conditions (red line, blue line). The error bars show the standard error of the mean. The two conditions do not elicit significantly different activity at any single voxel (overlapping error bars). Therefore a univariate mapping technique would not mark a single voxel. Multivariate analysis of variance, however, shows that the patterns are significantly different ( $p < 0.05$ ). The simulated noise is Gaussian here and there is no statistical dependency between the voxels.

(b) Distributed effects can be inhomogeneous, i.e. they may contain univariate effects of opposite sign. The figure shows a case where adjacent voxels have opposite selectivities. This suggests that spatial smoothing and intersubject averaging, instead of merely removing spatial noise, eliminate essential information in the high-spatial-frequency band of the fMRI signal. The figure shows data obtained while the subject viewed face and object photos (block design, 45 photos per 30s block, Siemens Sonata 1.5 T).

### High-spatial-frequency information is discarded in the classical activation-based approach

What is required to detect the effect in the example of Figure 3.2a is local spatial integration of the univariate effects. In the classical approach to brain mapping, local spatial integration is performed by spatial smoothing of the data. Note, however, that spatial smoothing (i.e. low-pass filtering) would obscure the effects in the example of Figure 3.2a because they reside in the high-spatial-frequency band: adjacent voxels often have opposite selectivities and the spatial mean activity levels are very similar between the two conditions. This situation is frequently observed in actual fMRI data (e.g. Figure 3.2b). As opposite selectivities encode the same information, such voxels should arguably be considered as belonging to the same functional unit. To detect weak-at-every-voxel distributed effects with frequent opposite selectivities in adjacent voxels requires a form of local spatial integration that does not simply average across space. More general methods of considering the voxels jointly are provided by classical multivariate statistics such as multivariate analysis of variance (MANOVA, see e.g. Krzanowski, 1988), which reveals the effect in the example of Figure 3.2a.

### Locally multivariate brain mapping finds focally distributed effects

We propose the use of a spherical multivariate “searchlight” (Figure 3.3) for enhancement of the sensitivity of detection of focally distributed effects. First we

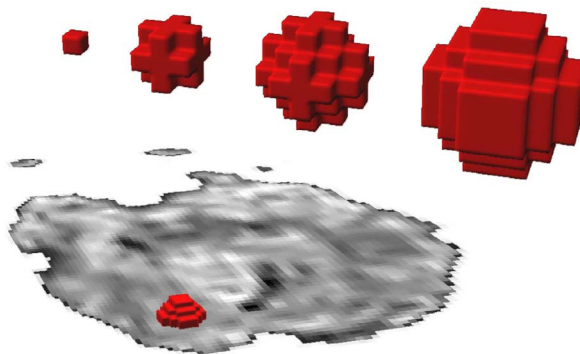


Figure 3.3: **Spherical multivariate searchlight**

A spherical cluster of voxels is used as a region for joint multivariate consideration of distributed effects. The spherical region is moved through the volume like a searchlight. At each voxel, a multivariate statistic is computed for its contents. This results in a continuous map of the local pattern effect.

choose a radius for the multivariate searchlight. The radius defines how many voxels will be considered jointly at each location. The searchlight is centered on each voxel in turn. At each location, a local multivariate effect statistic is computed integrating the signals from all the voxels falling into the spherical searchlight. As in the classical approach, local signals are combined to enhance statistical sensitivity—but in a fundamentally different way: Instead of spatially smoothing the *activity signal* (which would remove high-spatial-frequency information), we locally combine *effects* (abstracting from their direction).

For each voxel in the volume, the resulting map shows how well the multivariate signal in the local spherical region differentiates between the experimental conditions. In this section, we demonstrate that spatial integration by means of this scheme actually enhances the sensitivity for localization of focally distributed effects in an extended space.

## Simulation

We simulated fMRI data as might be obtained in a 20-minute slow-event-related experiment with two conditions. The data were simulated by embedding Gaussian random patterns (a distributed code) in Gaussian white noise. The disc-shaped activity patterns (one for each condition, Figure 3.4) appeared and faded according to a realistic hemodynamic response model (Boynton et al., 1996). Spatially, the data set consisted in a 128-by-128-voxel slice. The details of the simulation are defined at the end of this section.

We mapped the simulated slice using three different statistics: (1) univariate linear regression contrast, (2) local pattern difference, and (3) local pattern information. Statistic (1) corresponds to the classical approach. To give this method its best chance, spatial smoothing has been omitted. (Spatial smoothing would only decrease sensitivity here, because the patterns to be detected are composed of white noise and their local average thus is 0 in the limit as the integration kernel increases in size.) Statistics (2) and (3) are multivariate

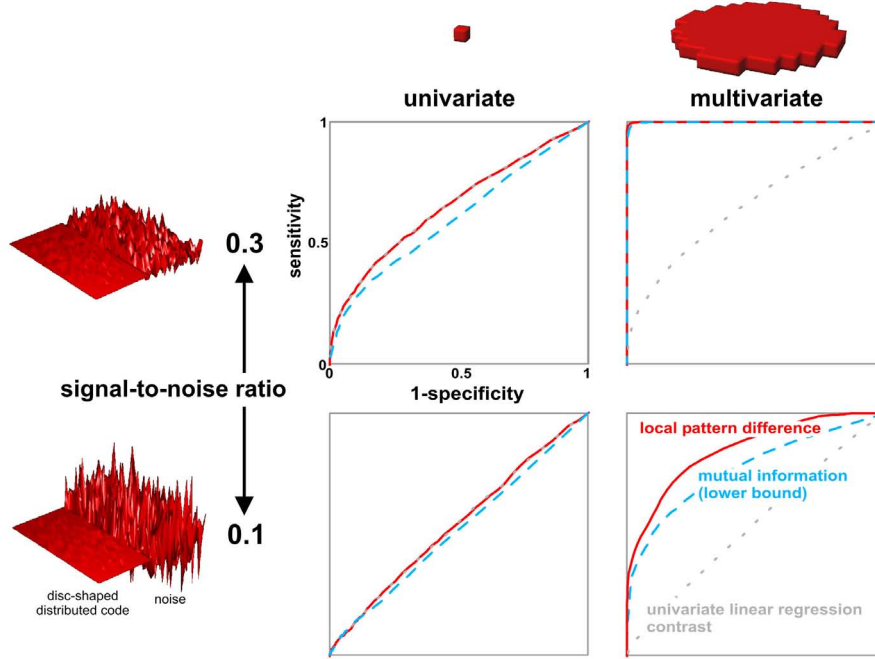


Figure 3.4: **Local-pattern-effect mapping—receiver operating characteristics (simulated data)**

Local-pattern-effect mapping involves local combination of distributed effects. Each panel shows the sensitivity of different multivariate effect statistics as a function of specificity. The figure contrasts multivariate local-pattern-effect mapping (right column) and its degenerate univariate counterpart (left column) for high (top row) and low (bottom row) signal-to-noise ratios. The red surfaces to the left of the four panels show snapshots of the spatial patterns of signal (left) and noise (right) side by side. The signal is a simulated disc-shaped distributed code. In the simulation, signal and noise are superimposed. The gray dotted lines representing univariate regression contrast have been copied to the right column as a visual reference.

statistics defined below. They were used in conjunction with a spherical search-light. All the methods employed predictor functions generated according to the hemodynamic response model of Boynton et al. (1996).

*Local pattern difference.* The local pattern difference has been defined here as the Euclidean distance between the estimated event-related spatial response patterns. The pattern responses are estimated using the least-squares technique with a design matrix composed of hemodynamic response predictors. This estimation is fully defined in Equations (3.1) and (3.2) below.

*Local pattern information.* As a measure of a region's information content, we estimated a lower bound on the mutual information between stimulus and response. The mutual information between two variables (each having a single or multiple dimensions) is a function of the separate and joint entropies of the two variables as illustrated in Figure 3.8 and defined in Equation (3.5).

To obtain a lower bound on the mutual information for the sake of this simulation study, we performed minimum-distance classification of the region's spatial response patterns, predicting the experimental conditions from the noisy data. From the classification matrix, we computed the mutual information between

the experimental conditions (stimulus) and the classifications (response). This method is efficient enough for whole-volume mapping of the information lower bounds.

*Results of simulation.* To be able to compare the three mapping methods as a function of noise, we varied the noise level across the slice and repeated the distributed codes at the same amplitude (the spatial pattern standard deviation is 1 at the maximal hemodynamic response amplitude). The maps shown in Figure 3.5 illustrate that the locally multivariate mapping techniques reveal the focally distributed effects down to much lower signal-to-noise ratios than univariate linear-regression-contrast mapping.

To avoid the bias introduced by the choice of threshold, we computed receiver-operating characteristics (i.e. sensitivity as a function of specificity) for each mapping technique and for two values of the signal-to-noise ratio (0.3 and 0.1). Figure 3.4 shows that at a signal-to-noise ratio of 0.3, spherical-searchlight mapping (121 local voxels considered together) discriminates almost perfectly between regions containing a focally distributed effect embedded in noise and pure noise regions. Univariate linear-regression-contrast mapping performs much worse.

At a signal-to-noise ratio of 0.1, spherical-searchlight mapping can still localize informative regions, whereas the linear-regression-contrast mapping performs at chance level. Within the spherical-searchlight techniques, the information mapping turned out to have slightly lower discrimination power than the local pattern distance mapping.

*Simulation details.* We simulated a slow event-related experiment with two conditions. Each event was simulated to last 500ms and there were 48 trials per condition (96 trials over all). The stimulus sequence was random. Each trial lasted 12s. The simulated temporal resolution of measurement was one volume per 2.4s, so there were 5 measurement volumes per trial. The entire simulated experiment consisted in 482 measured volumes ( $96 \times 5 + 2$  volumes preceding the first stimulus presentation). This corresponds to a run length of 19.28 minutes.

The simulated dataset had two spatial dimensions along each of which it measured 128 voxels (a large square slice, see Figure 3.5). For each of the two conditions, a disc-shaped focally distributed response pattern was generated.

The disc-shaped response patterns were 32 voxels in diameter. Inside the disc region, they were composed of Gaussian noise of standard deviation 1. Outside the disc region, their value was 0. In order to be able to study detectability of the simulated focally distributed code as a function of noise level, the disc patterns were arranged in two horizontal rows (each made up of four discs) within the 128-by-128 simulated data slice (blue circles in Figure 3.5) and the noise level was varied from 1 (standard deviation) at the left edge of the first disc in the top row to 9 at the right edge of the last disc of the bottom row.

The signal-to-noise ratio was defined as the ratio between the standard deviation of the patterns across space (at the points in time when they are maximal) and the standard deviation of the Gaussian white noise.

The spatiotemporal data matrix was computed according to  $VT = VC \times CT + noise$ , where  $VT$  is the voxel-by-time data matrix,  $VC$  is the voxel-by-condition matrix of disc-shaped Gaussian spatial response patterns (one for each of the

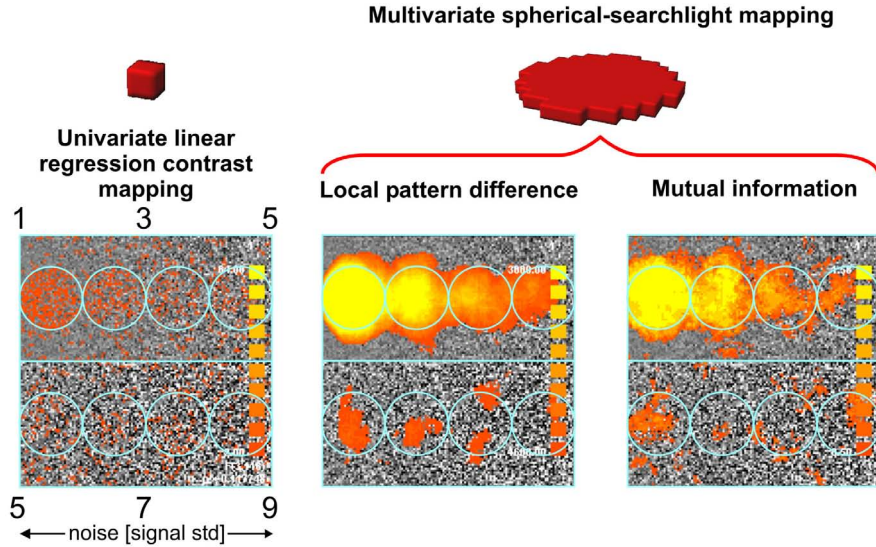


Figure 3.5: **Local-pattern-effect mapping under varying noise (simulated data)**

Descriptive results for the simulated data obtained by linear-regression-contrast mapping and spherical-searchlight mapping. The noise level increases from left to right in two rows, starting at 1 (signal-to-noise ratio = 1) in the top left corner of each panel and ending at 9 (signal-to-noise ratio = 1/9) in the bottom right corner. The blue circles indicate the locations of the simulated disc-shaped distributed codes, which are repeated along the two rows to cover all noise levels. The grayscale background image is a snapshot of the noise pattern. The statistical map is superimposed in color. The threshold has been chosen arbitrarily here. An analysis for all thresholds employing receiver-operating characteristics is shown in Figure 3.4. The maps illustrate that local-pattern-effect mapping using a spherical searchlight can reveal distributed effects with greater sensitivity than conventional univariate mapping.

two conditions),  $CT$  is the condition-by-time matrix defining the time course of fading in and out of the spatial response pattern for each condition, which is obtained by convolution of the stimulus function with the hemodynamic impulse response according to the model of Boynton (1996), and  $noise$  is a  $128 \times 128$ -by-482 matrix of Gaussian white noise, whose standard deviation varied from 1 to 9 with space (but not time) according to the two-row ramp function described above.

### Real fMRI data

The real fMRI data were acquired while subjects passively viewed face and object photos while fixating a central cross. The object images were arranged in blocks of 30s (face and mixed-object blocks alternating), which were separated by 20-s fixation blocks, during which a black background was presented behind the central fixation cross.

We performed local-pattern-effect mapping using a spherical searchlight of 6-mm radius (comprising 19 voxels as shown in Figure 3.6) and the Euclidean distance as a rapidly computable test statistic. Significance thresholds ( $p < 0.05$ , corrected for multiple comparisons) were obtained by a randomization test. This

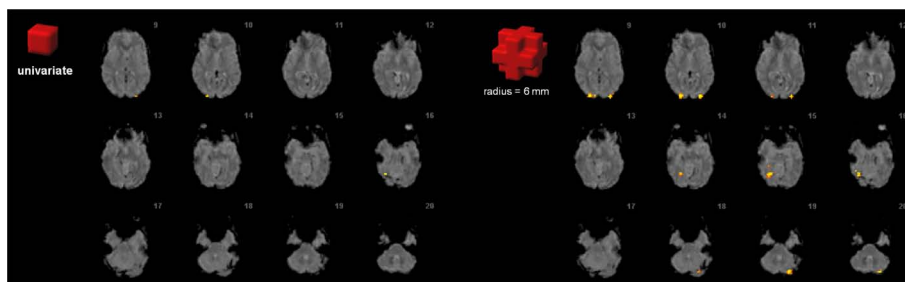


Figure 3.6: **Local-pattern-effect mapping (real fMRI data)**

We performed local-pattern-effect mapping using a spherical searchlight as shown. Significance thresholds ( $p < 0.05$ , corrected for multiple comparisons) were obtained by a randomization test. To elucidate the effect of spatial combination of signals, we used exactly the same method for the multivariate (right) and the degenerate univariate case (left), where the searchlight comprises only a single voxel. Local spatial combination of signals reveals additional regions in left inferotemporal and right occipital cortex. Spatial low-pass filtering prior to univariate analysis did not reveal these regions.

approach is explained in the section Nonparametric approaches of the Discussion. To elucidate the effect of spatial combination of signals across the 19 voxels at each location, we used exactly the same method for the degenerate univariate case, where the searchlight comprises only a single voxel. This method is very similar to univariate regression-contrast mapping—conceptually, mathematically and in terms of the results (not shown).

Both methods highlighted regions in occipital and inferotemporal cortex. The signal in the highlighted regions reflects whether the objects perceived are faces or not. The locally multivariate technique detected two additional informative foci (Figure 3.6) in left inferotemporal and right occipital cortex. This indicates that local spatial combination of signals increases sensitivity.

Would spatial low-pass filtering, the conventional method of local spatial combination of signals, give the same increase of sensitivity? Spatial low-pass filtering with a 4-mm Gaussian kernel did enlarge the foci found univariately, but it did not reveal the additional regions. An analysis of the single-voxel effects (not shown here, but see Figure 3.2b for a similar case) revealed that the selectivity of the voxels in the informative regions is not homogeneous (i.e. there are opposite effects within the same region). Low-pass filtering therefore obscured the distributed effects detected in these two regions by local-pattern-effect mapping.

## Conclusions

Multivariate phenomena including combinatorial neural codes can be completely invisible to univariate analysis—no matter how much data is acquired for sensitivity to weak effects. In fMRI analysis, the structure of the noise is usually assumed to be the same across all conditions (homoscedasticity). Under this widespread and reasonable assumption, multivariate phenomena necessarily leave a trace in univariate projections, but multivariate analysis can still greatly enhance statistical sensitivity. Such an enhancement of sensitivity is needed to detect combinatorial effects in fMRI data particularly when they are

weak at every voxel. Univariately insignificant weak-at-every-voxel distributed effects may be jointly significant.

An enhancement of sensitivity can result from local spatial integration of effects. In conventional analysis, spatial integration is performed by spatial low-pass filtering of the data, which discards information contained in the high-spatial-frequency band of the response patterns and renders the analysis sensitive to spatially homogeneous effects (larger activation blobs), but insensitive to distributed effects with opposite selectivities in adjacent voxels. To increase sensitivity through spatial integration without discarding high-spatial-frequency information, a spherical multivariate searchlight can be used. We obtain a continuous map reflecting focally distributed effects by moving the searchlight through the volume and performing a local multivariate analysis at each location. Simulation showed that this novel approach to functional brain mapping can be much more sensitive to focally distributed effects. Application of the spherical-searchlight method to real fMRI data revealed informative regions not found using conventional voxelwise-univariate regression analysis. This suggests that focally distributed effects actually occur in fMRI data and that our method is more sensitive at detecting them not only within white noise but also within real fMRI noise.

## Method: Parametric information-based functional brain mapping

In this section we formally define a parametric method of information-based functional brain mapping. Our method combines the idea of a spherical multivariate searchlight introduced in the previous section with the power of classical parametric multivariate statistics.

Whereas above we used the Euclidean distance as a measure of the multivariate effect, we will use the Mahalanobis distance here. The Mahalanobis distance (Equation (3.6), see Figure 3.9) takes the covariance structure of the noise across voxels into account. Using the Euclidean distance above served our purpose of studying the effect of spatial integration with a spherical multivariate searchlight. Since the noise used in the simulation was white, i.e. not spatially correlated, the Euclidean distance yielded results equivalent to those we would have obtained using the Mahalanobis distance. For real fMRI data, however, it is known that voxels close to each other have correlated noise. In this situation, modeling the multivariate structure of the noise can bring a further increase in sensitivity. Consider, for instance, a case where the multivariate effect vector is orthogonal to the first principal component of the noise (this is approximately true in Figure 3.9a). In this case a smaller multivariate effect will be significant.

The case of multivariate orthogonality of noise and effect is important because it is realistic and prototypical of the kind of focally distributed effect our analysis targets. If each condition elicits a unique distributed pattern of activity within a region, then the spatial means of the response patterns may be very similar. If a combinatorial code relies on the combinatorics of unit activities (which it owes its name to), the resulting multivariate effect is unlikely to have a major component along the all-one vector (i.e. a major spatial-mean component). At

the same time, as mentioned above, the noise is often positively correlated for voxels within the same neighborhood. The noise is therefore expected to have its principal component along the all-one vector. In this situation of multivariately orthogonal noise and effect, modeling multivariate noise covariance promises increased sensitivity.

Note that this argument is in exact opposition to the logic that justifies the conventional activation-based mapping approach, where the low-spatial-frequency component (including the local average) is considered as the effect of interest and the high-spatial-frequency component is considered as noise. In the activation-based approach, spatial averaging is thus used for noise reduction, which would work optimally if the noise were uncorrelated.

## Estimation of event-related spatial response patterns

To estimate the *event-related spatial response pattern* elicited by each stimulus in a contiguous region of interest (e.g. the fusiform face region), we fit the hemodynamic-predictor model described in the previous paragraph to each voxel's time course. For the region as a whole, the linear model can be stated as follows:

$$Y = XB + E, \quad (3.1)$$

where  $Y$  is the time-by-voxel data matrix,  $X$  is the time-by-condition design matrix,  $B$  is the condition-by-voxel matrix of beta weights (containing a spatial response pattern in each row), and  $E$  is the time-by-voxel matrix of errors. Ordinary least-squares minimization of  $\text{sum}(\text{diag}(E^T E))$  leads to

$$\hat{B} = (X^T X)^{-1} X^T Y, \quad (3.2)$$

where  $\hat{B}$  is the condition-by-voxel matrix of beta estimates. Each row of  $\hat{B}$  contains an estimated spatial response pattern.

## Visualization of event-related spatial response patterns

For a predefined contiguous cortical region in a single subject, the event-related spatial response pattern elicited by a particular stimulus is visualized as a surface (see Chapter 4, Figures 4.3 and 4.5), where the vertical dimension represents the response amplitude (in proportion signal change) and the horizontal plane represents a flatmap of the cortical region. These flatmaps are obtained by means of a self-organizing process that defines a projection from the three-dimensional voxel space onto the unit square. Each voxel is initially placed at a random location on the unit square. The self-organization is driven by two forces that act on the locations of the voxels on the unit square: an attractive force that pulls voxels that are close in 3D together on the unit square and a repulsive force that pushes neighboring voxels on the unit square apart to ensure equal spacing in 2D. The strength of each of the forces is a Gaussian function of the respective distance multiplied by rate factor that decreases exponentially in time. First, each force is computed for each pair of voxels. Then, for each voxel, all the forces acting upon it are integrated and its location on the unit square is adjusted. This sequence is applied iteratively.

The resulting projection preserves 3D-neighborhood relationships in 2D. When the shape of the region of interest reflects the shape of the cortical sheet (as is frequently the case, particularly when the region has been defined on the basis of 7T functional data), then the resulting flatmap constitutes a local approximation to a cortical flatmap as obtained by flattening a polygon-mesh reconstruction of the cortical sheet based on anatomical measurements. Our local-flatmap approach has the advantage that it does not require anatomical measurement, cortex reconstruction, and coregistration of functional and anatomical data sets.

### Multivariate normal noise model

For each region of interest, we model the spatial dependencies of the errors as a multivariate normal distribution defined by the covariance of the errors between voxels across time. The covariance matrix  $\Sigma$  is estimated as  $\hat{\Sigma} = (\hat{E}^T \hat{E})/n$ , where  $\hat{E} = Y - X\hat{B}$  and  $n$  is the number of time points. If there is no cross-lagged correlation, the covariance between two beta estimates (beta estimate for condition  $c1$  and voxel  $v1$  and the beta estimate for condition  $c2$  and voxel  $v2$  is

$$\text{cov}(\hat{B}_{c1,v1}, \hat{B}_{c2,v2}) = \Sigma_{v1,v2} \cdot (X^T X)^{-1}_{c1,c2}, \quad (3.3)$$

where the subscript notation of the form  $M_{i,j}$  denotes the scalar value in row  $i$  and column  $j$  of matrix  $M$ . As the equation shows, for a given condition  $c = c1 = c2$ , the orientation and shape of the multivariate normal distribution describing the variability of the spatial response-pattern estimates in the space spanned by the voxels depends only on the covariance  $\Sigma$  of the errors. The design matrix and condition determine merely a scaling factor of the distribution.

### Multivariate test of difference

In order to test in a single subject whether two stimuli elicit a different response pattern in a predefined region of interest, we use run-A data to formulate a subject-specific hypothesis as to the multivariate dimension and the direction of the effect. We then test this hypothesis using run-B data. More precisely, we estimate the Fisher-linear-discriminant dimension (henceforth Fisher dimension, Equation (3.4), Figure 3.7; Fisher, 1936; see also: Bishop, 1995) based on run-A data and perform a one-sided t-test on run-B data projected onto the run-A Fisher-dimension estimate.

The Fisher dimension is the dimension best separating the two distributions (the dimension, on which the ratio of between-class and within-class variance is maximal). For multivariate normal distributions of equal covariance, the optimal decision boundary for classification is a hyperplane orthogonal to the Fisher dimension (Figure 3.7). Assuming that the predictors of the two conditions (columns of the design matrix  $X$ ) have the same energy and are nearly orthogonal, it follows from (3) that each condition's response-pattern estimate will have a multivariate variability described by the same multivariate normal distribution (homoscedasticity). The Fisher dimension defined by vector  $w$ , then, depends only on the spatial response patterns and the covariance structure of the errors:

$$w = (p_2 - p_1) \Sigma^{-1}, \quad (3.4)$$

where  $p_1$  and  $p_2$  are the two spatial response patterns (for which the rows of  $\hat{B}$  provide estimates). The noise in fMRI time series is usually not white but temporally autocorrelated to some degree. Under such conditions the ordinary-least-squares fitting (which our t-test is based on) may not provide minimum-variance unbiased estimates and, more importantly, the variability of the estimates may be underestimated. The actual false positives rate can then substantially exceed the nominal  $\alpha$  level. We handle this well-known problem by an iterative prewhitening method, the Cochrane-Orcutt procedure (Cochrane and Orcutt, 1949; see also: Bullmore et al., 2001), which we apply to run-B data after projection onto the run-A Fisher dimension.

The approach of testing run-B effects on the run-A Fisher dimension, in our hands, appeared to be more robust to violations of multivariate normality than multivariate analysis of covariance (MANCOVA) performed on the data of both runs combined. Greater robustness to such violations appears plausible because, by the central limit theorem, the projection onto the Fisher dimension (a weighted summing of many random variables) is likely to yield scalar variables whose distribution is closer to normal than that of the original set of variables. Furthermore a significant effect on the Fisher dimension indicates a replicated effect of the same sign on the Fisher dimension.

Using only half of the data for the significance test suggests a loss of sensitivity, but this loss is at least partially compensated by the fact that the hypothesis tested is much more specific. The inclusion of information extracted from run A allows a univariate one-sided t-test to be performed—as opposed to a general test of multivariate difference (e.g. MANCOVA), whose sensitivity is compromised by the immensity of the space of possible effects.

Using run-B data to test for a multivariate effect also provides a solution to the multiple comparisons problem arising below, when the volume is continuously mapped for local pattern effects. Run A is then used to formulate a subject-specific hypothesis about the location and precise shape of the region as well as the multivariate effect dimension and direction.

## The pairwise stimulus information as a multivariate effect measure

As a measure of the information a region’s activity pattern contains about the experimental condition it is associated with, we use the mutual information. We will refer to the condition as the stimulus  $S$  and to the activity pattern as the response  $R$  here. However, the mutual information is symmetrical with respect to the two variables and there is no assumption of a directed influence. Any kind of experimental-condition variable can be used. For example, the conditions could be defined post hoc on the basis of behavioral measures.

We estimate the mutual information (Figure 3.8) between the stimulus and the multivariate response it elicits *on a single trial* (one complete hemodynamic response). Although the experiment as a whole (comprising many trials) provides much more information, we use the mutual information for a single-trial response as a measure because it is less dependent on accidental properties of the experiment such as the amount of data acquired and the efficiency of the

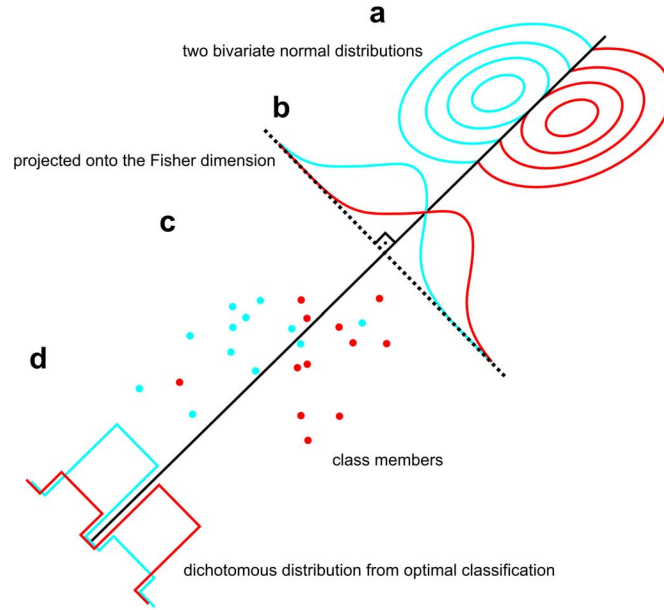


Figure 3.7: **Fisher linear discriminant and distributions obtained by projection and classification**

Some basic facts about two multivariate normal distributions of equal covariance are illustrated for the case of a two-dimensional response variable. The two classes could be two conditions of an fMRI experiment and the response space could be the multivariate response-pattern space of a region of interest. The facts stated below require the distributions to be multivariate normal with equal covariance, but are general with respect to the dimensionality of the response space. **(a)** Iso-probability-density contours for two bivariate normal distributions (red, blue) of equal covariance. The Fisher linear discriminant (dotted black line, henceforth: Fisher dimension) is orthogonal to the Bayes-optimal decision boundary (solid black line) for classifying a point in the space as belonging to one of the two distributions. The optimal decision boundary is the boundary along which the likelihood ratio of the two classes is 1. This is a hyperplane for multivariate normal distributions of equal covariance. **(b)** The probability-density functions of the two classes marginalized onto the Fisher dimension are univariate normal distributions (red, blue). The likelihood ratio of the two original multivariate normals is constant along the hyperplanes orthogonal to the Fisher dimension. Projecting a data point onto the Fisher dimension, thus, preserves all information about its class. Mutual information between the class variable (dichotomous here) and the multivariate response can therefore equivalently be computed either before or after projection of the joint probability density function onto the Fisher dimension. **(c)** Set of class members (e.g. trials of an fMRI experiment) generated on the basis of the multivariate normal distributions shown in a). **(d)** Classification of the members shown in (c) has been performed using the optimal decision boundary (black line). The resulting dichotomous distributions (red, blue) for the two classes can be used to estimate the mutual information between the class variable and the response (both dichotomous now). However, the estimate is inconsistent: For an infinite number of class members, the mutual information computed will be less than the true value despite the fact that the optimal decision rule is used. This is because the likelihood ratio changes continuously along the Fisher dimension and this information is lost by dichotomization.

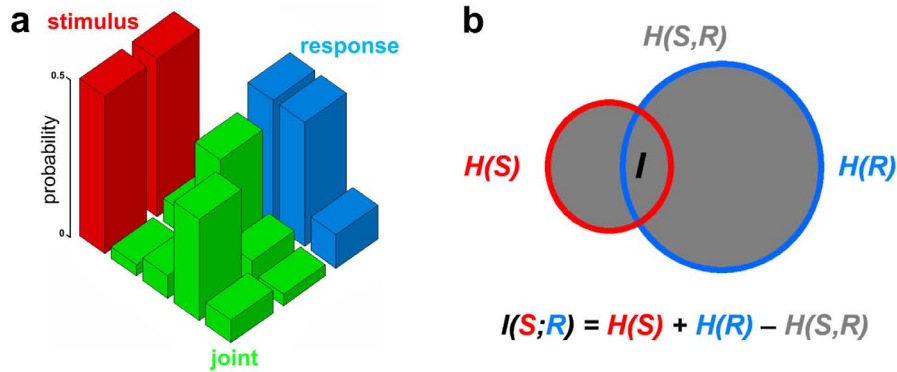


Figure 3.8: **Mutual information**

(a) The mutual information  $I(S; R)$  between two variables (e.g. stimulus  $S$  and response  $R$ ) can be computed from their joint distribution. Mutual information is a completely general measure of the relationship between two variables. If  $I(S; R) = 0$ , the two variables are independent and knowing one of them does not at all affect an ideal agent's expectation about the other. Independence means that the conditional probability  $p(s|r)$  is the same function for all values of  $r$  (or equivalently vice versa). This is the case exactly when  $p(s, r) = p(s) \times p(r)$ , i.e. when the joint distribution is the product of the marginals. Either of the variables or both can be multivariate. For the discrete case, all rows of the multidimensional distribution can simply be appended to yield a more extended one-dimensional distribution, to which Equation (3.5) can be applied.

(b) Mutual information can be thought of as entropy overlap between the two variables. Intuitively, mutual information is information shared between the variables. If the variables convey partly redundant information, then the whole information (joint entropy  $H(S, R)$ ) they convey together will be less than the sum of the information they convey singly (marginal entropies  $H(S)$  and  $H(R)$ ) by the amount of shared information (mutual information  $I(S; R)$ ).

design. One single-trial bit of information would imply perfect postdictability of the stimulus from a single trial's full hemodynamic response.

The mutual information estimate of a region is computed separately for each pair of stimuli. The stimulus variable is dichotomous (two stimuli) and uniform (each stimulus has the same probability of occurrence). The response has one dimension for each voxel in the region. We assume that the response distribution is multivariate normal and has the same covariance structure for each of the stimuli. The latter assumption (homoscedasticity) follows from the assumption of noise stationarity, which is usually made in univariate multiple-regression analysis of fMRI data.

Given multivariate normality and equal covariance structure, the multivariate response can be projected onto the Fisher dimension (see above) without loss of information about the stimulus. (This is because the likelihood ratio is constant on hyperplanes orthogonal to the Fisher dimension. See legend of Figure 3.7.) If the means and the covariance matrix are known, the mutual information can thus equivalently be computed from the one-dimensional response distributions of the two stimuli on the Fisher dimension (Figure 3.7). However, as the Fisher dimension maximizes class separation, estimation of the Fisher dimension and the distributions from the same data can give strongly positively biased information estimates. To reduce this bias, we use data from run A to determine the Fisher dimension and estimate the mutual information on the basis of the scatter and separation of the means of run-B data on the run-A Fisher dimension.

The covariance matrix describing the spatial structure of the noise is estimated from the residuals of the least-squares fit of the linear model with hemodynamic response predictors (Boynton, 1996), as described above. To estimate the covariance matrix describing the expected multivariate variability of the response-pattern estimates obtained *for single trials*, we assume a design matrix  $X$  with two nonoverlapping predictors, each of which describes a complete hemodynamic response (time window considered: 20 s). Equation (3.3) then provides the multivariate normal distributions (estimated from run-B data) which we project onto the run-A Fisher dimension. This results in two univariate normal probability density functions (one for each stimulus), from which we compute the mutual information  $I(S; R)$  between stimulus and response, which is defined as follows:

$$I(S; R) = H(S) + H(R) - H(S, R) = \sum_{s,r} p(s, r) \cdot \log \frac{p(s, r)}{p(s) \cdot p(r)} \quad (3.5)$$

where  $H(X) = - \sum_x p(x) \cdot \log p(x)$  is the entropy of a variable  $X$  with particular values  $x$ ,  $S$  is the stimulus variable (dichotomous here),  $R$  is the response variable (continuous here),  $s$  and  $r$  are particular values of  $S$  and  $R$ , respectively,  $p(x)$  is the discrete probability mass function of random variable  $X$ , and  $\log$  is the base-2 logarithm.

## Group analysis by meta-analytical integration

Local-pattern-effect analysis targets the spatial microstructure of the fMRI signal (1-2mm resolution in this study). The widespread method of using Talairach space to define a spatial point-to-point mapping between subjects for group-statistical analysis is not applicable in our case for two reasons. First, Talairach correspondence does not have millimeter precision. As Talairach correspondence does not even necessarily match functional regions of interest across subjects, it definitely cannot provide an accurate correspondence mapping for a region's intrinsic response pattern.

Second, a more fundamental concern, there may not exist a point-to-point mapping between brains that matches the response patterns representing the same entity (e.g. an object image) across subjects. These representations probably arise in a process of self-organization driven by an interaction between a developmental program and experience. Such processes are usually informed by a random element. Since the same representational space can be spanned by an infinite number of basis systems, a random choice of the basis system might not entail any functional disadvantages. The response patterns, thus, may be unique like the fingerprint.

We therefore first analyze local pattern effects idiographically. For each region of interest, the individual subjects' data are then statistically integrated at the level of the multivariate effect estimates, which abstract from the idiosyncratic layout of the local combinatorial representation. The statistical integration across subjects is performed for regions of interest defined functionally in each single subject.

The pairwise-stimulus-information effects reported for the group are averages across subjects. In order to test a multivariate effect for the group, we use the  $t$  value obtained from run-B data after projection onto the run-A Fisher dimension (see above) as a single-subject multivariate effect measure. These  $t$  values are combined meta-analytically, by performing an exact Monte-Carlo test to determine whether the mean of the absolute  $t$  values is significantly different from 0.

## Local-pattern-effect mapping

In order to find combinatorial codes, we scan the measured volume using a spherical multivariate searchlight. We consider the spherical neighborhood of each voxel of the volume and compute the local pattern effect for each pair of conditions. Consider two conditions 1 and 2 (e.g. corresponding to two face images) and the local spherical neighborhood of a single voxel. As a measure of the local pattern effect (the multivariate difference between the two local response patterns), we use the Mahalanobis distance, which is defined as

$$\Delta^2 = (p_2 - p_1) \Sigma_c^{-1} (p_2 - p_1)^T, \quad (3.6)$$

where  $p_1$  and  $p_2$  (as above) are two spatial activity patterns present within the spherical local neighborhood during conditions 1 and 2, respectively (for which the rows of  $\hat{B}$  provide estimates) and  $\Sigma_c$  is the covariance matrix describing the variability of the response-pattern estimate  $p_c$ , where  $c$  is the condition index. The use of the Mahalanobis distance is intuitively motivated in Figure 3.9. From Equation (3.3) we see that  $\Sigma_c$  is a scaled version of the error covariance matrix  $\Sigma$ :

$$\Sigma_c = \Sigma \cdot (X^T X)_{c,c}^{-1}. \quad (3.7)$$

The experimental design must be symmetrical with respect to the conditions corresponding to  $p_1$  and  $p_2$ , such that the response-pattern-estimate covariance is the same for  $p_1$  and  $p_2$ :  $\Sigma_c = \Sigma_1 = \Sigma_2$ . The Mahalanobis distance is computed for the local spherical neighborhood of each voxel in the measured volume, such that a continuous map of the local pattern effect is obtained. We automatically threshold this map to select a predefined number of voxels. The resulting contiguous regions are expanded with a spherical kernel of the same size as that used for mapping, such that each region includes all voxels that contributed to it in the mapping. The regions are then ranked by their size and tested multivariately (see above) in the order of their size on independent data (run B) obtained with the same experimental paradigm. The resulting  $p$  values are Bonferroni-corrected for multiple comparisons based on the number of tests performed.

## Discussion

### Examining the assumptions

#### **Interindividual consistency: high-resolution group analysis without the assumption of precise spatial correspondence**

Neuroimaging has demonstrated that the patterns of activity associated with a wide range of cognitive processes are consistent across individual brains. Like the organs of the body, different brain regions serve specific functions. At the macroscopic scale of centimeters, the spatial layout of the functional regions appears to be the same in every individual. This allows a spatial point-to-point mapping to be defined between individual brains. Talairach-space provides a rough approximation to such a mapping based on brain anatomy. Incorporating more detailed anatomical and functional information in defining the correspondence mapping will help increase its spatial precision. However, the problem goes beyond methodology. It concerns the fundamental neuroscientific question how closely different brains are related in terms of their functional structure and its spatial layout. To what level of precision does a functional correspondence *exist* between different human brains?

Different brains have different numbers of neurons, so a functional correspondence does not exist at the level of single neurons. The spatial precision of the functional correspondence will be different for different parts of the brain. In early visual areas, for example, one might expect a precision of correspondence in the millimeter range, because of retinotopic spatial organization. For the frontal lobe, on the other end of the continuum, it is not even known whether the functional component structure is completely consistent across individuals and one might expect a much lower precision of functional correspondence.

From a theoretical perspective, submillimeter-precise functional correspondence seems unlikely for all but early sensory and motor cortices. As the same feature space can be represented equally well by an infinite number of basis systems, it may not matter to the function of the system, which particular equivalent representation is actually used. Even if the same system of representational basis functions is used, the precise spatial layout of the units may not entirely be determined by functional constraints. Computational models of self-organization of representational maps suggest that there is a random element in the ontogenetic development of the spatial layout: Repetition of the developmental process yields a map of equivalent functionality but different spatial layout. Much like the fingerprint, the tertiary folding structure of the cortex, and, in fact, human personality, the details of cortical representation at the columnar level are probably unique to each individual.

Functional brain mapping currently faces a challenge: While human fMRI already operates at a resolution of one to two millimeters, the spatial correspondence mapping across brains provided by Talairach space has a much lower spatial precision. Information in the fine-grained structure of the data, thus, averages out in conventional group analysis, entailing a reduction of the spatial complexity of the measured signal by at least an order of magnitude.

Even with the ideal correspondence mapping, the spatial precision of correspondence is likely to be lower than the measurement resolution.

We therefore approach group analysis without the assumption of precise spatial correspondence. While allowing for the fine-grained spatial response patterns to be individually unique, we do assume that corresponding macroscopic regions encode the same information. For each subject separately, each region is functionally defined and its fine-grained response pattern is analyzed multivariately. Group analysis is performed at a coarser scale by combining all subjects' multivariate effects for a given macroscopic region. This two-scale approach allows us to statistically integrate individually unique fine-grained multivariate effects to obtain group-level characterizations of the distribution of a given type of information across brain regions.

### Contiguity of the distributed codes

Using a multivariate searchlight to find distributed representations implies the assumption that the representations are *focally* distributed, i.e. contiguous. Is this assumption of contiguity necessary? In order to find widely distributed, discontinuous representations, multivariate analysis could, in principle, be applied to each subset of the set of all voxels. This approach, however, is computationally intractable because the size of the set of all subsets of voxels is exponential in the number of voxels. The contiguity assumption is thus necessary, unless one is willing to settle for heuristic methods of selection of promising subsets of voxels (cf. Cox and Savoy, 2003).

Beyond considerations on the computational tractability of statistical analysis, the contiguity assumption is also *desirable* for conceptual reasons. A great wealth of neuroscientific evidence from cytoarchitectonic studies, lesion studies, electrophysiology and neuroimaging supports the idea of contiguous macroscopic regions with nameable functions. In the context of a representational framework, the contiguity assumption suggests that a given region contains a certain type of behaviorally relevant information (e.g. object-category information or individual-face information).

From a theoretical perspective, the spatial layout of the functional components of the brain is not arbitrary, because components located more closely can interact at a given bandwidth with lower costs in terms of exchange latency and neuronal connections. Neurons in the same region, therefore, are expected to participate in a common component function.

The above arguments suggest that the contiguity assumption is in line with the way the brain is organized. In other words, they suggest that assuming contiguity is descriptively correct. According to the following line of argument, contiguity should be assumed independent of empirical evidence—as a pre-empirical paradigm or norm. Cognitive neuroscience has attempted to localize in the brain cognitive component functions suggested by cognitive psychology. However, a complex cognitive function (e.g. object recognition) can be decomposed into component functions in many different ways. The task of cognitive neuroscience is not merely to localize component functions, but also to determine the correct decomposition of cognition into component functions. A decomposition should

be viewed as correct if the component functions turn out to be localized in contiguous regions. This paradigm implies that if a component function appears widely distributed, then the decomposition assumed by the researcher in designing the experiment is to be questioned. The early stages of primate vision are an example of a subsystem, for which neuroscience has already determined the global functional component structure.

### Shape and size of the regions containing distributed codes

Our spherical searchlight approach implies more than mere contiguity of the distributed codes to be located. By choosing a spherical searchlight of a particular radius, the analysis is focused on regions whose shape is well approximated by a superposition of spheres that have the same radius as the searchlight. If a target region is much smaller than the searchlight or, for instance, composed of elongated parts that are much thinner than the searchlight, the analysis will not be optimally sensitive to distributed effect, because at any location the searchlight will contain many voxels outside the region. These extraneous voxels will contribute to the noise, but not to the multivariate effect, and will thus lower sensitivity.

The target region most sensitively detected is spherical and of the same size as the searchlight. More generally, the spherical-searchlight approach focuses sensitivity on convex regions. Ideally, the searchlight should be given the shape and size of the target region, but these are usually not known *a priori*. Choosing the searchlight to be spherical implies minimal assumptions about the shape of the target region. Another advantage of sphericity is rotational invariance. If a rotationally noninvariant shape (e.g. a cylinder) were used as a searchlight, the multivariate statistic would have to be computed for all orientations at each location.

Choosing the radius of the sphere requires an informed guess about the size and type of shape of the target region. If the radius is too large or too small, sensitivity suffers—in the former case because too much noise is included in the searchlight when it is centered on the target region and in the latter case because spatial integration is restricted. The more separate spheres are needed to approximate the shape of the target region, the less of the distributed code is seen in any single location.

For finding *cortical* distributed codes, we consider an anatomically-informed method below. The approach we suggest takes the individual subject's cortical folding structure into account. The spherical searchlight is replaced by a roughly disc-shaped cortical patch.

### Multivariate normality

Our method assumes that the noise for a given contiguous region is multivariate normal. This assumption seems both reasonable and necessary. It seems reasonable because univariate normality is widely accepted for fMRI noise and conventionally assumed in most activation-based statistical mapping techniques. While multivariate normality implies univariate normality, however, the opposite implication does not hold, so strictly speaking our assumption of multivariate

normality needs to be tested. However, even if the assumption of multivariate normality is violated to some degree, it might still be a good choice for two reasons:

1) By choosing a multivariate normal noise model, we already estimate many more noise parameters than voxelwise univariate approaches from the same amount of data. For a region of  $d$  voxels the multivariate normal distribution has  $(d^2 + d)/2$  parameters. For large regions the number of voxels can exceed the number of time points per voxel, rendering the covariance matrix defining the multivariate normal distribution singular. In this case the covariance matrix cannot be inverted and computation of the Mahalanobis distance and the Fisher dimension is thus algebraically impossible. This suggests that it is necessary to consider relatively small sets of voxels multivariately, as is the case in our locally multivariate approach. It also suggests that the amount of data acquired in a typical fMRI experiment is not sufficient to estimate the parameters of a noise model substantially more complex than the multivariate normal model. Unless there is an alternative model of similar complexity that provides a better fit, the multivariate normal model may thus be the best practically viable compromise—even should it turn out to be violated to some degree.

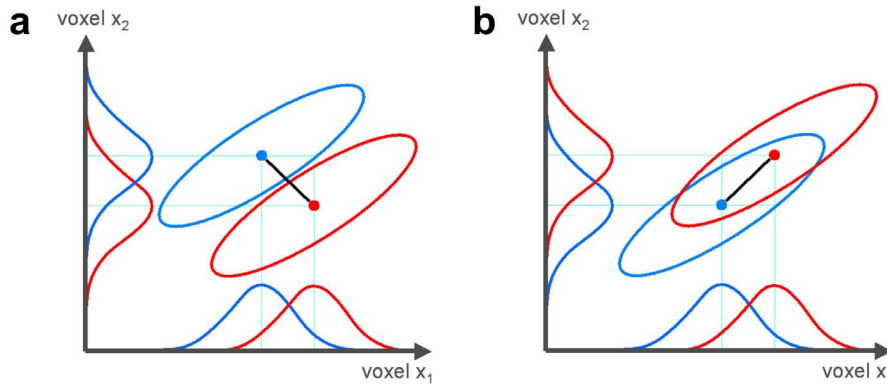
2) Even if multivariate normality is violated, our significance test is still likely to be valid, because the test is performed after projection of the data onto a single dimension of the multivariate space (the run-A Fisher dimension). This projection amounts to a weighted summing of many variables. Each of them can be assumed to have a normal distribution; this assumption is generally accepted in conventional analysis. Although the variables are not independent and may deviate from multivariate normality, their weighted sum may still approximate a normal distribution. This expectation should be tested before statistical inference. If it is violated, a nonparametric technique can be used to perform a valid test of mean difference for run-B data on the run-A Fisher dimension.

Such a significance test will be valid despite the fact that multivariate normality is assumed in defining the Fisher dimension, because a violation of multivariate normality will not affect the rate of false positives, i.e. the specificity. However, if multivariate normality is violated, the Fisher dimension may no longer be the optimal filter, so the rate of true positives, i.e. the sensitivity of the test, will be adversely affected.

## Alternative statistical approaches to information-based functional brain mapping

### Nonparametric approaches

Information-based functional brain mapping can also be approached nonparametrically, i.e. without assuming a particular shape for the probability density function describing multivariate noise. We have experimented with one particular nonparametric method of local-pattern effect mapping, which we will briefly outline to illustrate the kind of trade-offs involved in taking a nonparametric approach.



**Figure 3.9: Euclidean versus Mahalanobis distance**

This figure illustrates the advantages of the Mahalanobis distance as a multivariate effect statistic in comparison to the Euclidean distance or univariate effects. Two bivariate normal distributions and their marginal (univariate normal) distributions are shown. Each bivariate normal is represented by an iso-probability-density contour (red and blue ellipses). The marginal distributions are shown as probability density functions (red and blue Gaussians). The distance between the multivariate means (large dots in red and blue) is the Euclidean distance (black line). Panels (a) and (b) differ only in the orientation of the multivariate effect vector. The Euclidean distance is identical in both panels and so are the effects in the univariate projections (e.g. in terms of separation in noise standard deviations). The bivariate distributions, however, have a much larger overlap of their probability mass in panel (b). The Mahalanobis distance (Equation (3.6), not directly visualized here) takes the multivariate covariance structure of the noise into account and is much larger in panel (a). From a given starting point, the Mahalanobis distance is equal on iso-probability-density contours of a multivariate normal distribution centered on the starting point. It is the squared Euclidean distance measured after linearly transforming the space, such that the hyperellipsoid iso-probability-density contours become hyperspheres.

Our nonparametric method assumes that the experimental design is slow event-related, allowing an independent estimate of the response amplitude to be obtained for each single trial using the Boynton (1996) linear model of the hemodynamic response and least-squares fitting. For a given region of interest or the contents of the spherical searchlight at a particular position, this provides a noisy estimate of the spatial response pattern for each trial. All further computations are performed on the basis of these trial response estimates. Since each trial corresponds to a number of time points, this entails a substantial reduction of the data (e.g. by a factor of 5), making the computer-intensive simulation of null distributions tractable.

The trial response patterns can be thought of as points in the multivariate space spanned by the voxels of the region (Figure 3.1a, small dots). They reflect both the effects and multivariate noise. Condition-average response patterns can be obtained by averaging all trial response patterns belonging to the same condition (Figure 3.1a, large dots). Patterns can be compared by using any multivariate similarity or dissimilarity metric such as the Euclidean distance, the Mahalanobis distance (which implies assumptions of multivariate normality, however) or the spatial correlation.

Since the design is slow event-related, the trials are widely spaced in time and can be considered statistically independent despite the serial autocorrelation

that is present in fMRI time series. A multivariate test can therefore be performed by randomization. The condition labels are randomly shuffled and the multivariate effect statistic is recomputed many times allowing a characterization of its distribution under the null hypothesis of no difference between conditions. By determining the percent rank of the effect statistic for the actual condition labeling in the null distribution, a significance test can be performed for the multivariate effect.

To address the problem of multiple comparisons for a local-pattern-effect mapping using a multivariate searchlight, a randomization test can be performed at the level of whole maps by determining the distribution of map-maximum effect statistics under the null hypothesis (Nichols and Holmes, 2002). In this way the family-wise error rate, i.e. the probability of getting a single marked voxel in the map under null hypothesis, can be controlled.

Note that Bonferroni correction is not an option in the case of multivariate-searchlight mapping as it assumes that the multiple tests are independent. In our case, tests at neighboring positions are dependent not only because of the correlation between time courses of neighboring voxels (as in conventional univariate mapping) but also because of the large overlap of the searchlight in neighboring locations. Bonferroni correction, while validly controlling the family-wise error rate, would be out-of-control conservative. A randomization test at the level of whole maps obviates the need to split the data into two subsets (called run A and run B in our parametric method).

Given the trial response patterns for a region of interest, any pattern-classification algorithm (for a good introduction, see Duda et al., 2001) can be used to study predictability of the condition from the response pattern. The Fisher linear discriminant represents one such algorithm—the one that is Bayes-optimal for a two-class problem when the data are homoscedastic and have a multivariate normal density. Other algorithms including neural networks and support vector machines (Cox and Savoy, 2003) do not imply assumptions of multivariate normality or homoscedasticity and can in principle reveal response-pattern information encoded in more complex ways. A lower bound on the mutual information between the stimulus and the multivariate response can be obtained from any classifier’s confusion matrix.

The main advantages of the nonparametric method we outlined are that it does not require the assumption of multivariate normality and allows control of the family-wise error rate without splitting of the data. Despite these advantages, we do not favor this nonparametric method. The reasons are as follows.

*Restriction to slow event-related designs.* The restriction to slow event-related designs is necessary to ensure the independence of the trial response estimates despite serial autocorrelation of the time series. Independence of the trial response estimates confers exchangeability, allowing randomization testing.

Our parametric method, in contrast, handles slow as well as rapid event-related designs. Beyond greater freedom of experimental design, using rapid event-related designs has an important statistical advantage. A greater number of trials entails larger effect energy and, thus, better estimates of the condition-average responses. This will increase the sensitivity of the local-pattern-effect mapping.

*The small number of data points limits the characterization of multivariate phenomena.* The nonparametric method outlined above bases all multivariate analysis on the trial response estimates. This is problematic because an fMRI experiment can only accommodate relatively few trials. The problem is aggravated by the fact that only slow event-related designs can be used. A full hemodynamic response to a single event requires at least 12s. This means that no more than 5 trials can be presented per minute. For a 10-minute experiment, this gives us merely 50 trials overall. As stated above, the nonparametric approach potentially allows more complex models of the multivariate encoding to be used. But on the basis of only 25 data points per condition (assuming merely 2 conditions), it seems wise to work with very simple models. Even assuming homoscedasticity and multivariate normality, the covariance matrix will be singular if more than 50 voxels are considered jointly. Our parametric method makes these assumptions and uses the residual *time courses* instead of the reduced data set of trial responses to estimate covariance between voxels. Since the time course consists of about 5 times as many data points as the trial response estimates, the algebraic limit to the number of voxels that can be considered jointly is 5 times greater in our parametric method. It should be noted that using the time course residuals implies a different model of the noise sources. We will not discuss this issue here.

*Time complexity of whole-map randomization testing dictates simple effect statistics.* In the parametric method outlined above, significance thresholds for the local-pattern-effect maps are obtained by randomization testing. This method allows nonparametric inference accounting for multiple comparisons. However, it requires a large amount of computation because the local-pattern-effect mapping (involving computation of a multivariate statistic at every voxel location) needs to be repeated thousands of times (e.g. 10 000 times) with random condition labelings in order to obtain a null-hypothesis distribution of map maxima (Nichols & Holmes 2002).

In practice these computational demands limit the complexity of the multivariate effect statistics that can be used. The Euclidean distance, for instance, is an attractive measure of multivariate separation because its time complexity is linear in the number of dimensions, i.e. voxels in the region. The drawback of the Euclidean distance is that it is insensitive to the multivariate structure of the noise. For this reason our parametric method employs the Mahalanobis distance. The Mahalanobis distance could also be used in the nonparametric method. This would not imply the assumption of multivariate normality since the effect statistic's null distribution is obtained by randomization. However, using the Mahalanobis distance would almost certainly increase the sensitivity of the test. But it would take too long to compute because at each location the covariance matrix of the local population of voxels would need to be computed and inverted.

In summary, for information-based functional brain mapping based on fMRI data the most attractive feature of the nonparametric approach, i.e. that it has the potential to capture more complex multivariate encodings of information, appears largely illusory under the real-world constraints of limited data and limited computational resources. While the nonparametric approach allows valid inference (i.e. control of test specificity) without the assumption of normality, sensitivity may suffer as a result of the need to revert to very simple effect statis-

tics. We have examined one particular nonparametric method in contrast to our parametric method for the purpose of illustrating the kinds of tradeoff involved in designing information-based fMRI analysis techniques. Many other nonparametric implementations are possible. Within the nonparametric approach, however, we are not aware of solutions to the problems described.

### **Suprathreshold-cluster testing**

The classical repertoire of brain mapping techniques includes suprathreshold-cluster testing, a method exploiting the assumption of spatial contiguity of effects to increase sensitivity in detecting activated regions while taking account of multiple comparisons. This method can be adapted for information-based functional brain mapping and promises some advantages over the techniques defined so far.

In suprathreshold-cluster testing, first a primary threshold is defined arbitrarily. Voxels whose effect statistic exceeds the primary threshold are marked. Then a cluster-effect statistic is computed for each contiguous cluster of marked voxels. The cluster-effect statistic could be the size of the cluster or a more complex statistic, which takes the strength of the effects into account (e.g. the exceedance mass, see Poline et al., 1997; Bullmore et al., 1999). From the null distribution of cluster-effect statistics, a secondary threshold is determined to assess cluster significance. The primary threshold regulates the distribution of test sensitivity across different kinds of cluster effect. If the primary threshold is high, only strong effects are considered at all, but large spatial extent will not be required for a cluster to be detected. Lowering the primary threshold will shift sensitivity to weaker, more extended effects.

By adjusting the cluster-effect statistic, this method can be used for local-pattern-effect mapping. In the simplest implementation, the size of contiguous clusters of inhomogeneous univariate effects (opposite effect directions) could be used as a cluster statistic. We have experimented with a more complex implementation (Kriegeskorte and Goebel, 2002), using a spherical multivariate searchlight of very small radius to compute the initial map to be subjected to primary thresholding. As a cluster-effect statistic we used the Mahalanobis distance. By using a smaller multivariate searchlight, the method presupposes less assumptions about the shape of the region containing the distributed effect. The reduced degree of spatial integration resulting from the smaller searchlight is compensated for by using a low primary threshold. The primary thresholding defines the shape of the candidate regions very precisely. This makes the method sensitive to regions whose shape is not well captured by a superposition of spheres substantially larger than the small searchlight. Nevertheless the cluster statistic is computed for quite extended regions, bringing the benefit of ample spatial integration. The reduction to a few clusters per map afforded by the primary thresholding allows the application of more complex effect statistics even in the context of a nonparametric approach involving randomization testing at the level of whole maps.

## Future challenges

### Cross-lagged correlation

High-resolution fMRI voxels sample the venous part of the vascular system at different levels between capillary bed and large draining veins. A neuronal macropopulation's response pattern will first produce a hemodynamic effect pattern for us to measure in capillaries close to its origin. The flow of the blood will then wash out the effect pattern, displacing, distorting and blurring it in the process.

The local averaging resulting from convergence of vessels will reduce effects as well as noise in the high-spatial frequency band, but local-pattern information may still be present at intermediate levels of the venous draining tree, in venules and small veins. The slight spatial offset will not hamper global localization of functional regions at the coarse spatial scale, at which interindividual consistency is to be expected.

However, the responses in venules and veins are likely to peak with a slight delay as compared to the initial responses in the capillary bed. So far we have assumed a constant delay. Ideally, the delay should be estimated from the data and integrated into the model. This would slightly improve estimation of the response amplitudes.

Beyond amplitude estimation, delays introduce cross-lagged correlation to a region's multivariate response, i.e. correlations between voxels at a temporal lag. So far we have taken into account spatial correlations (i.e. correlations between voxels across time) and serial autocorrelation within voxels, but not cross-lagged correlation. Explicitly modeling the response delay between voxels may help in addressing this issue.

### More complex effect statistics

In this paper we have considered the prototypical case of comparing the spatial response patterns associated with two conditions. In many cases, a single type of information (e.g. face-identity information) may be better represented by a greater number of conditions, whose response pattern estimates are to be combined into a single statistic. It is not generally advisable to group the conditions into two subsets and to average the response patterns within each group.

Multivariate effect measures are scalar and always positive. They therefore do not behave linearly, as scalar differences or difference vectors do. Consider four response pattern vectors  $a_1$ ,  $a_2$ ,  $b_1$ , and  $b_2$ . The difference between two averages does not in general equal the average of two pairwise differences:

$$\frac{a_1 + a_2}{2} - \frac{b_1 + b_2}{2} \neq \frac{(a_1 - b_1) + (a_2 - b_2)}{2}, \quad (3.8)$$

where the minus sign “-” denotes a statistic representing the multivariate difference (e.g. the Euclidean or Mahalanobis distance). The pairing implied in taking the average of multivariate differences can be conceptually well motivated. For example a face-identity effect can separately be estimated for each of two views

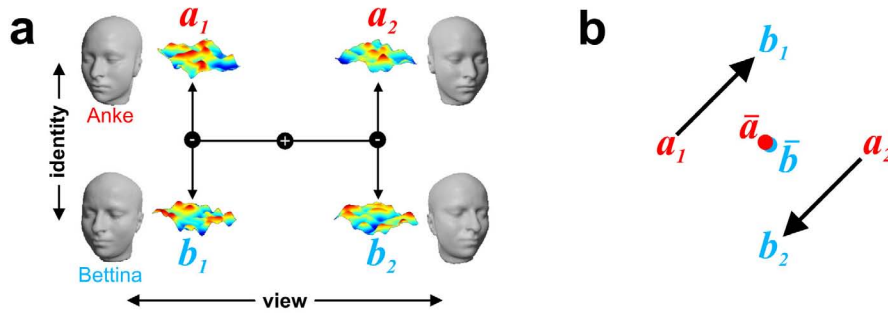


Figure 3.10: **Nonlinear-composite multivariate statistics**

The multivariate statistic used for local-pattern-effect mapping need not simply compare two spatial response patterns. For some research questions, a single type of information (e.g. face-identity information) may be better represented by a greater number of conditions. The response pattern estimates for multiple conditions can then be combined into a single composite multivariate statistic. In doing so it is not generally advisable to group the conditions into two subsets and to average the response patterns within each group. (a) The figure shows hypothetical stimuli of an fMRI experiment (four face images, two views of each of two persons) and the imaginary associated spatial response patterns of a region of interest (colored surface plots). The diagram indicates how a face-identity effect can separately be estimated for each of the two views and the two effects averaged. The minus (“-”) denotes some multivariate effect measure (e.g. the Euclidean or Mahalanobis distance). The plus (“+”) denotes the average. Panel b) illustrates why averaging the response patterns across views first can obscure the effect. (b) Multivariate effect measures are scalar and always positive. They therefore do not behave linearly like scalar differences or multivariate difference vectors. The points  $a_1$ ,  $a_2$ ,  $b_1$ , and  $b_2$  represent spatial-response-pattern vectors, corresponding to the stimulus faces of Anke (a) and Bettina (b) in each of the two views (indices 1 and 2). For the hypothetical multivariate arrangement shown, averaging response patterns across views first and considering the difference between the averages yields a multivariate effect of zero. This is because the two identity effect vectors add up to form the null vector, but the multivariate effects are a scalar positive value whose average reveals the actual average identity effect.

and the two effects averaged (Figure 3.10a). In contrast, averaging the response patterns across views first can obscure the effect (Figure 3.10b).

Such more complex, nonlinear-composite multivariate statistics may often better reflect specific research questions. In object recognition, for example, they can help in addressing questions as to the location of category, identity, and view information.

### Continuous group-statistical mapping

We have yet to define a fully automatic method for continuous group-statistical local-pattern-effect mapping. In our present method, we perform the local-pattern-effect mapping separately for each subject. The regions defined by this functional mapping are then matched across subjects based on their anatomical location. This matching currently requires some user interaction but could be automated (and thus made completely objective) on the basis of the location of the regions in a common spatial reference frame (e.g. Talairach space or a cortex-based common space). In the common space, the superposition of the corresponding individual regions could be used to compute a probabilistic map characterizing the location and shape of the functional region.

In the above approach, the shape of the region is first defined separately in each subject. Alternatively, the spherical-searchlight mapping defining the shape of the region could be performed concurrently for all subjects. At each location defined by the common space, the individual subjects' multivariate effects for the voxels falling within the searchlight would be combined to yield a group statistical map.

### Cortex-based local-pattern-effect mapping

Several advanced statistical techniques in fMRI analysis utilize an explicit representation of the spatial folding structure of the cortical sheet of each individual subject (see Chapter 1 or Kriegeskorte and Goebel, 2001). Local-pattern-effect mapping as well can benefit from such anatomical information when the objective is to find a focally distributed *cortical* code.

As an extension of the general approach of cortex-based statistics proposed by Goebel and Singer (1999), cortex-based local-pattern-effect mapping would benefit from a reduction of the set of voxels considered to voxels representing the gray matter. This reduction (by a factor of up to 5) alleviates the problem of multiple comparisons and reduces the computational load.

More importantly, the shape of the multivariate searchlight can be adapted to the shape of the cortex at each location. This will reduce the proportion of noise voxels within the searchlight and increase sensitivity.

We have discussed the choice of a sphere as the shape of the multivariate searchlight above. A spherical searchlight may be the best choice when nothing is known about the shape of the target region. When an explicit representation of the cortex is available, the sphere can be replaced by a roughly disc-shaped cortical surface patch. More precisely, the cortical sheet will be represented by a polygon mesh. The searchlight will be centered on each vertex of the polygon mesh in turn. At each vertex, all surrounding vertices within a predefined on-surface distance will be included in the patch. To select all gray-matter voxels corresponding to the patch, a range of distances along the normal vectors of the polygon mesh can be defined.

With a spherical searchlight, choosing a diameter greater than 4 mm (radius  $> 2$  mm) will entail inclusion of white-matter voxels at every location. With a cortical-patch searchlight, the diameter of the patch can exceed the thickness of the cortex, but the patch will still be restricted to the gray matter at every location. A larger patch promises greater benefits of spatial combination of signals. The anatomically-informed definition of the contiguous regions to be analyzed multivariately promises greater statistical power.

## Conclusion

We have motivated and defined a method of locally multivariate functional brain mapping that is suited for the localization of focally distributed representations. Our method represents a generalization of conventional brain mapping

techniques, which target macroscopic activated regions, in that it is additionally sensitive to information contained in the high-spatial-frequency band of regional response patterns. In other words, our method localizes regions whose multivariate response contains *information* about the experimental condition, whereas the conventional approach localizes *activation* and is most sensitive for extended regions that become activated as a whole. In this sense our method is information-based, whereas the conventional approach is activation-based.

We address an important challenge functional brain mapping is currently faced with: While human fMRI already operates at a resolution of one to two millimeters, the spatial reference frames relating locations in different individual brains (e.g. Talairach space) have a much lower spatial precision. We propose an approach involving two spatial scales. At the fine spatial scale of millimeters, response patterns are assumed to be unique to each individual and therefore analyzed separately for each subject. At the coarse spatial scale of centimeters, single-subject multivariate statistics are combined to increase statistical power and obtain a group-statistical summary of the individual results.



## Chapter 4

# Single-object-image response patterns in inferotemporal cortex

**Corresponding publication:** Kriegeskorte N, Formisano E, Sorger B, van de Moortele P-F, Ritter J, Adrian G, Kim D-S, Ugurbil K, and Goebel R. Single-object-image response patterns in human inferotemporal cortex. *Submitted*.

**Abstract:** Neuroimaging of visual object recognition has focused on the representation of object categories. Here we explore human inferotemporal response patterns to single object images (two faces and two houses) with functional magnetic resonance imaging. We estimate the amount of between- and within-category information present in spatial response patterns and search the measured volume for regional macropopulation codes by a novel method of local-pattern-effect mapping. Fusiform face and parahippocampal place regions both contain category information not only in the spatial-mean response usually studied but also in the spatial *shape* of their response patterns. The combinatorial representations in these posterior inferotemporal regions seem to emphasize category at the expense of visual appearance of individual exemplars. Unexpectedly, face-exemplar information appeared absent in the fusiform face region, but was found more anteriorly in inferotemporal cortex, casting doubt on the view that the fusiform face region is the main locus of individual-level face representation.

## Introduction

Visual objects are represented by a combinatorial code in primate inferotemporal cortex. This widely accepted view is the product of a paradigm shift in electrophysiology, which took place in the 1970s and was incited on the empirical front by the lack of evidence for single cells that exclusively respond to particular complex objects and on the theoretical front by the representational power afforded by combinatorial codes. Neuroimaging of human visual object recognition may be undergoing a similar paradigm shift at a coarser spatial scale of analysis.

Early neuroimaging studies on object vision established the existence of macroscopic inferotemporal regions selectively responsive to certain object categories. Two such well established regions are the fusiform face region and the parahippocampal place region (Kanwisher et al., 1997; McCarthy et al., 1997; Epstein et al., 1999; see also Puce et al., 1995; for electrophysiological background, see Perrett et al., 1982; Baylis et al., 1985; Rolls, 1992; Perrett et al., 1992; Allison et al., 1994).

More recent studies showed that there is a continuous spatial organization encompassing the category-selective regions (Levy et al., 2001; Hasson et al., 2002; Hasson et al., 2003a) and that category information is widely (Edelman et al., 1998; Ishai et al., 1999; Haxby et al., 2001; Cox and Savoy, 2003; Carlson et al., 2003), though not equally (Spiridon et al., 2002) distributed in human inferotemporal cortex (for electrophysiological background on distributed object representation, see e.g. Tanaka et al., 1991; Tsunoda et al., 2001). These studies suggest that combinatorial coding may be an essential principle not only at the microscopic scale of single cells but at the macroscopic scale accessible to functional magnetic resonance imaging (fMRI) as well.

A common feature of the cited neuroimaging studies is their focus on category effects. Whether employing block or event-related designs, these studies have analyzed differences between category-average responses. Here, we explore human inferotemporal spatial response patterns to single images of objects with event-related high-resolution fMRI. The single-image approach eliminates the somewhat arbitrary grouping of images into categories, brings us closer to what happens on a single trial of perception of a particular image, and allows us to investigate how exemplars within the same category are represented.

To obtain a good estimate of the spatial response pattern elicited by an object image from fMRI data, it is necessary to average across many trials of perception. We therefore estimate responses to only four particular object images: two face and two house images (Figure 4.1) in rapid and slow event-related fMRI designs. The argument for studying responses to a few particular images instead of average responses across many images can be made in precise analogy to the argument for studying a few subjects singly in depth instead of studying average measures obtained for groups.

In comparison to the category-average approach, our single-image approach moves the focus to a finer conceptual scale of object representation: into the categories. To match this move in terms of measurement, we target activity patterns at a finer spatial scale than previous studies, utilizing the high spatial

resolution (1-2 mm) of fMRI at high magnetic field strength (3T and 7T). In our analysis, we look into each region of interest, considering not merely its spatial-average response but the shape of its spatial response pattern. Furthermore, we search the measured volume for combinatorial codes by means of local-pattern-effect mapping. This method, which we introduce in this paper, allows us to find regions where the local response patterns differ multivariately (e.g. between two faces).



Figure 4.1: **Stimuli and anomaly-detection task.**

Panel (a) shows the four particular images whose inferotemporal response patterns are investigated in this study. Each image was processed to have a precisely uniform histogram. The images were presented sequentially while subjects fixated a central cross (not shown). Subjects performed an anomaly-detection task. About 12 percent of the images were variations (b) of the four originals (a), in which the global shape of the object as well as details had been slightly distorted. Anomalies were subtle and unpredictable as several anomalous versions were used for each original. The task ensured that subjects attended to each image presentation even after many repetitions.

## Results

### Local-pattern information in key regions of interest

#### Early visual areas

It is known that visually perceived images have a detailed retinotopic representation in primary visual cortex and other early visual areas. Each of the four images therefore should elicit a unique response pattern in retinotopic cortex. The extent to which the response patterns can be distinguished in the fMRI signal, tells us how much of the information is present in the relatively low spatial-frequency band accessible to fMRI.

The fMRI response pattern across early visual areas reflected the differences between all image pairs except the two faces (Figure 4.2). As a measure of multivariate effect between the response patterns to two images, we estimate

the pairwise stimulus information (the mutual information between the stimulus and the multivariate regional response, see Chapter 3) in single-trial bits. One single-trial bit of information would imply perfect postdictability of the stimulus from a single trial's full hemodynamic response.

For a bilateral region anatomically defined in each subject to encompass V1 and adjacent early visual areas (see Methods), the pairwise stimulus information present in the multivariate fMRI response was 0.012 single-trial bits or lower for all image pairs. Lowest information estimates were obtained for the two face exemplars. Consistently, the multivariate effect of face exemplar was not significant for the group (11 subjects), whereas the other pairwise multivariate effects were significant in early visual areas.

### **Parahippocampal place region**

In the parahippocampal place region, all between-category multivariate effects were highly significant and extremely strong. On average across subjects, the multivariate responses of the place region contained about 0.1 single-trial bits for between-category comparisons. This estimate includes information contained in the shape of the response patterns as well as information in the spatial-mean response.

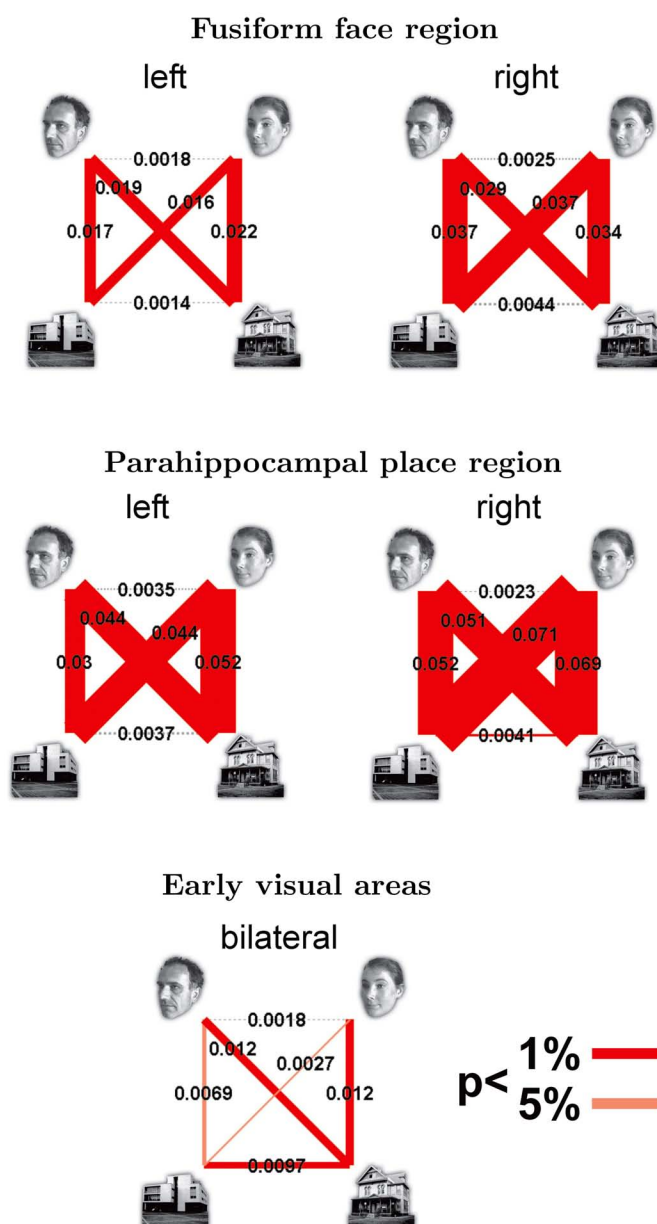
The spatial-mean response is expected to contain category information because the region is known to be selective for scenes and buildings (Epstein and Kanwisher, 1998) and was functionally defined here as the parahippocampal peak of the contrast map showing where the houses elicit greater activity than the faces (see Methods).

However, after subtracting the spatial-mean time course of the place region from each voxel's time course, half of the between-category information (about 0.05 single-trial bits on average across subjects and between-category image pairs) was still present in this region (Figure 4.2). This shows that it is not merely the spatial-mean response usually analyzed in fMRI studies, but also the shape of the response pattern within the region that contains between-category information.

Within-category effects were much less pronounced in the place region. The house-exemplar effect was small (similar in both hemispheres: about 0.004 single-trial bits) and significant for the group in the right hemisphere. The face-exemplar effect was not significant. In contrast to the between-category effects, the house-exemplar effect was only marginally affected by removal of the spatial mean, suggesting that it is the shape of the response pattern, not its spatial mean, that reflects which of the two houses is perceived.

### **Fusiform face region**

The pattern of information effects in the fusiform face region was similar to that of the parahippocampal place region with between-category effects very strong and within-category effects very weak (Figure 4.2). All between-category multivariate effects are highly significant for the group but somewhat smaller



**Figure 4.2: Local-spatial-response-pattern effects in key regions (group results, 3T experiment).**

For each region of interest, a local-pattern-effect icon shows the multivariate effects for each pair of images. A red (or light red) line between two images indicates a one-percent-significant (or five-percent-significant) multivariate difference between the response patterns elicited by the two images. A dotted gray line indicates a nonsignificant effect. The thickness of each line reflects the multivariate effect size in terms of the pairwise stimulus information (see Chapter 3), which is also given explicitly in single-trial bits (numbers on lines). To focus the analysis on genuine combinatorial effects, the spatial-mean effect has been removed from the data before the multivariate analysis (significance testing and information estimation) by subtracting the spatial-mean time course of the region from each single-voxel time course.

than in the place region. This might be related to the fact that the face region was smaller in spatial extent than the place region on average.

The qualitative pattern of effects does not depend on the spatial-mean effect. As for the place region, removal of the spatial mean reduced the between-category information to about half, showing that between-category information is present in the shape of the response pattern as well as in its spatial mean. Multivariate effects were generally stronger in the right hemisphere by about 0.01 single-trial bits.

None of the within-category effects were significant for the group in the fusiform face region. Note, however that the house-exemplar effect in the right face region is comparable to that found in both hemispheres' parahippocampal place region (about 0.004 single-trial bits). Importantly, no significant face-exemplar effects were found in either the right or the left face region in the group analysis. Consistently, the face-exemplar information estimates for the face region were extremely small (around 0.002 single-trial bits).

### **Zooming in: The fine-grained spatial response in the fusiform face region**

The fusiform face region responds selectively to faces (Kanwisher et al., 1997; Kanwisher et al., 1998; Kanwisher et al., 1999) and has been implicated in individual-level processing of faces (e.g. Gauthier et al., 2000; Haxby et al., 2000). It is the major candidate region for face-exemplar representation. However, imaging at 2-mm isotropic resolution in the 3T experiments did not reveal face-exemplar information in the fusiform face region. We therefore proceeded to image the response patterns of this region at even higher spatial resolution ( $1 \times 2 \times 1.6\text{mm}^3$ ) and with greater spatial specificity (Ugurbil et al., 2003) by fMRI at 7T. The measurement volume was restricted to the right posterior fusiform gyrus.

The pattern of multivariate effects we found in two subjects (two 20-min runs per subject, two data sets were discarded because of head motion) is identical to that found in the 3T experiments. All pairwise between-category effects are highly significant. The within-category effects are not significant for either faces or houses.

To exclude the possibility of missing face-exemplar-selective voxels because of our particular definition of the region of interest (choice of threshold, face-house contrast using the same two face images also used to estimate the face-exemplar effect), we redefined the fusiform face region, additionally including less strongly face-selective voxels by either of two methods: 1) lowering the threshold in defining the region and 2) expanding the original region isotropically by convolution with a spherical kernel of 2.1-mm radius. However, just like the original regions the resulting larger regions did not display a multivariate face-exemplar effect.

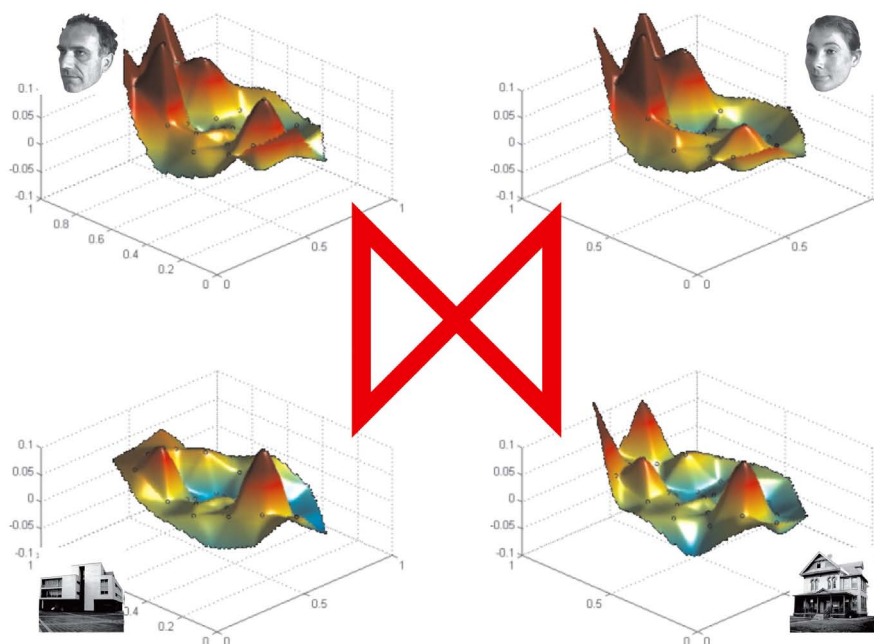


Figure 4.3: **Zooming in on the fusiform face region (single-subject results, 7T experiment)**

The fusiform face region's response pattern does not distinguish the two faces at a resolution of  $2 \times 1 \times 1.6 \text{ mm}^3$ . Each of the four panels shows the event-related spatial response pattern elicited by one of the four images. In the graphs, the vertical dimension represents response amplitude and the horizontal plane represents a flatmap of the region of interest (see Chapter 3, Visualization of spatial response patterns). The local-pattern-effect icon at the center of the figure is explained in the legend of Figure 4.2. However, in this figure line thickness does not reflect the pairwise stimulus information, which could not be estimated here because the two runs could not be coregistered with sufficient precision.

### Zooming out: continuous mapping of local pattern effects

Since face-exemplar information proved elusive in the fusiform face region's fMRI signal even at the higher resolution and specificity afforded by 7T fMRI, we decided to search for it throughout inferotemporal cortex.

Our method of local-pattern-effect mapping (see Chapter 3) allows us to find regions where the local response patterns differ multivariately (e.g. between face 1 and face 2). For each subject, a continuous map of the local pattern effect is obtained by scanning the volume with a spherical multivariate searchlight. At each voxel, the searchlight selects a spherical cluster of neighboring voxels, for which the Mahalanobis distance between the response patterns is computed. This analysis was performed on the 3T data, which afforded an extended field of view including all of inferotemporal cortex in both hemispheres. (In the 7T experiments, in contrast, the imaging was confined to a smaller volume centered on the expected location of the fusiform face region.)

Local-pattern-effect mapping revealed a region in anterior inferotemporal cortex that contains face-exemplar information. We first descriptively mapped the local

pattern effect in each subject using run-A data (searchlight of 3-mm radius, corresponding to 19 voxels at 2-mm isotropic resolution). The maps suggested the presence of face-exemplar information—not in the fusiform gyrus, but in anterior inferotemporal cortex.

To test the effect at the group level, we defined the most likely candidate anterior inferotemporal face-exemplar region in each subject based exclusively on run-A data. We excluded clusters containing obviously artefactual effects. Although some subjects only showed weak effects, we included all subjects in order to avoid biasing the group-level test by selection of subjects.

To determine the significance of the anterior face-exemplar effect, we then tested the candidate regions on independent data (run B used for single-subject and group-level multivariate tests, see Chapter 3). The face-exemplar effect in this region was significant for the group and in four single subjects.

Even though the region may be incorrectly defined or absent in some subjects (Table 1), the group-average face-exemplar effect is still significant and twice as strong (0.005 single-trial bits) as in the fusiform face region (0.0025 single-trial bits). In subject TS (Figure 4.5), the anterior inferotemporal face-exemplar effect is eight times as strong (0.021 single-trial bits) as in the fusiform face region.

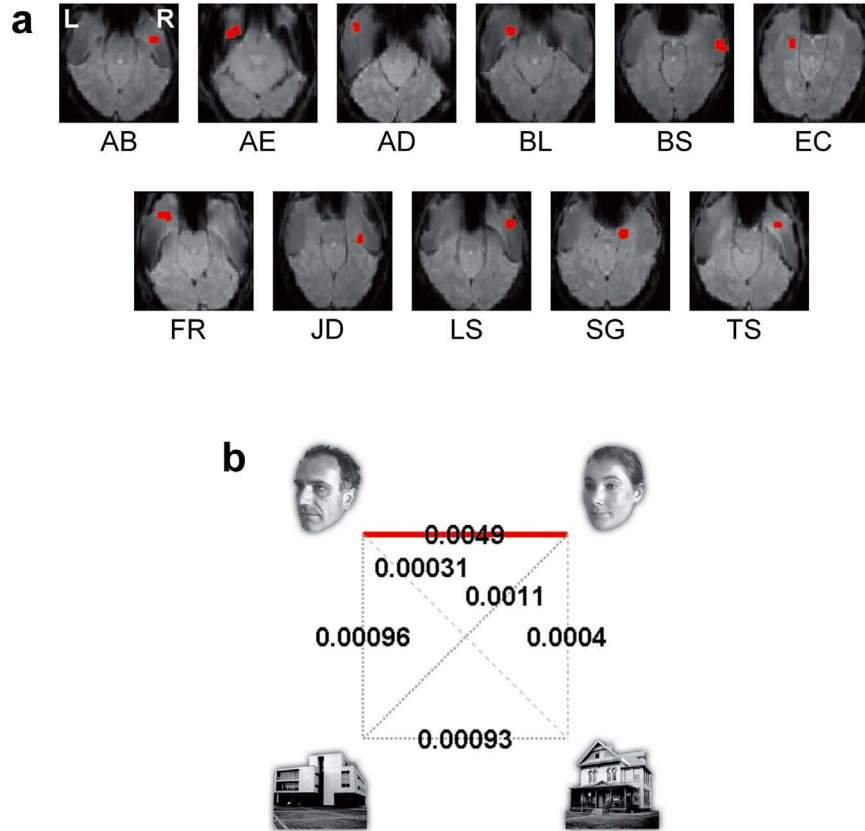
In the group analysis, the face-exemplar effect, but none of the other pairwise effects was significant (Figure 4.4b). The single subject analyses showed the same pattern in the four subjects, for whom the face-exemplar effect was significant (Table 1, Figure 4.5a).

Importantly, removal of the spatial mean (see Chapter 3) did not affect the pattern of results and the spatial-mean face-exemplar effect by itself was not significant. This rules out a spatially homogeneous (e.g. attentional) effect and supports an interpretation in terms of a combinatorial code for faces.

## Discussion

### Does the fusiform face region contain face-exemplar information?

The fusiform face region is frequently attributed the function of representing faces at the individual level (e.g. Gauthier et al., 2000; Hoffman and Haxby, 2000; Haxby et al., 2000). Consistent with this view, fMR-adaptation studies (see also Gauthier et al., 2000; I. Gobbini and J. Haxby, in preparation) suggest the presence of face-exemplar information in the fusiform face region (for a similar finding in the posterior fusiform gyrus, see Grill-Spector et al., 2001). In these studies, the summed response to different subsequent face stimuli was found to be greater in the fusiform face region than the summed response to a repetition of the same face stimulus. From this it is inferred that the region knows the difference between the different particular face stimuli—more precisely, that a different subpopulation of neurons was driven by each face resulting in less neuronal adaptation when different faces are presented in sequence.



**Figure 4.4: Anterior inferotemporal face-exemplar region (group results, 3T experiment)**

Local-pattern-effect mapping (see Chapter 3) revealed an anterior inferotemporal region whose response pattern appears to reflect which of the two faces is perceived. (a) A single slice with the face-exemplar region marked is shown for each subject. The anatomical background slices were obtained by averaging the functional volumes across time. The right hemisphere is shown on the right side. The region has been defined here in all 11 subjects based on half of the data (run A). (b) When tested multivariately on independent data (run B) not used for defining the region, the face-exemplar effect alone was significant at the group level and in four single subjects (see Table 1; for precise definition of the statistical test, see Chapter 3). For an explanation of the local-pattern-effect icon, see legend of Figure 4.2.

In contrast to the fMR-adaptation studies, we did not find face-exemplar information in the fusiform face region by analyzing fMRI spatial response patterns to different faces. We will consider two explanations for this apparent contradiction. The first explanation suggests the presence, the second the absence of face-exemplar information in the fusiform face region.

First, the face-exemplar information may reside in a higher spatial-frequency band not currently accessible to fMRI. Our approach of direct measurement of the spatial response patterns is necessarily limited by the spatial resolution of the measurement (in our case:  $1 \times 2 \times 1.6 \text{ mm}^3$ ). This restriction does not apply to the fMR-adaptation technique (for a thorough introduction, see Grill-Spector

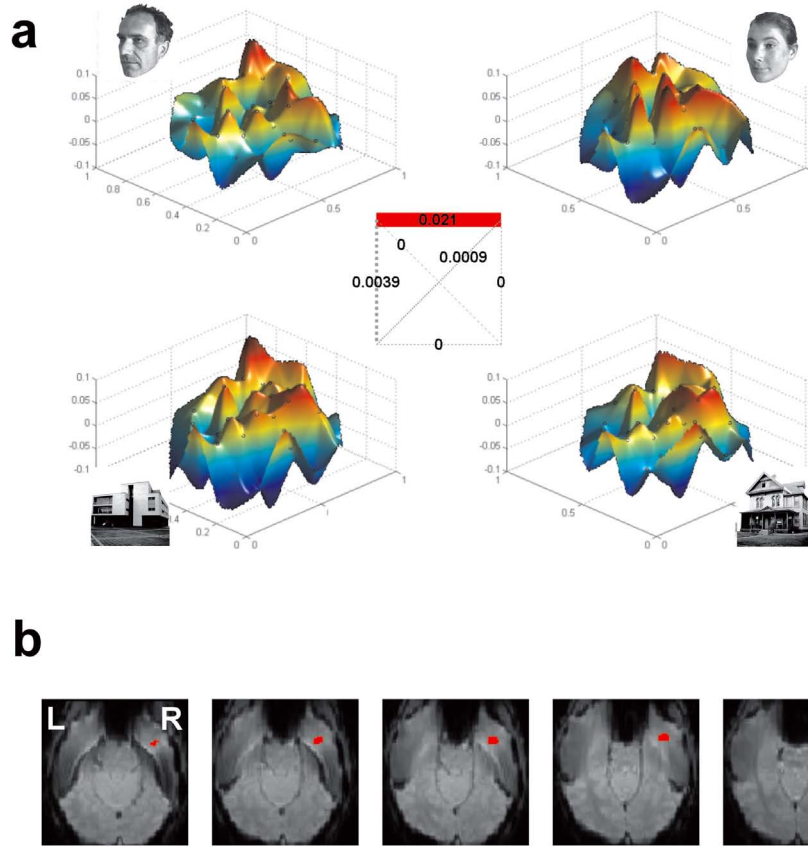


Figure 4.5: **Anterior inferotemporal face-exemplar region (subject TS, 3T experiment)**

(a) Event-related spatial response patterns elicited by the four images in the anterior temporal face-exemplar region of subject TS. The face-exemplar effect, but none of the other pairwise effects, is significant in this subject. The spatial-mean effects have been removed by subtracting the spatial-mean time course of the region from each single-voxel time course before the analysis. (Omitting this step yields the same pattern of multivariate effects with negligible changes to the effect sizes.) Equivalent results were obtained for subject EC and for the group. For an explanation of the central local-pattern-effect icon, see legend of Figure 2. (b) The anatomical location of the face-exemplar region in subject TS (Talairach coordinates: 33, -8, -33). The anatomical background slices were obtained by averaging the functional volumes across time. Slices progress from inferior to superior (left to right).

et al., 2001), which has the potential to detect a change of activity pattern even within a single voxel. If this explanation holds, then the fusiform face region may serve the function of individual-level face representation, which is often attributed to it.

As a second explanation for the apparently contradictory findings, the fMR-adaptation results may not reflect different patterns of response to different individual faces in the fusiform face region. In the fMR-adaptation technique,

the difference in the activity pattern is not directly measured as in our approach but inferred on the basis of the assumption that any reduction in fMRI signal will be due to neuronal adaptation effects in the region of interest.

Alternative explanations cannot be ruled out with certainty. Let us consider the possibility that the fusiform face region is a face detector (Kanwisher, 2000; Tong et al., 2000) in the context of a more complex face recognition system. A sequence of two different faces may still elicit more activity in the fusiform face region than a repetition of the same face. For example, in the case of a repetition, the visual system as a whole is primed to the particular appearance of the face to be detected. The fusiform face detector may thus converge more rapidly and expend less energy when the same face is repeated, resulting in a smaller summed hemodynamic response on face-repetition trials.

As another possibility (not excluding the previous one), a separate face-exemplar region (e.g. in anterior inferotemporal cortex, see below) may represent faces at the individual level. Such a region would be expected to be in strong interaction with the fusiform face region. The face-exemplar region would show a genuine adaptation effect, responding less strongly when the same face is repeated. This lesser response to a repetition may then carry over to the fusiform face region, even if the fusiform face region's response pattern is not specific to individual faces.

Lesion studies do not resolve whether the fusiform face region simply detects faces or represents individual face exemplars. Prosopagnosia, a recognition deficit restricted to the domain of faces, is often associated with damage to ventral occipito-temporal regions including the fusiform gyrus (Damasio et al., 1982). The fusiform face region may thus be necessary for face identification. However, its role in the face identification process may be to detect faces and to initiate face-exemplar representation elsewhere. This would explain why prosopagnosia can also be caused by lesions to anterior temporal cortex (references and discussion below) in the absence of damage to the fusiform face region. In fact, there are prosopagnosics whose fusiform face region has been shown by fMRI to be functionally intact, responding selectively to faces (Marotta et al., 2001; Hasson et al., 2003b; Rossion et al., 2003). This shows that the fusiform face region is not sufficient for face identification.

In summary, the fusiform face region probably contributes to the face identification process, but its primary role in that process may be face detection rather than face-exemplar representation. Our results do not rule out the possibility that the fusiform face region contains face-exemplar information (for example in the high spatial frequency band or in the temporal structure of its response), but the absence of a significant multivariate face-exemplar effect at a reasonably high resolution is at least unexpected under this view. That a significant face-exemplar effect is absent in the fusiform face region but present in anterior inferotemporal cortex does cast doubt on the view that the fusiform face region is the main region serving face-exemplar representation.

## **A face-exemplar region in anterior inferotemporal cortex**

By local-pattern-effect mapping, we found a region in anterior inferotemporal cortex whose response pattern reflects the difference between the two faces.

The face-exemplar effect consisted not in a greater spatial-mean response of the region to either of the faces, but in a distinct pattern of response associated with each face. The effect, thus, cannot be accounted for by a globally acting mechanism such as attention and is consistent with a combinatorial code for faces (response patterns from subject TS are shown in Figure 4.5). The anterior inferotemporal face-exemplar region appeared to contain little or no between-category or house-exemplar information either in its spatial-mean response or in the shape of its spatial response pattern.

A face-exemplar local pattern effect is of particular interest, not only because face identification is a socially crucial function of the human visual system, but because faces are very similar in a low-level sense—more so than individual exemplars of other categories. In our experiment, for instance, the two face images were much more similar in terms of the correlation of the image intensity across space than the two house images or, in fact, any other pair of images. As a consequence, the response patterns associated with the two houses could be distinguished in early visual areas and posterior inferotemporal cortex, whereas those elicited by the two faces could not.

In contrast to the house-exemplar effect, the face-exemplar effect is restricted to a single region within our measured volume (which included the entire ventral stream bilaterally). The presence of face-exemplar information in the low spatial-frequency band accessible to fMRI implies that a very subtle visual difference is blown up to detectability as it comes to be represented in a macroscopic combinatorial code.

Our findings, thus, suggest that the anterior inferotemporal region represents visual objects by a basis system of visual-shape dimensions that magnifies the subtle differences between faces. One caveat needs to be considered: we found this region based on measurement of response patterns to only two particular face images, one of a young woman, the other of a slightly more mature man. Two obvious confounding factors, thus, are sex and age. Additional research employing single-image designs is needed to eliminate these confounds. However, it seems implausible that an entire region should be devoted exclusively to the representation of the dichotomous variable of sex or the one-dimensional variable of age. This suggests that face-exemplar representation is at least one of the functions of the region we found.

Another important question to be considered is whether the region contains a high-level perceptual representation as suggested above or a higher cognitive, more task-dependent representation. In the latter case, the region might distinguish objects on the basis of the task at hand, abstracting from currently irrelevant differences.

Our anomaly-detection task (see Methods), however, did not bias the perceptual process in favor of face discrimination as compared to house or between-category discrimination. In fact, the task did not require any comparisons between the four images to be made and instead encouraged independent perception of each presented image. Nevertheless the region showed a face-exemplar, but no house-exemplar or between-category effect.

One might argue that face identification but not house identification is automatic. To eliminate this possibility, a future study might counter the face-identification automatism by a house-identification task. Note, however, that

categorization is probably automatic as well and did not lead to a between-category effect in the anterior inferotemporal region. This supports the view that it is a perceptual region specific to the visual domain of faces.

Consistent with our finding, neuroimaging and lesion studies suggest the existence of face and identity representations in anterior temporal cortex. Sergent et al. (1992) found bilateral anterior temporal activation associated with performance of a face-identity task. Kuskowski and Pardo (1999) found right temporal polar cortex active during face perception and recognition. Lesion studies, as well, suggest that the right temporal polar region is involved in face identification (e.g. Tranel et al., 1997).

Right anterior temporal atrophy is frequently associated with progressive prosopagnosia (e.g. Evans et al., 1995). Gainotti et al. (2003) describe a patient with a right anterior temporal focal atrophy associated with impaired identification of familiar people from their faces or voices. The authors suggest that the impairment may be caused by damage to face-recognition units (Bruce and Young, 1986; Burton et al., 1990) in right anterior temporal cortex. Left anterior temporal cortex may serve the representation of unique entities including people and landmarks (Damasio et al., 1996; Grabowski et al., 2001).

Anterior temporal activity is sometimes found to be evoked by familiar but not by unfamiliar faces and is therefore often interpreted to reflect access to memory about people (for further evidence, see Sugiura et al., 2001) rather than a visual representation optimized for the domain of faces. Haxby et al. (2000) suggest that anterior temporal cortex may contain representations of person identity, name, and biographical information.

Person-identity representation is likely to be rooted in face-exemplar representation. In language and writing, purely conventional symbols can evolve from iconic or onomatopoeic representations that are similar to the entity they represent. Analogously a person-identity representation may develop as an abstraction of an individual-face representation.

More generally, face-exemplar representation, person-identity representation, and memory encoding and recall concerning people are closely related functions, all of which may be subserved by anterior temporal cortex—possibly each in a separate region.

The region we found contains information about the perceived face exemplar in its intrinsic pattern of response. However, it does not respond more strongly to faces than to other objects in terms of its spatial-average activity. Since such behavior is not revealed by conventional contrast mapping, our region may be separate from the anterior temporal regions described in previous neuroimaging studies. To localize it may require the conjunction of single-image event-related design, high-spatial-resolution measurement, and local-pattern-effect mapping that distinguishes this study.

In summary, the anterior inferotemporal region is likely to be part of the face recognition network and may serve individual-level face or person representation. Its precise function and possible invariances need to be addressed in future studies.

## The posterior inferotemporal macropopulation code may reflect category more than visual appearance

Between-category information was barely detectable in early visual areas but very strong in the parahippocampal place and the fusiform face region. This information resides in the shape of the response pattern as well as in the spatial-mean response usually studied. In the parahippocampal place region, for example, the amount of category information present in the shape of the fMRI response was about an order of magnitude greater than in early visual areas. This suggests that the intervening processing projects category information into the low spatial-frequency band accessible to fMRI.

The house-exemplar information present in the fMRI response, in contrast, is reduced to less than half on its way from early visual areas (about 0.01 single-trial bits) to the inferotemporal regions (between 0.001 and 0.005 single-trial bits). Note that the two house images are not similar visually. Their image intensity is essentially uncorrelated across space, just as for the between-category pairs. Visual dissimilarity, thus, appears to be abstracted from by the fusiform face and parahippocampal place regions when the images belong to the same category. Note, however, that exemplar information was still present in the posterior inferotemporal regions, and that its strength appeared dependent on visual dissimilarity (house-exemplar information present, but no face-exemplar information detectable).

In summary, the present results suggest that category (as conventionally defined in this field) more than visual appearance determines the posterior inferotemporal response pattern—at least in the spatiotemporal frequency band accessible to fMRI. Note that our findings address the organization of the multivariate response space in each region and not the large-scale spatial layout of inferotemporal cortex. They, thus, support the widespread view that category constitutes an important organizing principle intrinsic to visual representation in inferotemporal cortex, but do not address the question whether the category-selective regions are modules (i.e. functional units with natural, clearly defined boundaries) or selectivity peaks of a continuous representational map.

Although our study takes a step in the direction of reducing a-priori assumptions about category representation, it is still biased because of our choice of stimuli and regions of interest. Future studies may attempt to further reduce this bias to place the notion of category as organizing principle on a more solid foundation. The single-image approach may help in defining empirically the particular category structure implicit to the inferotemporal representation of visual objects.

## Methods

### Stimuli

The basic set of stimuli consisted of four photographs (Figure 4.1a), depicting a woman's face, a man's face, a traditional house in Minneapolis and a modern building in Maastricht. The images were in 8-bit grayscale and had a

resolution of 512 by 512 pixels. Each image was processed to have a precisely uniform histogram. The images, thus, had identical light and spatial-signal energy. Anomalous versions were constructed using image processing software (see Task, below).

## Task

Before the experiment, subjects were familiarized with the four original images (Figure 4.1a). They were instructed to continually fixate a central cross, which was always visible, and to perform an anomaly-detection task during the experiment. On a small number of trials of each experimental run (12% in the rapid event-related 3T experiment, 11% in the slow-event-related 7T experiment), subtle variations of the four original images were presented. In each anomalous version, the global shape of the object as well as several details were slightly distorted (Figure 4.1b). The particular changes were unpredictable to the subjects as several anomalous versions were used for each original. Subjects were asked to press the button placed underneath their right index finger on a regular trial and the button underneath their left index finger when they detected an anomaly. The task required them to attend to each image presentation even after many repetitions and allowed us to monitor attentive viewing.

## Experimental design

In the 3T experiments, we used a rapid event-related design with a basic trial duration of 3s (minimal stimulus onset asynchrony), corresponding to two functional volumes of  $TR = 1500\text{ms}$ . The stimulus sequence was optimized for estimation of the contrasts between the responses to the four original images by a method based on a genetic algorithm (Wager et al., 2003). Each image was presented for 400ms. There were 63 presentations of each of the four original images, 33 presentations of anomalous versions of the images (see Task) and nine null trials. The total number of 3-s time slots was, thus,  $4 \times 63 + 33 + 9 = 294$ , and the duration of the run including two empty time slots at the end was  $(294 + 2) \times 3\text{s} = 14.8\text{min}$ .

In the 7T experiments, we used a slow event-related design with a random stimulus sequence. The trial duration was 16.779s (stimulus onset asynchrony), corresponding to seven functional volumes at a time to repeat (TR) of 2397ms. Each image was presented for 3000ms. There were 16 presentations of each of the four original images, eight presentations of anomalous images (see Task) and an initial empty time slot. The total number of time slots was, thus, 73 and the duration of the run was 20.42min.

## Subjects

### 3T experiments

Eleven subjects between 18 and 30 years of age participated in the 3T experiments (average age: 24.5 years). All had normal or corrected-to-normal vision. Five of them were female, six male. Ten of them were right-handed, one was

left-handed. After receiving information about magnetic resonance imaging they gave their informed consent by signing a form. The experimental techniques used in this study and the consent form were approved by the ethical committee CWOM of the Academisch Ziekenhuis (university hospital) associated with the Katholieke Universiteit Nijmegen (The Netherlands).

### 7T experiments

Two subjects who provided written consent by signing a form approved by the institutional review board at the University of Minnesota participated in the 7T experiments. Their ages were 21 and 22. Both had normal or corrected-to-normal vision. One was female, the other male.

## Magnetic resonance imaging

### 3T measurements

We measured 15 transversal functional slices with a Siemens Magnetom Trio scanner (3 Tesla) using a single-shot gradient-echo echo-planar-imaging (EPI) sequence. The pulse-sequence parameters were as follows: in-plane resolution:  $2 \times 2mm^2$ , slice thickness: 2mm, gap: 0mm, slice acquisition order: interleaved, field of view (FoV):  $256 \times 256mm^2$ , acquisition matrix:  $128 \times 128$ , TR: 1500ms, time to echo (TE): 32ms, flip angle (FA): 75 deg. A functional run lasted 14.8min. Each subject underwent two such runs in the same imaging session. Each subject also underwent a high-resolution T1-weighted anatomical MPRAGE scan in the same magnet, lasting 9.8min (192 slices, slice thickness: 1mm, TR: 2300ms, TE: 3.93, FA: 8 deg, FoV:  $256 \times 256mm^2$ , matrix:  $256 \times 256$ ). Functional and anatomical imaging was performed at the Donders Centre for Cognitive Neuroimaging (Nijmegen, The Netherlands).

### 7T measurements

Functional imaging was performed using a 90-cm bore 7-Tesla magnet (MagneX Scientific, Abingdon, UK) operated from a Varian console (Palo Alto, CA, USA). A quadrature surface coil (with two overlapping loops covering the temporal and occipital lobes of the right hemisphere) was used for radiofrequency transmission and reception. High-resolution functional images of the right temporal lobe were collected with a single-shot gradient-recalled blipped EPI sequence (TE: 27ms, FA: 90 deg). Localized adjustments of the first-order shims, and where needed of the second-order shims, were performed in the temporal cortex with FAST(EST)MAP (Gruetter and Tkac, 2000). In order to acquire single-shot high-resolution EPI images, an outer-volume suppression technique with adiabatic pulses (Luo et al., 2001; Pfeuffer et al., 2002) was used, yielding a reduced field-of-view and a nominal resolution of  $1.0 \times 2.0 \times 1.6mm^3$ . The TR was 2397ms for each volume consisting of 14 slices (interleaved order, no gap between slices). These measurements were performed at the Center for Magnetic Resonance Research (Minneapolis, USA).

## Local-pattern-effect analysis, mapping, and visualization

Local-pattern-effect analysis, mapping, and visualization has been performed using the method of parametric information-based functional brain mapping as defined in Chapter 3. The radius chosen for the searchlight was 3 mm. At our isotropic functional resolution of 2 mm, this radius results in a spherical searchlight comprising 19 voxels. Unless otherwise stated the spatial pattern mean has been removed prior to the analysis, so the multivariate effects reflect the information contained in the shape of the pattern, not the average regional activity level.

**Table 1: Anterior inferotemporal face-exemplar region**

| Subject | Talairach location |     |     | number of voxels | effect icon |
|---------|--------------------|-----|-----|------------------|-------------|
|         | x                  | y   | z   |                  |             |
| AB      | -32                | -14 | -30 | 86               |             |
| AE      | -28                | -6  | -29 | 103              |             |
| AD      | -44                | 1   | -29 | 71               |             |
| BL      | -25                | -3  | -26 | 74               |             |
| BS      | 60                 | -21 | -20 | 103              |             |
| EC      | -25                | -15 | -20 | 72               |             |
| FR      | -33                | 7   | -25 | 91               |             |
| JD      | 32                 | -19 | -22 | 56               |             |
| LS      | 47                 | 0   | -26 | 77               |             |
| SG      | 24                 | -9  | -16 | 68               |             |
| TS      | 33                 | -8  | -33 | 68               |             |
| group   | +35                | -8  | -24 | 79               |             |

## Conventional statistical analysis

### Preprocessing

The fMRI data sets were subjected to a series of preprocessing operations: 1) Slice-scan-time correction was performed by resampling the time courses with linear interpolation such that all voxels in a given volume represent the signal at the same point in time. 2) Small head movements were automatically detected and corrected by utilizing the anatomical contrast present in functional MR images. The Levenberg-Marquardt algorithm was used to determine translation and rotation parameters (6 parameters) that minimize the sum of squares of the voxelwise intensity differences between each volume and the first volume of the run. Each volume was then resampled in 3-D space according to the optimal parameters using trilinear interpolation. 3) Temporal high-pass filtering was performed to remove temporal drifts of a frequency below 8 cycles

per run ( $8/14.8\text{min} = 0.009\text{Hz}$ ) in the 3T experiments and 5 cycles per run ( $5/20.42\text{min} = 0.004\text{Hz}$ ) in the 7T experiments. Preprocessing was performed using the BrainVoyager 2000 software package (version 4.8).

### Definition of regions of interest

*Early visual areas.* In each subject, the region referred to as “early visual areas” was defined anatomically. The region consisted of two cuboids, one for each hemisphere, aligned with the coordinate system of the slices as measured. The two cuboids shared a boundary along the medial plane and were independently located to be centered on the calcarine sulcus and to start at the occipital pole. The cuboids had equal size in each dimension in both hemispheres in all subjects, extending 8mm (4 voxels) laterally from the medial plane, 10mm (5 voxels) vertically around the calcarine sulcus, and 16mm (8 voxels) anteriorly from the occipital pole.

*Parahippocampal place and fusiform face regions.* The parahippocampal place region was defined by univariate multiple-linear-regression contrast mapping (see below) as the region within the parahippocampal gyrus where the mean response to the two houses was maximally greater than the mean response to the two faces. The fusiform face region was defined as the region within the fusiform gyrus where the reverse contrast was maximal. Thresholds were initially set to correspond to a  $p$  value of 5% with Bonferroni correction for multiple comparisons. When the resulting region of interest appeared implausibly large, the threshold was adjusted to yield a region of plausible size ( $p < 5\%$ , corrected).

*Multiple linear regression at every voxel.* Single-subject analyses were performed by multiple linear regression of the response time course at each voxel. For each of the four original images, there was one predictor for the regular version and one predictor for the anomalous versions presented (see Task). The predictor time courses were computed using a linear model of the hemodynamic response (Boynton et al., 1996) and assuming an immediate rectangular neural response during each condition of visual stimulation. Multiple linear regression was performed using the BrainVoyager 2000 software package (version 4.8).

## Acknowledgments

This research was funded by the Universiteit Maastricht (Maastricht, The Netherlands), the Donders Centre for Cognitive Neuroimaging (Nijmegen, The Netherlands), and the Center for Magnetic Resonance Research (University of Minnesota Medical School, Minneapolis, USA). Work performed at the University of Minnesota was supported by NIH grants RR08079, EB00331, NS44825, the Keck Foundation and the MIND Institute.

# References

- Albright TD. Direction and orientation selectivity of neurons in visual area mt of the macaque. *J Neurophysiol*, 52:1106–30, 1984.
- Allison T, Ginter H, McCarthy G, Nobre AC, Puce A, Luby M, and Spencer DD. Face recognition in human extrastriate cortex. *J Neurophysiol*, 71(2):821–5, 1994.
- Amedi A, Malach R, Hendler T, Peled S, and Zohary E. Visuo-haptic object-related activation in the ventral visual pathway. *Nat Neurosci*, 4:324–30, 2001.
- Andersen RA and Bradley DC. Perception of three-dimensional structure from motion. *Trends Cogn Sci*, 2:222–228, 1998.
- Baylis GC, Rolls ET, and Leonard CM. Selectivity between faces in the responses of a population of neurons in the cortex in the superior temporal sulcus of the monkey. *Brain Res*, 342:91–102, 1985.
- Bishop CM. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.
- Boynton GM, Engel SA, Glover GH, and Heeger DJ. Linear systems analysis of functional magnetic resonance imaging in human V1. *J Neurosci*, 16:4207–21, 1996.
- Braddick OJ, O’Brien JM, Wattam-Bell J, Atkinson J, and Turner R. Form and motion coherence activate independent, but not dorsal/ventral segregated, networks in the human brain. *Curr Biol*, 10:731–4, 2000.
- Bradley DC, Chang GC, and Andersen RA. Encoding of three-dimensional structure-from-motion by primate area mt neurons. *Nature*, 392:714–7, 1998.
- Bruce V and Young A. Understanding face recognition. *Br J Psychol*, 77 (Pt 3):305–27, 1986.
- Bullmore E, Long C, Suckling J, Fadili J, Calvert G, Zelaya TA, Carpenter F, and Brammer M. Colored noise and computational inference in neurophysiological (fMRI) time series analysis: resampling methods in time and wavelet domains. *Hum Brain Mapp*, 12:61–78, 2001.

- Bullmore ET, Suckling J, Overmeyer S, Rabe-Hesketh S, Taylor E, and Brammer MJ. Global, voxel, and cluster tests, by theory and permutation, for a difference between two groups of structural mr images of the brain. *IEEE Trans Med Imaging*, 18(1):32–42, 1999.
- Burton AM, Bruce V, and Johnston RA. Understanding face recognition with an interactive activation model. *Br J Psychol*, 81 (Pt 3):361–80, 1990.
- Carlson TA, Schrater P, and He S. Patterns of activity in the categorical representations of objects. *J Cogn Neurosci*, 15:704–717, 2003.
- Carman GJ, Drury HA, and Van DC Essen. Computational methods for reconstructing and unfolding the cerebral cortex. *Cerebral Cortex*, 5:506–517, 1995.
- Cochrane D and Orcutt GH. Application of least squares regression to relationships containing autocorrelated error terms. *J Am Stat Assoc*, 44:32–61, 1949.
- Corbetta M, Akbudak E, Conturo TE, Snyder AZ, Ollinger JM, Drury HA, Linenweber MR, Petersen SE, Raichle ME, Van DC Essen, and Shulman GL. A common network of functional areas for attention and eye movements. *Neuron*, 21:761–73, 1998.
- Corbetta M, Shulman GL, Miezin FM, and Petersen SE. Superior parietal cortex activation during spatial attention shifts and visual feature conjunction. *Science*, 270:802–5, 1995.
- Cox DD and Savoy RL. Functional magnetic resonance imaging (fMRI) brain reading: detecting and classifying distributed patterns of fMRI activity in human visual cortex. *Neuroimage*, 19:261–70, 2003.
- Culham J, He S, Dukelow S, and Verstraten FA. Visual motion and the human brain, what has neuroimaging told us? *Acta Psychol (Amst)*, 107:69–94, 2001.
- Culham JC, Brandt SA, Cavanagh P, Kanwisher NG, Dale AM, and Tootell RB. Cortical fMRI activation produced by attentive tracking of moving targets. *J Neurophysiol*, 80:2657–70, 1998.
- Dale AM, Fischl B, and Sereno MI. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *NeuroImage*, 9:179–194, 1999.
- Damasio AR, Damasio H, and Van GW Hoesen. Prosopagnosia: anatomic basis and behavioral mechanisms. *Neurology*, 32:331–41, 1982.
- Damasio H, Grabowski TJ, Tranel D, Hichwa RD, and Damasio AR. A neural basis for lexical retrieval. *Nature*, 380:499–505, 1996.
- DeAngelis GC and Newsome WT. Organization of disparity-selective neurons in macaque area mt. *J Neurosci*, 19:1398–415, 1999.

Duda RO, Hart PE, and Stork DG. *Pattern Classification*. Wiley and Sons Inc, 2001.

Edelman S, Grill-Spector K, Kushnir T, and Malach R. Towards direct visualization of the internal shape space by fMRI. *Psychobiology (special issue on Cognitive Neuroscience of Object Representation and Recognition)*, 26:309–321, 1998.

Epstein R, Harris A, Stanley D, and Kanwisher N. The parahippocampal place area: recognition, navigation, or encoding? *Neuron*, 23:115–25, 1999.

Epstein R and Kanwisher N. A cortical representation of the local visual environment. *Nature*, 392(6676):598–601, 1998.

Evans JJ, Heggs AJ, Antoun N, and Hodges JR. Progressive prosopagnosia associated with selective right temporal lobe atrophy—a new syndrome? *Brain*, 118 (Pt 1):1–13, 1995.

Fischl B, Sereno MI, and Dale AM. Cortical surface-based analysis. II. Inflation, flattening, and a surface-based coordinate system. *NeuroImage*, 9:195–207, 1999.

Fisher RA. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7:179–188, 1936. Reprinted in Contributions to Mathematical Statistics, John Wiley: New York (1950).

Formisano E, Kim DS, Di F Salle, Van PF de Moortele, Ugurbil K, and Goebel R. Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron*, 40(4):859–869, 2003.

Friston KJ, Holmes AP, Worsley KJ, Poline JP, Frith CD, and Frackowiak S R J. Statistical parametric maps in functional imaging: a general linear approach. *Human Brain Mapping*, 2:189–210, 1995.

Friston KJ, Worsley KJ, Frackowiak RSJ, Mazziotta JC, and C A Evans. Assessing the significance of focal activations using their spatial extent. *Human Brain Mapping*, 1:214–220, 1994.

Gainotti G, Barbier A, and Marra C. Slowly progressive defect in recognition of familiar people in a patient with right anterior temporal atrophy. *Brain*, 126:792–803, 2003.

Gauthier I, Tarr MJ, Moylan J, Skudlarski P, Gore JC, and W A Anderson. The fusiform “face area” is part of a network that processes faces at the individual level. *J Cogn Neurosci*, 12:495–504, 2000.

Goebel R. A fast automated method for flattening cortical surfaces. In *Proceedings of the 6th International Conference on Functional Mapping of the Human Brain, June 12-16, 1999, San Antonio, Texas, USA*, 2000. NeuroImage, 11, S680.

- Goebel R, Khorram-Sefat D, Muckli L, Hacker H, and Singer W. The constructive nature of vision: direct evidence from functional magnetic resonance imaging studies of apparent motion and motion imagery. *Eur J Neurosci*, 10:1563–73, 1998.
- Goebel R and Singer W. Cortical surface-based statistical analysis of functional magnetic resonance imaging data. In *Proceedings of the 5th International Conference on Functional Mapping of the Human Brain, June 22-26, 1999, Düsseldorf, Germany*, 1999. NeuroImage, 9: S64.
- Grabowski TJ, Damasio H, Tranel D, Ponto LL, Hichwa RD, and Damasio AR. A role for left temporal pole in the retrieval of words for unique entities. *Hum Brain Mapp*, 13:199–212, 2001.
- Grill-Spector K, Kushnir T, Edelman S, Avidan G, Itzhak Y, and Malach R. Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron*, 24:187–203, 1999.
- Grill-Spector K and Malach R. fmr-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychol (Amst)*, 107:293–321, 2001.
- Gruetter R and Tkac I. Field mapping without reference scan using asymmetric echo-planar techniques. *Magn Reson Med*, 43(2):319–323, 2000.
- Hasson U, Avidan G, Deouell LY, Bentin S, and Malach R. Face-selective activation in a congenital prosopagnosic subject. *J Cogn Neurosci*, 15:419–31, 2003b.
- Hasson U, Harel M, Levy I, and Malach R. Large-scale mirror-symmetry organization of human occipito-temporal object areas. *Neuron*, 37:1027–41, 2003a.
- Hasson U, Levy I, Behrmann M, Hendler T, and Malach R. Eccentricity bias as an organizing principle for human high-order object areas. *Neuron*, 34:479–90, 2002.
- Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, and Pietrini P. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293:2425–30, 2001.
- Haxby JV, Hoffman EA, and Gobbini MI. The distributed human neural system for face perception. *Trends Cogn Sci*, 4:223–233, 2000.
- Hoffman EA and Haxby JV. Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nat Neurosci*, 3:80–4, 2000.
- Ishai A, Ungerleider LG, Martin A, Schouten JL, and Haxby JV. Distributed representation of objects in the human ventral visual pathway. *Proc Natl Acad*

*Sci USA*, 96:9379–84, 1999.

Kanwisher N. Domain specificity in face perception. *Nat Neurosci*, 3:759–6, 2000.

Kanwisher N, McDermott J, and Chun MM. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci*, 17:4302–11, 1997.

Kanwisher N, Stanley D, and Harris A. The fusiform face area is selective for faces not animals. *Neuroreport*, 10:183–7, 1999.

Kanwisher N, Tong F, and Nakayama K. The effect of face inversion on the human fusiform face area. *Cognition*, 68:B1–11, 1998.

Kiebel SJ, Goebel R, and Friston KJ. Anatomically informed basis functions. *NeuroImage*, 11:656–667, 2000.

Kourtzi Z, B HHülthoff, Erb M, and Grodd W. Object-selective responses in the human motion area mt/mst. *Nat Neurosci*, 5:17–8, 2002.

Kourtzi Z and Kanwisher N. Representation of perceived object shape by the human lateral occipital complex. *Science*, 293:1506–9, 2001.

Kriegeskorte N and Goebel R. An efficient algorithm for topologically correct segmentation of the cortical sheet in anatomical mr volumes. *NeuroImage*, 14:329–346, 2001.

Kriegeskorte N and Goebel R. Information-based multivariate functional brain mapping. In *Proceedings of the 8th International Conference on Functional Mapping of the Human Brain, June 2-6, 2002, Sendai, Japan*, 2002. NeuroImage, 16, No 2, Part 2.

Krzanowski WJ. *Principles of Multivariate Analysis: A User's Perspective*. Clarendon Press: Oxford, 1988.

Kuskowski MA and Pardo JV. The role of the fusiform gyrus in successful encoding of face stimuli. *Neuroimage*, 9:5, 1999.

Lee JS. Digital image smoothing and the sigma filter. *Computer Vision, Graphics, and Image Processing*, 24:255–269, 1983.

Levy I, Hasson U, Avidan G, Hendler T, and Malach R. Center-periphery organization of human object areas. *Nat Neurosci*, 4:533–9, 2001.

Linden DE, Kallenbach U, Heinecke A, Singer W, and Goebel R. The myth of upright vision. a psychophysical and functional imaging study of adaptation to inverting spectacles. *Perception*, 28:469–481, 1999.

Luo Y, de RA Graaf, DelaBarre L, Tannus A, and Garwood M. Bistro: an

outer-volume suppression method that tolerates rf field inhomogeneity. *Magn Reson Med*, 45(6):1095–1102, 2001.

MacDonald D, Kabani N, Avis D, and Evans AC. Automated 3-d extraction of inner and outer surfaces of cerebral cortex from MRI. *NeuroImage*, 12:340–356, 2000.

Malach R, Levy I, and Hasson U. The topography of high-order human object areas. *Trends Cogn Sci*, 6:176–184, 2002.

Malach R, Reppas JB, Benson RR, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady TJ, Rosen BR, and Tootell RB. Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc Natl Acad Sci U S A*, 92:8135–9, 1995.

Marotta JJ, Genovese CR, and Behrmann M. A functional mri study of face recognition in patients with prosopagnosia. *Neuroreport*, 12:1581–7, 2001.

Maunsell JH and Van DC Essen. Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. *J Neurophysiol*, 49:1127–47, 1983a.

Maunsell JH and Van DC Essen. Functional properties of neurons in middle temporal visual area of the macaque monkey. II. Binocular interactions and sensitivity to binocular disparity. *J Neurophysiol*, 49:1148–67, 1983b.

McCarthy G, Puce A, Gore JC, and Allison T. Face-specific processing in the human fusiform gyrus. *J Cogn Neurosci*, 9:605–610, 1997.

Nichols TE and Holmes AP. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum Brain Mapp*, 15(1):1–25, 2002.

Orban GA, Dupont P, De B Bruyn, Vogels R, Vandenberghe R, and Mortelmans L. A motion area in human visual cortex. *Proc Natl Acad Sci USA*, 92:993–97, 1995.

Orban GA, Sunaert S, Todd JT, Van P Hecke, and Marchal G. Human cortical regions involved in extracting depth from motion. *Neuron*, 24:929–40, 1999.

Pack CC and Born RT. Temporal dynamics of a neural solution to the aperture problem in visual area mt of macaque brain. *Nature*, 409:1040–2, 2001.

Paradis AL, Cornilleau-Peres V, Droulez J, -F Van P de Moortele, Lobel E, Berthoz A, Le D Bihan, and Poline JB. Visual perception of motion and 3-d structure from motion: an fMRI study. *Cereb Cortex*, 10:772–83, 2000.

Perrett DI, Hietanen JK, Oram MW, and Benson PJ. Organization and functions of cells responsive to faces in the temporal cortex. *Philos Trans R Soc Lond B Biol Sci*, 335:23–30, 1992.

- Perrett DI, Rolls ET, and Caan W. Visual neurones responsive to faces in the monkey temporal cortex. *Exp Brain Res*, 47(3):329–42, 1982.
- Petit L, Dubois S, Tzourio N, Dejudin S, Crivello F, Michel C, Etard O, Denise P, Roucoux A, and Mazoyer B. Pet study of the human foveal fixation system. *Human Brain Mapping*, 8:28–43, 1999a.
- Petit L and Haxby JV. Functional anatomy of pursuit eye movements in humans as revealed by fMRI. *J Neurophysiol*, 81:463–71, 1999b.
- Pfeuffer J, -F Van P de Moortele, Yacoub E, Shmuel A, Adriany P, andersen G, Merkle H, Garwood M, Ugurbil K, and Hu X. Zoomed functional imaging in the human brain at 7 tesla with simultaneous high spatial and high temporal resolution. *NeuroImage*, 17:272–286, 2002.
- Poline JB, Worsley KJ, Evans AC, and Friston KJ. Combining spatial extent and peak intensity to test for activations in functional imaging. *NeuroImage*, 5:83–96, 1997.
- Puce A, Allison T, Gore JC, and McCarthy G. Face-sensitive regions in human extrastriate cortex studied by functional MRI. *J Neurophysiol*, 74(3):1192–9, 1995.
- Rolls ET. Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas. *Philos Trans R Soc Lond B Biol Sci*, 335:11–20, 1992.
- Rossion B, Caldara R, Seghier M, -M Schuller A, Lazeyras F, and Mayer E. A network of occipito-temporal face-sensitive areas besides the right middle fusiform gyrus is necessary for normal face processing. *Brain*, 126:1–15, 2003.
- Sáry G, Vogels R, and Orban GA. Cue-invariant shape selectivity of macaque inferior temporal neurons. *Science*, 260:995–7, 1993.
- Sereno ME, Trinath T, Augath M, and Logothetis NK. Three-dimensional shape representation in monkey cortex. *Neuron*, 33(4):635–52, 2002.
- Sereno MI, Dale AM, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, Rosen BR, and Tootell RB. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science*, 268(5212):889–893, 1995.
- Sergent J, Ohta S, and MacDonald B. Functional neuroanatomy of face and object processing a positron emission tomography study. *Brain*, 115:15–36, 1992.
- Singh KD, Smith AT, and Greenlee MW. Spatiotemporal frequency and direction sensitivities of human visual areas measured using fMRI. *NeuroImage*, 12:550–64, 2000.
- Smith AT, Greenlee MW, Singh KD, Kraemer FM, and Hennig J. The processing of first- and second-order motion in human visual cortex assessed by func-

- tional magnetic resonance imaging (fMRI). *J Neurosci May*, 15;18(10):3816–30, 1998.
- Spiridon M and Kanwisher N. How distributed is visual category information in human occipito-temporal cortex? an fMRI study. *Neuron*, 35:1157–65, 2002.
- Sugiura M, Kawashima R, Nakamura K, Sato N, Nakamura A, Kato T, Hatano K, Schormann T, Zilles K, Sato K, Ito K, and Fukuda H. Activation reduction in anterior temporal cortices during repeated recognition of faces of personal acquaintances. *Neuroimage*, 13:877–90, 2001.
- Talairach J and Tournoux P. *Co-Planar Stereotaxic Atlas of the Human Brain*. NY: Thieme Medical Publishers, 1988.
- Tanaka K. Mechanisms of visual object recognition studied in monkeys. *Spat Vis*, 13:147–63, 2000.
- Tanaka K, Saito H, Fukada Y, and Moriya M. Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *JNeurophysiol*, 66(1):170–189, 1991.
- Tong F, Nakayama K, Moscovitch M, Weinrib O, and Kanwisher N. Response properties of the human fusiform face area. *Cognitive Neuropsychology*, 17 (1/2/3):257–279, 2000.
- Tootell RB, Mendola JD, Hadjikhani NK, Ledden PJ, Liu AK, Reppas JB, Sereno MI, and Dale AM. Functional analysis of V3a and related areas in human visual cortex. *J Neurosci*, 17(18):7060–78, 1997.
- Tootell RB, Reppas JB, Kwong KK, Malach R, Born RT, Brady TJ, Rosen BR, and Belliveau JW. Functional analysis of human mt and related visual cortical areas using magnetic resonance imaging. *J Neurosci*, 15:3215–30, 1995.
- Tranel D, Damasio H, and Damasio AR. A neural basis for the retrieval of conceptual knowledge. *Neuropsychologia*, 35:1319–27, 1997.
- Treue S, Husain M, and Andersen RA. Human perception of structure from motion. *Vision Res*, 31:59–75, 1991.
- Tsunoda K, Yamane Y, Nishizaki M, and Tanifuji M. Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. *Nat Neurosci*, 4(8):832–838, 2001.
- Ugurbil K, Toth L, and Kim DS. How accurate is magnetic resonance imaging of brain function? *Trends Neurosci*, 26:108–14, 2003.
- Ungerleider LG and Mishkin M. Two cortical visual systems. in: Analysis of visual behavior (ingle dj, ed), pp 549–586. Cambridge, MA:MIT Press, 1982.
- Van Essen DC and Drury HA. Structural and functional analyses of human

cerebral cortex using a surface-based atlas. *J Neurosci*, 17:7079–7102, 1997.

Van Essen DC, Drury HA, Joshi S, and Miller MI. Functional and structural mapping of human cerebral cortex: solutions are in the surfaces. *Proc Natl Acad Sc USA*, 95:788–795, 1998.

Van Oostende S, Sunaert S, Van P Hecke, Marchal G, and Orban GA. The kinetic occipital (ko) region in man: an fMRI study. *Cereb Cortex*, 7(7):690–701, 1997.

Vaughan JT, Garwood M, Collins CM, Liu W, DelaBarre L, Kim SG, Adriany G, Andersen P, Merkle H, Goebel R, Smith MB, and Ugurbil K. 7t versus 4t: Rf power, homogeneity, and signal-to-noise comparison in head images. *Magnetic Resonance in Medicine*, 46(1):24–30, 2001.

Wager TD and Nichols TE. Optimization of experimental design in fMRI: a general framework using a genetic algorithm. *Neuroimage*, 18:293–309, 2003.

Wallach H and O’Connell DN. The kinetic depth effect. *J Exp Psychol*, 45:205–217, 1953.

Xiao DK, Marcar VL, Raiguel SE, and Orban GA. Selectivity of macaque MT/V5 neurons for surface orientation in depth specified by motion. *Eur J Neurosci*, 9:956–64, 1997.

Zeki S, Watson JD, Lueck CJ, Friston KJ, Kennard C, and Frackowiak RS. A direct demonstration of functional specialization in human visual cortex. *J Neurosci*, 11:641–9, 1991.



# Summary

Motivated by the question how visual images of natural objects are processed by the human brain, this thesis combines empirical and methodological contributions to cognitive neuroscience. All contributions involve magnetic resonance imaging (MRI), a noninvasive method allowing the measurement of the anatomy and activity of the human brain. The methodological developments concern the analysis of anatomical and functional MRI (fMRI) data and are motivated by the neuroscientific questions addressed. Each methodological development (Chapters 1 and 3) is applied in the empirical investigation presented in the following chapter (Chapters 2 and 4), but the novel methodologies are generally applicable beyond the domain of object-vision studies.

## Chapter 1: Topologically correct cortex segmentation

To understand the function of the human cortex, we need to be interested in its anatomy. We present a method revealing the spatial structure of the cortex in anatomical magnetic resonance volumes. An explicit representation of the unique folding structure of each individual subject's cortex (Figure 1.1) allows the visualization of activity distributions on cortical flatmaps (see Chapter 2, Figure 2.5) and is required for a number of advanced methods of statistical analysis of functional neuroimaging data.

The human cortex is a folded sheet: Opposite banks of a sulcus are nowhere grown together through the pia mater to form rings. This interesting topological property of ringlessness is built into our method as an a-priori assumption. Our method ensures that the resulting representation of the cortex is topologically consistent with normal human neuroanatomy.

We start from a rough presegmentation of the white matter obtained by established techniques including intensity thresholding. Such presegmentations are 3D objects (defined in a voxel grid) that typically contain many ring structures, i.e. topological errors. Our method cuts each ring at its thinnest point or fills in its hole to obtain a ringless (and thus topologically correct) object with minimal changes to the presegmentation (Figure 1.3).

The core component of our algorithm is a region growing process that is seeded deep inside the white-matter voxel object provided by the presegmentation. The region grows inside the object, keeping maximal distance to the object surface, thus avoiding obstacles and becoming increasingly complex as its shape approaches that of the object. While the region grows, it is never allowed to touch

itself such that a ring would result (Figure 1.2). In the final state, the region represents a topology-corrected replacement for the white-matter presegmentation. The algorithm involves two such regions growing from within and without the white-matter object until their boundaries come to define the cortical sheet interactively.

## **Chapter 2: Object recognition from a visual motion flow-field**

Visual object recognition is known to involve the ventral stream of visual cortex. The dorsal stream, in contrast, processes information about object motion and location. Dorsal and ventral streams are traditionally conceptualized as separate subsystems, though they are known to be in interaction. When static photographs of objects are viewed, recognition may take place in the ventral stream with minimal involvement of the dorsal stream. This is the experimental situation in most object-recognition studies, including that presented in Chapter 4. In the natural world, in contrast, objects are often in motion relative to the observer, so the dorsal stream is coactivated.

To study the interaction between dorsal and ventral visual subsystems during object recognition, we constructed moving-dot stimuli that define object surfaces (Figure 2.1). Moving dots can evoke a percept of a 3-D object's spatial structure in the absence of other visual cues. This phenomenon, called structure from motion (SFM), suggests that the motion flowfield represented in the dorsal stream can form the basis of object recognition performed in the ventral stream. SFM processing is likely to contribute to object recognition whenever there is relative motion between the observer and the object viewed.

Our SFM stimuli elicit strong surface percepts and involve an extended dorso-ventral network of cortical regions (Figure 2.2 and Figure 2.5). From early visual areas, this network extends dorsally into the human motion complex and parietal regions, and ventrally into object-related cortex. When the surface encoded in the SFM stimulus was a face, a response was elicited in the fusiform face region, a face-selective ventral-stream region found in the fusiform gyrus.

Our results suggest that the motion-flowfield representation in the dorsal-stream human motion complex can indeed serve as the basis of ventral-stream object recognition. Furthermore, the motion complex appears to represent not merely the motion flowfield but also the surface structure of the object. The motion complex and a region in the intraparietal sulcus reflected the motion of the object implicit to the SFM stimulus—independent of the retinal motion flowfield. This suggests a higher-order dorsal-stream representation of object motion.

## **Chapter 3: Information-based functional brain mapping**

Visual objects are represented by a combinatorial code in ventral visual cortex. Combinatorial representation means that a perceived object's category or identity is not unambiguously encoded in any single cell's response, but can be inferred from the activity pattern across a population of cells. This widely accepted view is the product of a paradigm shift in electrophysiology, which took

place in the 1970s and was incited on the empirical front by the lack of evidence for single cells that exclusively respond to particular complex objects and on the theoretical front by the representational power afforded by combinatorial codes. Beyond the domain of object vision, combinatorial cell-population codes are believed to represent complex perceptual, cognitive, and motor information.

Recent studies suggests that population codes play a major representational role not only at the microscopic scale of individual neurons but at the macroscopic scale accessible to fMRI as well. To be explicit about the spatial scale, we refer to the latter variety as *macropopulation* codes. Neuroimaging may be undergoing a paradigm shift similar to that of electrophysiology—at a coarser spatial scale of analysis.

To be able to locate such macroscopic population codes in the brain, we propose a novel method of functional brain mapping based on fMRI data. The currently dominant brain mapping approach is *activation-based* in that it localizes regions that are activated as a whole in one condition as compared to a control condition. Our approach, by contrast, is *information-based* in that it localizes regions whose intrinsic spatial response pattern contains information about the experimental condition, thus addressing a more general question. The information in local contiguous response patterns is mapped out by scanning the imaged volume with a spherical “searchlight” (Figure 3.3), whose contents are analyzed multivariately at each location in the brain.

Our approach addresses an important current challenge in functional brain mapping: While human fMRI already operates at a resolution of one to two millimeters, the spatial correspondence mapping across brains provided by Talairach space has a much lower spatial precision. Information in the fine-grained structure of the data, thus, averages out in conventional group analysis, entailing a reduction of the spatial complexity of the measured signal by at least an order of magnitude.

A better (e.g. cortex-based) spatial correspondence mapping may provide greater precision, but the problem is not merely one of methodology. Instead it concerns the fundamental neuroscientific question how closely different brains are related in terms of their functional structure and its spatial layout. It is certain that a cross-subject spatial correspondence cannot be established at the level of single neurons. For most parts of the human brain, it is not known to what precision a functional spatial correspondence exists.

We therefore approach group analysis without the assumption of precise spatial correspondence. While allowing for the fine-grained spatial response patterns to be individually unique, we do assume that corresponding macroscopic regions encode the same information. Each macroscopic region’s fine-grained response pattern (comprising many voxels) is multivariately analyzed for each subject separately. Group analysis is performed at a coarser scale by combining all subjects’ information effects for a given macroscopic region. This two-scale approach allows us to statistically integrate individually unique fine-grained multivariate effects to obtain group-level characterizations of the distribution of a given type of information across brain regions.

## Chapter 4: Single-object-image response patterns in ventral visual cortex

To explore the representation of visual objects in ventral visual cortex with the method of information-based functional brain mapping, we measured brain activity with fMRI while subjects viewed grayscale photographs of faces and houses.

Previous neuroimaging studies of object recognition have focused on the representation of object categories. They demonstrated the existence of category-selective regions as well as the widely distributed representation of category information. These studies analyzed differences between category-average responses.

In this chapter, we explore ventral-stream response patterns to *single images* of objects (two faces and two houses, see Figure 4.1). The single-image approach eliminates the somewhat arbitrary grouping of images into categories, brings us closer to what happens on a single trial of perception of a particular object, and allows us to investigate how exemplars within the same category are (differentially) represented. To match this move of focus into the categories in terms of measurement, we target activity patterns at a finer spatial scale than previous studies, utilizing the high spatial resolution (1–2 mm) of fMRI at high magnetic field strength (3T and 7T). In our analysis, we look into each region of interest, considering not merely its spatial-average response but the shape of its spatial response pattern. Furthermore, we search the measured volume for macropopulation codes using information-based functional brain mapping.

Fusiform face and parahippocampal place regions both contain category information not only in the spatial-average response usually studied but also in the spatial shape of their response patterns (Figure 4.2). The combinatorial representations in these posterior ventral-stream regions seem to reflect mainly object category at the resolution provided by fMRI—possibly abstracting from the difference between exemplars of the same category. Whereas face-exemplar information appeared absent in the fusiform face region, it was found more anteriorly in ventral temporal cortex (Figure 4.5), casting doubt on the view that the fusiform face region is the main locus of individual-level face representation.

## Conclusion and future directions

This thesis demonstrates the advantages of developing theory, experiment, and statistical analysis methods in close interaction. This interaction is in evidence in the first half (Chapters 1 and 2), but even closer in the second half (Chapters 3 and 4). In the second half, a theoretical prediction (the existence of macropopulation codes) motivated the development of a new analysis technique (information-based functional brain mapping). In conjunction with a novel experimental approach (the single-object-image approach), the analysis technique lead to the challenge of a widespread assumption (that the fusiform face region represents faces at the individual level) and suggested the existence of a functional region that has not previously been described (the anterior inferotemporal face-exemplar region).

These advances merely represent a first step in the context of a more general research program of information-based functional brain mapping. Visual object recognition will be a major domain of application, but the method can be applied in any domain of brain function. In the currently dominant activation-based approach to brain mapping, information contained in the spatial microstructure of the fMRI signal is discarded—by explicit spatial smoothing, by Talairach-based intersubject integration, or by selective reporting of large foci of activation only. Since our approach has the potential to utilize fine-grained spatial information, while providing a coarse-grained intersubject summary of information effects, it is likely to become more important as fMRI spatial resolution increases. Many statistical, empirical, and theoretical challenges are yet to be confronted.



# Samenvatting

Gemotiveerd door de vraag hoe de menselijke hersenen visuele beelden van natuurlijke objecten verwerken, worden in deze dissertatie empirische en methodologische bijdragen aan cognitieve neurowetenschappen gecombineerd. Bij alle bijdragen, empirisch en methodologisch, is magnetic resonance imaging (MRI) betrokken, een methode waarmee op onschadelijke wijze de anatomie en activiteit van de menselijke hersenen gemeten kan worden. De methodologische ontwikkelingen hebben betrekking op de analyse van anatomische en functionele MRI (fMRI) gegevens en vinden hun oorsprong in de neurowetenschappelijke probleemstellingen die worden behandeld.

Elke nieuw ontwikkelde methode (Hoofdstuk 1 en 3) wordt toegepast in de empirische onderzoeken die worden gepresenteerd in het daaropvolgende hoofdstuk (Hoofdstuk 2 en 4), hoewel de nieuwe methoden ook algemener zijn toe te passen dan in het domein van object visie studies.

## Hoofdstuk 1: Topologisch correcte cortexsegmentatie

Om de functie van de menselijke cortex te begrijpen, dient men zijn anatomie nader te onderzoeken. Wij presenteren een methode die de ruimtelijke structuur van de cortex reconstrueert uit anatomische magnetische resonantie beelden. Een expliciete representatie van de unieke plooien van de cortex van iedere individuele proefpersoon (Figuur 1.1) maakt het mogelijk activatiepatronen op het tweedimensionale corticale oppervlak zichtbaar te maken (zie Hoofdstuk 2, Figuur 2.5). Deze representatie is vereist voor een aantal geavanceerde statistische analysemethoden van fMRI data.

De menselijke cortex is een geplooid oppervlak: tegenover elkaar liggende zijden van een sulcus buigen nergens over de pia mater om ringen te vormen. Deze interessante topologische eigenschap van ringloosheid is een a-priori veronderstelling in onze methode. Onze methode verzekert dat de resulterende representatie van de cortex topologisch consistent is met de normale menselijke neuro-anatomie.

We beginnen vanuit een ruwe presegmentatie van de witte stof die verkregen is met behulp van gevestigde technieken inclusief intensiteitsbegrenzing. Zulke presegmentaties zijn 3D objecten (gedefinieerd in een voxelrooster), die doorgaans vele ringstructuren vertonen, dwz. topologische fouten. Onze methode snijdt elke ring op het dunste punt door of vult de leemte om een ringloos (en dus topologisch juist) object te verkrijgen met minimale wijzigingen aan de presegmentatie (Figuur 1.3).

De kern van ons algoritme is een groeiproces dat wordt geïnitieerd diep in het witte stof voxelobject dat het resultaat is van de presegmentatie. De regio groeit in het object, dusdanig dat het maximale afstand houdt tot het objectoppervlak en zodoende obstakels vermijdt. De vorm wordt steeds complexer wanneer het de vorm van het object benadert. Terwijl de regio groeit, wordt het deze nooit toegestaan om zichzelf te raken zodat een ring zou ontstaan (Figuur 1.2). In de uiteindelijke toestand van het object representeert de regio een topologisch gecorrigeerde vervanging van de presegmentatie van de witte stof. Het algoritme gebruikt twee van dergelijke regio's die respectievelijk van binnenuit en van buitenaf naar het witte materie object toe groeien totdat hun grenzen op interactieve wijze het corticale oppervlak definiëren.

## Hoofdstuk 2: Objectherkenning vanuit een visueel bewegings-stroomveld

Het is bekend dat het ventrale pad van de visuele cortex betrokken is bij visuele objectherkenning. Dit in tegenstelling tot het dorsale pad, dat informatie over objectbeweging en -locatie verwerkt. Dorsale en ventrale paden worden traditioneel voorgesteld als twee gescheiden subsystemen, hoewel bekend is dat ze informatie uitwisselen. Wanneer statische foto's van objecten worden bekeken, kan het herkenningsproces plaatsvinden in het ventrale pad, met een minimale betrokkenheid van het dorsale pad. Dit is de experimentele opstelling in de meeste onderzoeken naar visuele objectherkenning, inclusief het onderzoek gepresenteerd in Hoofdstuk 4. In de natuurlijke wereld zijn objecten echter vaak in beweging ten opzichte van de waarnemer, zodat ook het dorsale pad geactiveerd is.

Om de interactie tussen de dorsale en ventrale visuele subsystemen gedurende objectherkenning te bestuderen, construeerden we stimuli van bewegende punten die tezamen een objectoppervlak definiëren (Figuur 2.1). Bewegende punten kunnen in de afwezigheid van andere visuele aanwijzingen de waarneming van de ruimtelijke structuur van een object oproepen. Dit fenomeen, structuur uit beweging (structure from motion, SFM) genoemd, suggereert dat het bewegings-stroomveld, gerepresenteerd in het dorsale pad, de basis kan vormen voor de objectherkenning die wordt uitgevoerd in het ventrale pad. Het is aannemelijk dat SFM verwerking bijdraagt aan objectherkenning wanneer er relatieve beweging is tussen de waarnemer en het object dat waargenomen wordt.

Onze SFM stimuli roepen sterke waarnemingen op van een oppervlak waarbij een uitgebreid dorsoventraal netwerk van corticale gebieden betrokken is (Figuur 2.2 en Figuur 2.5). Vanuit de vroege visuele gebieden strekt het netwerk zich dorsaal uit tot in het human motion complex (een gebied gevoelig voor bewegende stimuli) en pariëtale gebieden, en ventraal tot in de inferotemporale cortex, betrokken bij objectherkenning. Wanneer het oppervlak dat in de SFM stimulus werd getoond een gezicht definieerde, werd een reactie opgeroepen in het fusiform gezichtsgebied, een gebied dat gevoelig is voor gezichtsstimuli en dat gelegen is op de fusiforme gyrus in het ventrale pad.

Onze resultaten dragen bij aan het idee dat de representatie van het bewegings-stroomveld in het dorsale human motion complex inderdaad als de basis kan dienen voor ventrale objectherkenning. Bovendien blijkt het motion complex niet

enkel het bewegingsstroomveld te representeren, maar ook de oppervlaktestructuur van het object. Het motion complex en een gebied in de intrapariëtale sulcus gaven de beweging van het object weer die impliciet was in de SFM stimulus. Dit suggereert dat er een hogere order representatie van objectbeweging in het dorsale pad bestaat.

### Hoofdstuk 3: Informatiegestuurde functionele hersenkartering

Visuele objecten worden in de ventrale visuele cortex gerepresenteerd door een combinatorische code. Een combinatorische representatie houdt in dat de categorie of identiteit van een waargenomen object niet ondubbelzinnig wordt gecodeerd in de reactie van één individuele cel, maar afgeleid kan worden uit het gezamenlijke activatiepatroon van een groep cellen. Dit wijds geaccepteerde standpunt is het product van een paradigmaverschuiving in de elektrofysiologie, welke plaatsvond in de jaren '70. Deze werd aan het empirische front ingeleid door het gebrek aan bewijs voor individuele cellen die exclusief reageren op specifieke complexe objecten en aan het theoretische front door het vermogen van combinatorische codes om een veelvoudigheid aan objecten en hun eigenschappen efficiënt te representeren. Ook buiten het domein van objectwaarneming wordt aangenomen dat combinatorische celpopulatiecodes complexe perceptuele, cognitieve en motorische informatie representeren.

Recente onderzoeken impliceren dat populatiecodes een hoofdrol spelen in representatie, niet alleen op de microscopische schaal van individuele neuronen maar ook op de macroscopische schaal die toegankelijk is voor fMRI. Om expliciet te zijn over de ruimtelijke schaal, refereren we aan de laatstgenoemde variant als *macropopulatie* codes. Neuroimaging zou een gelijke paradigmaverschuiving kunnen ondergaan als die in de elektrofysiologie—op een grovere ruimtelijke schaal van analyse.

Om het mogelijk te maken dergelijke macropopulatie codes te lokaliseren, presenteren we een nieuwe methode voor het in kaart brengen van hersenfuncties gebaseerd op fMRI gegevens. De op dit moment dominerende benadering tot hersenkartering is gebaseerd op *activatie*, in dat het gebieden lokaliseert die in een conditie als een geheel zijn geactiveerd, vergeleken met de controle conditie. Onze benadering is echter gebaseerd op *informatie*, in dat het gebieden lokaliseert waarvan het intrinsieke ruimtelijke reactiepatroon informatie bevat over de experimentele conditie, en behandelt zodoende een algemenere kwestie. De informatie in lokale, aaneengesloten reactiepatronen is in kaart gebracht door het doorzoeken van het gescande volume met een sferisch “zoeklicht” (Figuur 3.3), waarvan de inhoud op elke locatie met multivariate statistische methoden wordt geanalyseerd.

Onze benadering behandelt een belangrijke actuele uitdaging in functioneel hersenkartering: terwijl fMRI voor mensen al mogelijk is op een resolutie van één of twee millimeter, is het in overeenstemming brengen van de ruimtelijke structuur van meerdere breinen met behulp van de Talairach ruimte veel minder precies. Informatie aanwezig in de fijne structuur van de data wordt zodoende uitgemiddeld in conventionele groepsanalyses, en de ruimtelijke complexiteit van het gemeten signaal wordt daardoor sterk gereduceerd.

Een betere (b.v. cortexgebaseerde) benadering tot het in kaart brengen van de ruimtelijke overeenkomst zou waarschijnlijk nauwkeuriger zijn, maar het probleem is niet alleen maar methodologisch van aard. In plaats daarvan betreft het de fundamentele neurowetenschappelijke vraag hoe nauw de verschillende hersenen aan elkaar gerelateerd zijn met betrekking tot hun functionele en ruimtelijke structuur. Het is zeker dat een ruimtelijke correspondentie over meerdere proefpersonen niet bereikt kan worden op het niveau van individuele neuronen. Voor de meeste delen van het menselijk brein is niet bekend tot op welke hoogte functionele ruimtelijke overeenstemming bestaat.

Daarom benaderen we groepsanalyse zonder de veronderstelling van nauwkeurige ruimtelijke overeenstemming. We gaan er vanuit dat overeenkomstige macroscopische gebieden dezelfde informatie coderen, terwijl we toestaan dat fijnkorrelige ruimtelijke activatiepatronen individueel uniek zijn. Het fijnkorrelige activatiepatroon van elk macroscopisch gebied (dat vele voxels bevat) wordt per proefpersoon met multivariate statistische methoden geanalyseerd. Op een grovere schaal wordt een groepsanalyse uitgevoerd door het combineren van alle informatie-effecten van de proefpersonen voor een gegeven macroscopisch gebied. Deze tweeledige benadering maakt het mogelijk om op groepsniveau de individueel unieke fijnkorrelige multivariate effecten statistisch te integreren om karakteristieken van de verdeling van een gegeven type van informatie over meerdere hersengebieden te verkrijgen.

## Hoofdstuk 4: Reactiepatronen op afbeeldingen van individuele objecten in de ventrale visuele cortex

Om de representatie van visuele objecten in de ventrale visuele cortex te onderzoeken met behulp van de informatiegestuurde functionele hersenkartering, hebben we de hersenactiviteit met fMRI gemeten terwijl de proefpersonen foto's zagen van gezichten en huizen in grijstinten.

Eerdere neuroimaging onderzoeken naar objectherkenning richtten zich op de representatie van objectcategorieën. Deze onderzoeken toonden het bestaan aan van gebieden die selectief zijn voor categorieën evenals de gedistribueerde representatie van de categorie-informatie. In deze onderzoeken werden verschillen in reacties tussen categoriegemiddelden geanalyseerd.

In dit hoofdstuk onderzoeken we reactiepatronen in het ventrale pad op beelden van *individuele* objecten (twee gezichten en twee huizen, zie Figuur 4.1). Deze aanpak elimineert de enigszins arbitraire groepering van plaatjes in categorieën, brengt ons dicht bij wat er plaatsvindt gedurende één enkele waarneming van een bepaald object en staat ons toe te onderzoeken hoe individuele objecten binnen eenzelfde categorie (verschillend) worden gerepresenteerd. Om deze focusverschuiving naar binnen de categorieën zelf door te voeren in de meetmethoden, richten we ons op activatiepatronen op een fijnere ruimtelijke schaal dan eerdere onderzoeken door het gebruik van de hoge spatiele resolutie (1–2 mm) in fMRI op een hoge magnetische veld sterkte (3T en 7T). In onze analyse bekijken we elk gebied van interesse, waarbij we niet alleen de over de ruimte gemiddelde respons in acht nemen maar ook de vorm van het ruimtelijke activatiepatroon. Verder zoeken we in het gemeten volume naar macropopulatie codes met gebruik van informatiegestuurde functionele hersenkartering.

Fusiform gezichts- en parahippocampale plaatsgebieden bevatten beide categorische informatie. Deze informatie betreft niet alleen het over de ruimte gemiddelde activatiepatroon dat gewoonlijk wordt bestudeerd maar ook de ruimtelijke vorm van deze activatiepatronen (Figuur 4.2). De combinatorische representaties in deze posterieure ventrale gebieden lijken, met de resolutie haalbaar met fMRI, voornamelijk de categorie van het object weer te geven—wellicht weg abstraherend van de verschillen tussen individuen in eenzelfde categorie. Waar de individuele gezichtsinformatie afwezig bleek in het fusiform gezichtsgebied, werd het meer anterieur gevonden in de ventrale temporele cortex (Figuur 4.5), wat twijfel oproept over de veronderstelling dat het fusiform gezichtsgebied de belangrijkste rol speelt in de representatie van gezichten op individueel niveau.

## Conclusie en toekomstige richtingen

Deze dissertatie demonstreert de voordelen van het interactief ontwikkelen van theorie, experiment en statistische analysemethoden. Deze interactie is evident in de eerste helft (Hoofdstuk 1 en 2), maar nog duidelijker aanwezig in de tweede helft (Hoofdstuk 3 en 4). In de tweede helft motiveerde een theoretische voorspelling (het bestaan van macropopulatie codes) de ontwikkeling van een nieuwe analysetechniek (informatiegestuurde functionele hersenkartering). Samen met een nieuwe experimentele aanpak (op grond van beelden van individuele objecten), leidde deze analysetechniek tot het ter discussie stellen van een wijdverspreide veronderstelling (dat het fusiform gezichtsgebied gezichten op individueel niveau representeert) en suggereerde het bestaan van een functioneel gebied dat voorheen nog niet is beschreven (het anterieure inferotemporale gebied waarin individuele gezichten gerepresenteerd zijn).

Deze ontwikkelingen representeren slechts een eerste stap in het kader van een algemener onderzoeksprogramma over informatiegestuurde functionele hersenkartering. Visuele objectherkenning zal een belangrijk toepassingsdomein zijn, maar de methode kan worden toegepast in elk domein van cognitieve en perceptuele functies. In de benadering die op dit moment dominant is wordt informatie die gevat is in de ruimtelijke microstructuur van het fMRI signaal genegeerd door expliciete spatiale filtering, door de Talairach-gebaseerde integratie tussen proefpersonen, of door het selectief noemen van enkel de grotere geactiveerde gebieden.

Daar onze benadering het potentieel heeft om fijnkorrelige ruimtelijke informatie te benutten en tegelijkertijd een grofkorrelige samenvatting van informatie-effecten over verschillende proefpersonen te bieden, wordt het waarschijnlijk belangrijker zodra de spatiële resolutie van fMRI toeneemt. Vele statistische, empirische en theoretische uitdagingen moeten nog worden aangegaan.



# Acknowledgments

I would like to thank Rainer Goebel for his friendship. He taught me how to do empirical science. His combination of vision, enthusiasm, and pragmatism is unique and has greatly inspired me. He gave me complete freedom, while at the same time actively contributing to my projects at all levels, from advice on difficult decisions down to hands-on analysis. The atmosphere of free scientific exploration and collaboration in our group has permanently shaped me.

In the course of these projects, scientific interaction has been most intense with my close friends Elia Formisano and Bettina Sorger. Elia shared an apartment with me in Maastricht and our cohabitation appears like a single extended conversation in retrospect. He has been a constant source of scientific wisdom. I have learned a lot from him. Most mental energy has probably been expended between Bettina and me—with forces opposed, in heated scientific discussions, as well as joining forces, in a long series of all-night sessions, preparing experiments to be conducted the following morning or analyzing and writing with abstract deadlines only hours away. Her ideas and raw energy have often challenged me and had a great positive impact.

I thank my friends and collaborators from the labs in Frankfurt and Maastricht: Lars Muckli, David Linden, Marcus Naumer, Miguel Castelo-Branco, Dirk Vorberg, Alard Roebroek, Steffen Egner, and Jens Schwarzbach, and in Minneapolis: Pierre-Francois van de Moortele, Gregor Adriany, Johannes Ritter, Dae-Shik Kim, and Kamil Ugurbil. Thanks also to Peter Bandettini for supporting the information-based functional brain mapping project from its early stages.

Special thanks to Judith Peters and Harry Smit for letting me use their faces as stimuli.

All members of the Maastricht Vision Group and the Neurocognitie department contributed to the great atmosphere I experienced there. I thank my “paranimfen” (thesis-defense supporters) Milene Bonte and Alard Roebroek for their friendship, for all their help with thesis and defense, and for enduring my stubbornness. For the latter I should probably also thank some others listed here.

Hester Breman translated the summary into Dutch and did the layout of the thesis. Much more than for these things, I’d like to thank her for everything else.

Many kinds of support came from my close friend Semfira Kamalova and among the undeniable contributors are my parents, Magnus and Ursula Kriegeskorte, and my grandmothers, Hannelore Kobier and Vroni Kriegeskorte, at which point I should really stop.

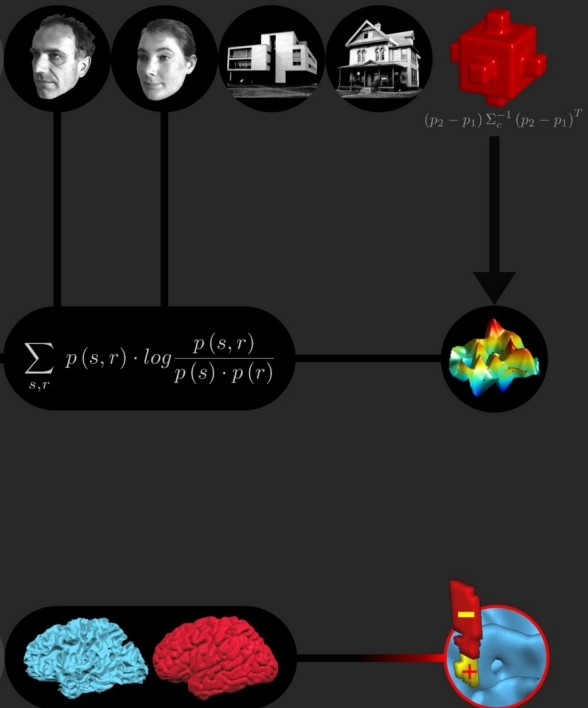
# Curriculum vitae

|           |                                                                                                                                                                                                                                     |
|-----------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1971      | born in Bergisch Gladbach, Germany                                                                                                                                                                                                  |
| 1986–1988 | lived in Kyoto and Tokyo, Japan                                                                                                                                                                                                     |
| 1990      |                                                                                                                                                                                                                                     |
| 1993–1998 | studied psychology and computer science at <i>Universität zu Köln</i> (Cologne, Germany)                                                                                                                                            |
| 1996–1997 | research assistant at the <i>Max Planck Institute for Psychological Research</i> (Munich, Germany. Departments of W. Prinz and G. Gigerenzer) and at <i>Universität Konstanz</i> (Konstanz, Germany. Advisor: J. Delius)            |
| 1998      | finished MA thesis on computational models of rotation-invariant object recognition at <i>Universität zu Köln</i> (Cologne, Germany)                                                                                                |
| 1999      | commenced PhD research on fMRI of visual object recognition in the <i>Department of Neurophysiology</i> (W. Singer) at the <i>Max Planck Institute for Brain Research</i> (Frankfurt, Germany. Advisor: R. Goebel)                  |
| 2000–2003 | PhD research at the <i>Universiteit Maastricht</i> (Maastricht, The Netherlands. Advisor: R. Goebel)                                                                                                                                |
| 2003      | postdoctoral fellow at the <i>Center for Magnetic Resonance Research</i> (K. Ugurbil) at the <i>University of Minnesota</i> (Minneapolis, MN, USA. Advisor: K. Ugurbil)                                                             |
| 2004      | visiting fellow in the <i>Unit on Functional Imaging Methods</i> (P. Bandettini) of the <i>Laboratory of Brain and Cognition</i> (L. Ungerleider), <i>National Institutes of Health</i> (Bethesda, MD, USA. Advisor: P. Bandettini) |

## Contact information

**Nikolaus Kriegeskorte** Unit on Functional Imaging Methods, Laboratory of Brain and Cognition, National Institute of Mental Health, National Institutes of Health, Building 10, Room 1D80, 10 Center Dr. MSC 1148, Bethesda, MD 20892-1148, USA  
**Work phone:** 301-594-9195    **Fax:** 301-402-1370    **Email:** niko@nih.gov  
**Home address:** 4707 Connecticut Ave NW, Apt. 204, Washington, DC 20008, USA  
**Cellphone:** +1-202 374 26 28

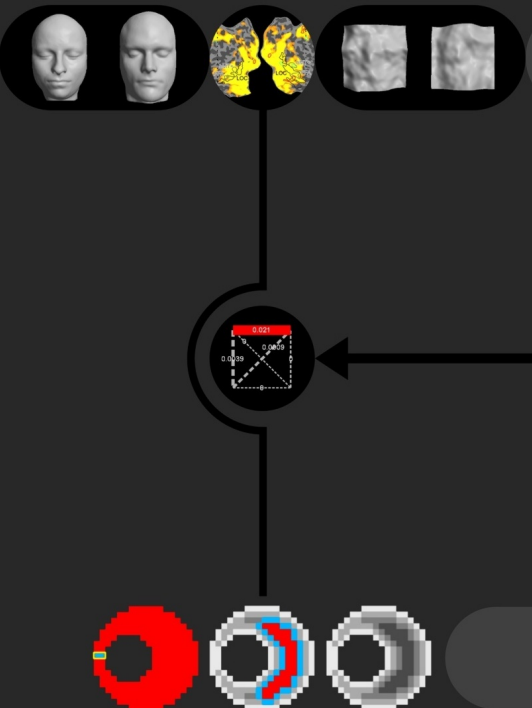
# Functional magnetic resonance **imaging** of the human **object-vision** system



$$\sum_{s,r} p(s,r) \cdot \log \frac{p(s,r)}{p(s) \cdot p(r)}$$

Nikolaus Kriegeskorte

Functional magnetic resonance **imaging** of the human **object-vision** system



Nikolaus Kriegeskorte  
methodological and empirical contributions

ISBN 90-9018729-4