

Human Cortical Object Recognition from a Visual Motion Flowfield

Nikolaus Kriegeskorte,¹ Bettina Sorger,^{1,2} Marcus Naumer,³ Jens Schwarzbach,^{1,4} Erik van den Boogert,⁴ Walter Hussy,² and Rainer Goebel^{1,4}

¹Department of Cognitive Neuroscience, Faculty of Psychology, Universiteit Maastricht, 6229 ER Maastricht, The Netherlands, ²Institute of Psychology, University of Köln, 50931 Köln, Germany; ³Max Planck Institute for Brain Research, Department of Neurophysiology, 60528 Frankfurt am Main, Germany, and ⁴F. C. Donders Centre for Cognitive Neuroimaging, NL-6500 HB Nijmegen, The Netherlands

Moving dots can evoke a percept of the spatial structure of a three-dimensional object 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 three-dimensional control shapes with matched curvature properties. We used two different types of an SFM stimulus with the dots either fixed to the surface of the object 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.

Key words: object recognition; motion processing; structure from motion; functional magnetic resonance imaging; human; cortex; face

Introduction

The primate visual system recovers the three-dimensional 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 three-dimensional structure of the object. The contribution of motion to structure perception has been recognized for a long time (Wallach and O'Connell, 1953) and studied extensively by using moving dot stimuli constructed to minimize other cues (Andersen and Bradley, 1998). The fact that primate observers can perceive the objects implicit to these structure-from-motion (SFM) stimuli suggests cross talk between the dorsal and the ventral visual pathway (Ungerleider and Mishkin, 1982). 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.

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 homolog of a complex of motion-sensitive regions in monkey cortex including the middle temporal area (MT) and its satellite regions, the medial superior temporal area and an area in the fundus of the superior temporal sulcus. Although many regions of primate visual cortex process motion information, hMT+ appears to have a special role (Maunsell and Van Essen, 1983a; Albright, 1984), in that it abstracts from other features of the visual input, including orientation and color, and integrates motion cues over small patches to solve 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). However, these studies 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 review, see Malach et al., 2002). LOC has been shown to be involved in many different types of object perception (Grill-Spector et al., 1999; Amedi et al., 2001) and is conjectured to represent perceived object shape (Kourtzi and Kanwisher, 2001). Notably, LOC and hMT+, although associated with ventral and dorsal stream, respectively, are very close together on the cortex. In anesthetized monkeys, the homologs of these regions and others have been found to be responsive to

Received Sept. 3, 2002; revised Nov. 5, 2002; accepted Nov. 29, 2002.

This research was funded by Universiteit Maastricht and the Donders Centre for Cognitive Neuroimaging. We thank Elia Formisano and two anonymous reviewers for their helpful comments on a draft of this paper.

Correspondence should be addressed to Nikolaus Kriegeskorte, Department of Cognitive Neuroscience, Faculty of Psychology, Universiteit Maastricht, Universiteitssingel 40, 6229 ER Maastricht, The Netherlands. E-mail: n.kriegeskorte@psychology.unimaas.nl

Copyright © 2003 Society for Neuroscience 0270-6474/03/231451-13\$15.00/0

three-dimensional geometrical objects defined by various visual cues, including motion (Serenio et al., 2002).

In this functional magnetic resonance imaging (fMRI) study, we use motion-defined face surfaces as stimuli, allowing us to trace SFM object recognition all the way to a category-specific response in the fusiform face area (FFA) of the ventral stream, a face-selective region found in the human fusiform gyrus anterior to pFs [for FFA, see Kanwisher et al. (1997, 1998, 1999); for previous studies describing face-selective responses, see Perrett et al. (1982), Allison et al. (1994), and Puce et al. (1995)].

Materials and Methods

General experimental rationale

Classical versus on-surface SFM. This study explores SFM object recognition for two radically different forms of SFM encoding. We used the conventional SFM encoding, in which the dots are fixed to the object surface (see Fig. 1A), as well as a novel SFM encoding, in which the dots move on the surface of the object (see Fig. 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 used motion-defined face surfaces to elicit a category-specific response in FFA (see Fig. 1C, left). As SFM control stimuli, we used random surfaces created to have curvature properties similar to those of the face surfaces (see Fig. 1C, right; for details, see below). 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 used moving and static non-SFM random-dot control stimuli closely matched in terms of low-level properties (details below).

Face–nonface categorization task. Subjects fixated on a central cross and performed a face–nonface categorization task, providing a behavioral control of the object-recognition process. Because 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 three-dimensional object) would have allowed subjects to classify the object as a nonface without having formed a continuous surface percept.

Circular aperture eliminates outer object boundary. We used a circular aperture to eliminate the outer boundary of the implicit objects in all stimuli (see Fig. 1A,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 nonmotion 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 (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. First, 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. Second, perceiving the object naturally triggers smooth pursuit, so fixation 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 point to a fixed position on the front surface of the object. This point on the surface of the object is the center of its motion and thus does not move in three-dimensional 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 appropriate stimulus conditions (see below, Experimental design and task). 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 (see Fig. 1A,B, respectively), as well as all random-dot control stimuli shared the following features. Displays consisted of ~1000 dots within a circular aperture. The dots (single pixels) were square and of a width of ~0.06° 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 ~15° visual angle.

For classical and on-surface SFM experimental stimuli, the three-dimensional 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 under 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 three-dimensions. 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 Fig. 1A). The two rotatory oscillations are harmonic, have a 90° phase shift relative to each other, and have an amplitude of 10°.

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 three-dimensional 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 initially were distributed randomly 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. Although 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° phase shift), causing the projections to describe cyclic trajectories on the surface. Motion is essential to the three-dimensional surface percepts evoked by these displays; when the dynamic stimuli are halted, the percept immediately disintegrates. The on-surface SFM stimuli were matched closely 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 that the dots were projected onto (face or random shape) was stationary. In the second, the implicit object, like the laser array, precessed (see Fig. 1B).

Encoded surfaces: faces and random shapes. The two face surfaces used (see Fig. 1C, left) were obtained by reconstructing two human heads as polygon-mesh surfaces on the basis of whole-head T1-weighted anatomical magnetic-resonance (MR) scans. Our SFM techniques (see above) ensured that the faces were always presented approximately in frontal view. The two matched random three-dimensional shapes (see Fig. 1C, right) were created to have curvature properties similar to the faces by randomizing the phases in the Fourier transform of the depth maps of the faces 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 depth maps of the faces were periodicized before 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 the motion of each dot 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 the trajectory of each single dot as well as the temporal phase relationship of the trajectories of the dots. The resulting display was restricted to the same circular aperture used in the experimental stimuli. Because of 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 and the other from an on-surface SFM face stimulus. Because the two faces and the random three-dimensional control shapes used had qualitatively and quantitatively similar curvature properties, these two motion control stimuli have motion-trajectory content similar to that of 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 252×252 pixel gray-scale 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 10×10 pixels (resulting in 25×25 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 ($\sim 15^\circ$ visual angle) as the SFM face stimuli.

Retinotopy mapping. Visual area borders were determined using a conventional polar mapping technique (Sereni et al., 1995; Goebel et al., 1998). Rotating checkerboard wedges were 22.5° in polar angle width and spanned $0.5\text{--}20^\circ$ in eccentricity. A fixation point was shown at the center of the screen. Wedges reversed in contrast 8.3 times per second and rotated counterclockwise starting at the upper vertical meridian. Each participant completed 10 cycles (64 sec per cycle) that lasted 10 min 40 sec, with an additional 20 sec of fixation at the beginning and end of each run.

Stimulus presentation in the scanner. The stimulus image signal was generated by a personal computer at a frame rate of 60 Hz. 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 (Tokyo, Japan) VPL-PX21 liquid crystal display projector equipped with a special lens. The subject viewed the stimuli via a mirror mounted to the head coil at an angle of $\sim 45^\circ$. In SFM and localization experiments, the stimuli had a size of $\sim 15^\circ$ visual angle.

Experimental designs and tasks

SFM experiment. The experiment comprised nine random-dot stimulus conditions (taxonomy in Fig. 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. Because each period had a duration of 16 sec, an experimental run lasted 8 min and 48 sec. 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 before 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 nonface 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. Because none of the subjects pressed the face button in any of the nonface 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 30 sec stimulus blocks and seven 20 sec blocks of fixation (resulting in 5 min and 20 sec of measurement time per run). In each block, 45 different photos were presented foveally at a rate of one every 670 msec. 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, and two subjects underwent one run.

Subjects

Seven subjects between 21 and 34 years of age participated in the study (average age, 25.3 years). They had normal (four subjects) or corrected-to-normal (three subjects) vision. Four of them were female, and three were male. Six of them were right-handed; one was left-handed. Potential subjects received information about MRI 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 of the Academisch Ziekenhuis (university hospital) associated with the Catholic University Nijmegen (Nijmegen, The Netherlands).

Functional and anatomical MRI

Functional measurements in the SFM and localization experiments. We measured 20 transversal slices at 1.5 T (Magnetom Sonata; Siemens, Erlangen, Germany) 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, 5 mm in the SFM experiment and 4 mm in the localization experiments; gap, 0 mm; slice acquisition order, interleaved; field of view (FOV), $200 \times 200 \text{ mm}^2$; acquisition matrix, 64×64 ; time to repeat (TR), 2000 msec; time to echo (TE), 60 msec; flip angle (FA), 90° . A functional run lasted 5 min and 20 sec in the localization experiments and 8 min and 48 sec in the SFM experiment.

Functional measurements in the retinotopy mapping experiment. We measured 25 transversal slices with $3 \times 3 \times 3 \text{ mm}^3$ isotropic voxels at 3 T scanner (Magnetom Trio; Siemens) (TR, 2000 msec; TE, 35 msec; FA, 70°). One scan lasted 11 min and 20 sec, yielding 340 vol.

Anatomical measurements. Each subject underwent a high-resolution T1-weighted anatomical scan at 1.5 T (Magnetom Sonata, see above), using either a three-dimensional magnetization-prepared-rapid-acquisition-gradient-echo sequence lasting 8 min and 34 sec (192 slices; slice thickness, 1 mm; TR, 2000 msec; TE, 3.93 msec; FA, 15° ; FOV, $250 \times 250 \text{ mm}^2$; matrix, 256×256) or a three-dimensional T1-fast-low-angle shot sequence lasting 16 min and 5 sec (200 slices; slice thickness, 1 mm; TR, 30 msec; TE, 5 msec; FA, 40° ; FOV, $256 \times 256 \text{ mm}^2$; matrix, 256×256).

Statistical analysis

Preprocessing. Before 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 detected automatically and corrected by using the anatomical contrast present in functional MR images. The Levenberg–Marquardt algorithm was used to determine translation and rotation parameters (six parameters) that minimize the sum of squares of the voxel-wise intensity differences between each volume and the first volume of the run. Each volume was then resampled in three-dimensional space according to the optimal parameters using trilinear interpolation. (3) Temporal high-pass filtering was performed to remove temporal drifts of a frequency below three cycles per run (3/528 sec). (4) The functional volumes were projected into Talairach space, using the position parameters of the scanner, 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 4 mm. The BrainVoyager 2000 software package (version 4.8; R. Goebel) 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 Fig. 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 (see Fig. 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 (see Figs. 2, 4, 5), we computed t statistics at each voxel on the basis of the b weights (β 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 localized individually 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 the blood-oxygen-level dependent (BOLD) response amplitude of one condition relative to the variability of the signal. The b weights obtained for

the individually localized regions were averaged across subjects, and their SEs were adjusted appropriately (see Fig. 3). This approach is preferred over averaging effect estimates in percentage-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 (6 sec 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 15 lags. Lags ranged from 0 to 14 TRs (i.e., 0–28 sec). 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 three-dimensional object, such as a face, even when the outer boundary is eliminated completely by superimposing a circular aperture, as was the case in all of 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 the surface of an object 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 structure of the object and its motion. We will refer to this type of SFM as classical SFM (Fig. 1A).

In a series of psychophysical pilot experiments, we found that three-dimensional surface structure can also be perceived in moving-dot stimuli constructed in a radically different way (Fig. 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 surface of the object (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 network of regions (Fig. 2). The network covers a large contiguous expanse of visual cortex (*blue* activation surface in Fig. 2, *bottom*), which is strikingly symmetrical and extends from early visual areas dorsally into hMT+, the intraparietal sulcus (IPS), and other parts of the 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 precentral gyrus (PCG) and midfrontal gyrus (MFG). Subcortically, there was strong bilateral thalamic activity, probably including the lateral geniculate nucleus as well as pulvinar.

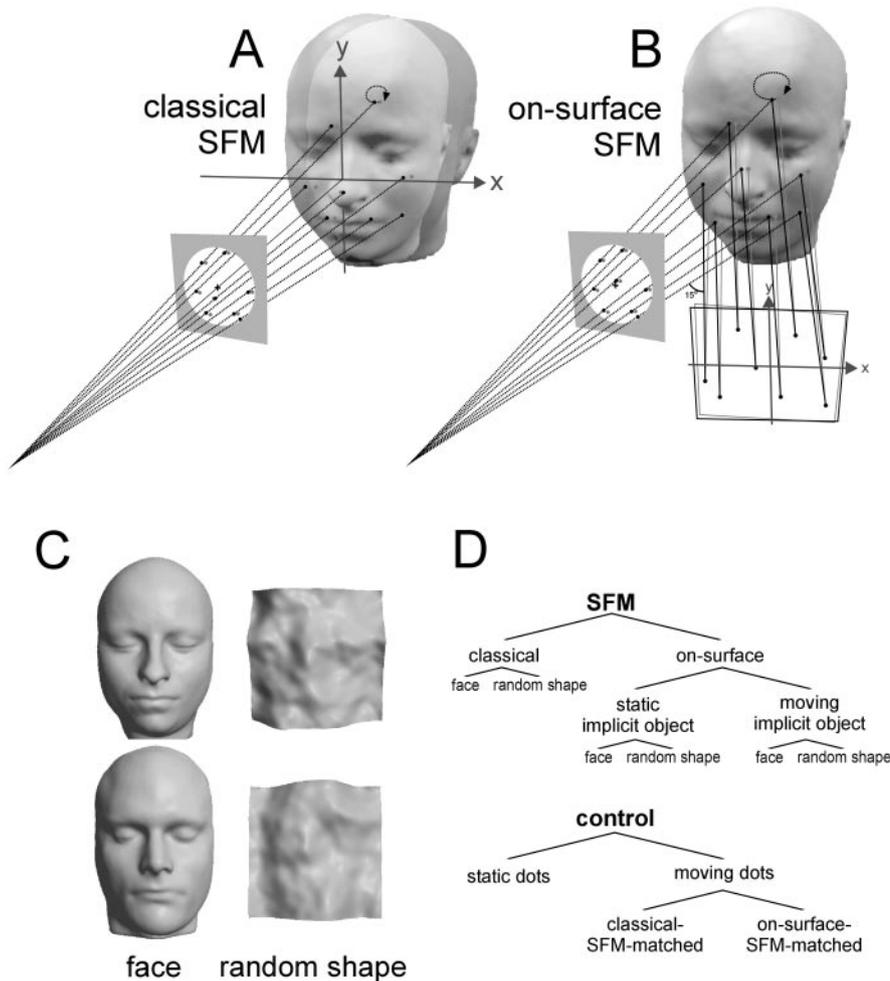


Figure 1. Construction of classical SFM and on-surface SFM stimuli. This figure contrasts the construction of the two types of SFM encoding that were used. *A*, In our version of the classical SFM stimulus, the dot locations are selected randomly on the surface of an object. These fixed locations are then polar-projected onto an image plane as the object moves in three-dimensional space. 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 surface of an object as if 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° 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 Fourier transforms of the depth maps. *D*, Taxonomy of the moving-dot conditions used in the SFM experiment (for details, see Stimuli and Experimental designs and tasks in Materials and Methods).

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 (Fig. 2). It was confirmed in subject J.S. by retinotopy mapping (Sereno et al., 1995), which allows precise determination of the boundaries of early visual areas (see Fig. 5*C,D*). Notably, the SFM network also includes motion-sensitive areas V3a (Tootell et al., 1997) and V3b, of which the latter is also known as the kinetic occipital (KO) region (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 (see Fig. 5*C,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. The response profile of V3a will be described below in the context of that of key region hMT+.

MFG and PCG

In addition, a large bilateral region in the PCG and bilateral regions of the 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 (Fig. 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. The activity of these regions was sustained for the full 16 sec of each condition period (Talairach-space group event-related average; data not shown). Although 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 of each 16 sec stimulus period. Transient activity, whose latency was consistent with the average button-press reaction time of our subjects, was found more posteriorly in PCG (data 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 the response of this bilateral PCG region was sustained suggests that it is the FEF, because the FEF is known to be active during fixation (Petit et al., 1999), 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. Because the FEF contributes to both fixation and control of dynamic eye movements, including smooth pursuit (Petit and Haxby, 1999), 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 did not differ between static and moving random-dot conditions or between static- and moving-implicit-object SFM conditions (data 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 caused by the suppression of stereotypical face-scanning patterns required by the instruction to fixate.

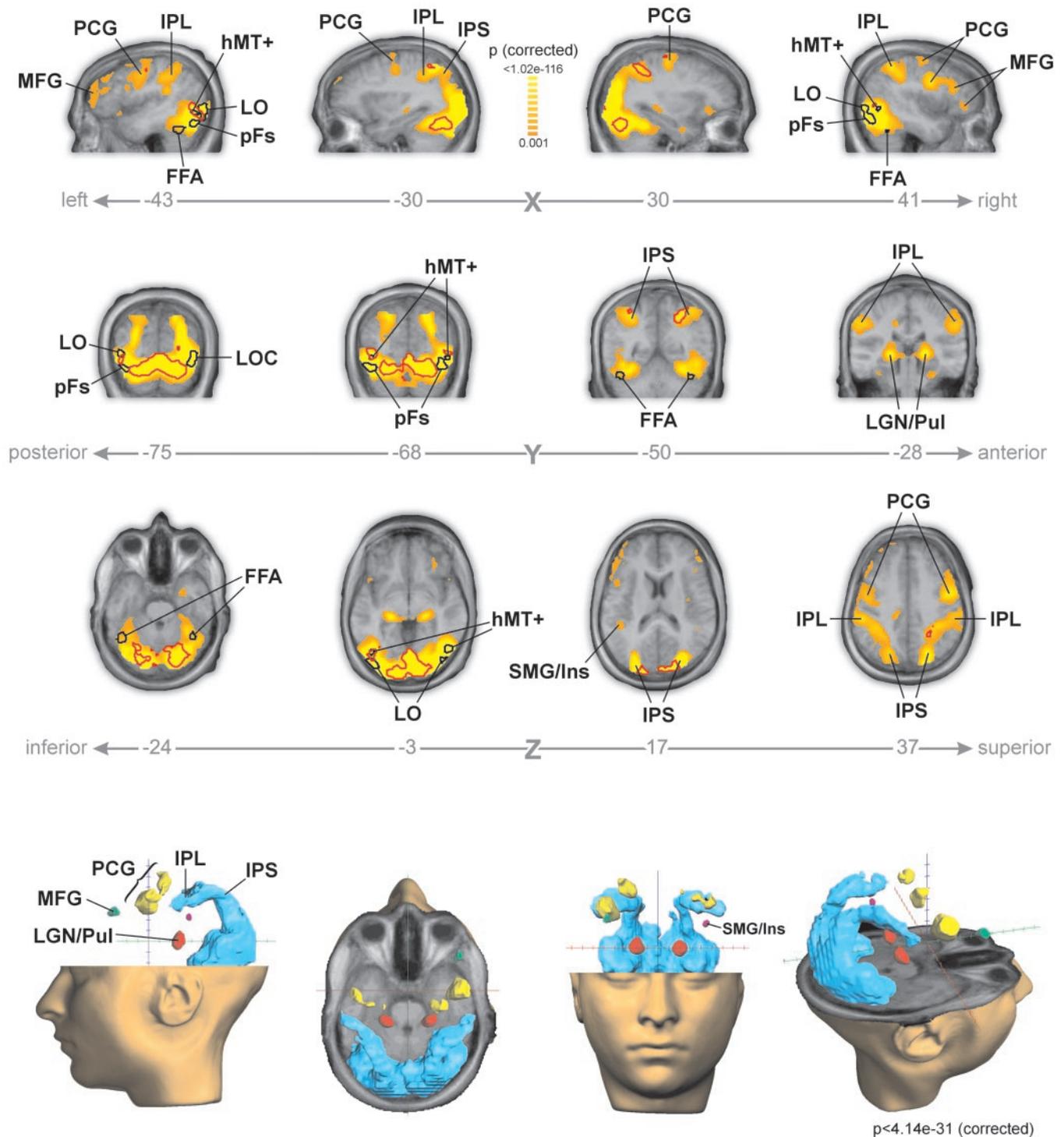


Figure 2. SFM-object-recognition network (group results). Brain regions active during SFM object recognition (Talairach-space group analysis) are shown. *Ins*, Insula; *IPL*, inferior parietal lobule; *LGN*, lateral geniculate nucleus; *Pul*, pulvinar; *SMG*, supramarginal gyrus. *Top*, Orange to yellow regions are significantly active during SFM object recognition (compared with 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 (i.e., LO and pFs), and FFA as defined by separate localizer contrasts using appropriate stimuli including photos (see Stimuli in 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; $p < 4.14 \times 10^{-31}$; 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.

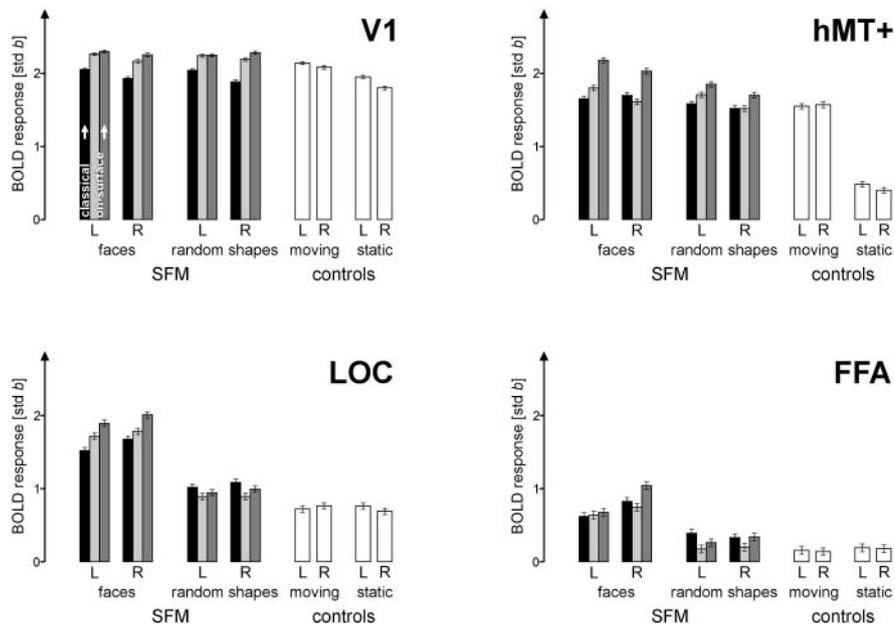


Figure 3. Key-region response profiles (group results). Responses to SFM object and control stimuli in regions of interest as reflected in the linear-regression standardized b weights (β estimates) averaged across subjects are shown. Group averaging is based not on Talairach correspondence but on individual localization of the key regions in each subject. V1 has been localized anatomically; hMT+, LOC, and FFA have been localized 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 SE 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* represent responses to on-surface SFM stimuli. *Light-gray bars* represent stationary-implicit-object on-surface SFM responses; *dark-gray bars* represent moving-implicit-object on-surface SFM responses. *White bars* represent responses to control stimuli as labeled (for details, see Statistical analysis in 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 (Fig. 2, *red outlines*, Figs. 3, 4, *red vs blue bars*). There were no regions significantly more strongly active during classical than during on-surface SFM perception.

Because 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, because 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 three-dimensional 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 Meth-

ods). Figure 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 behavior of the key regions during the SFM experiment were first analyzed for each subject individually. These individual analyses have been integrated in Figure 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, approximately 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) (Fig. 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 attributable 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, although this effect was not significant. SFM stimuli, on average, did not drive V3a more strongly than motion controls. Together, these results suggest that V3a is not particularly sensitive to the surface structure implicit to SFM stimuli.

HMT+

HMT+, as localized individually in each subject, showed an even more pronounced motion selectivity than V3a, responding very strongly to all moving-dot stimuli, whereas 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- vs light-gray bars* in Fig. 3). This effect was more pronounced for faces than for random shapes.

To assess the effect of the SFM encoding (classical vs 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) (Fig. 4, *bottom row*). 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

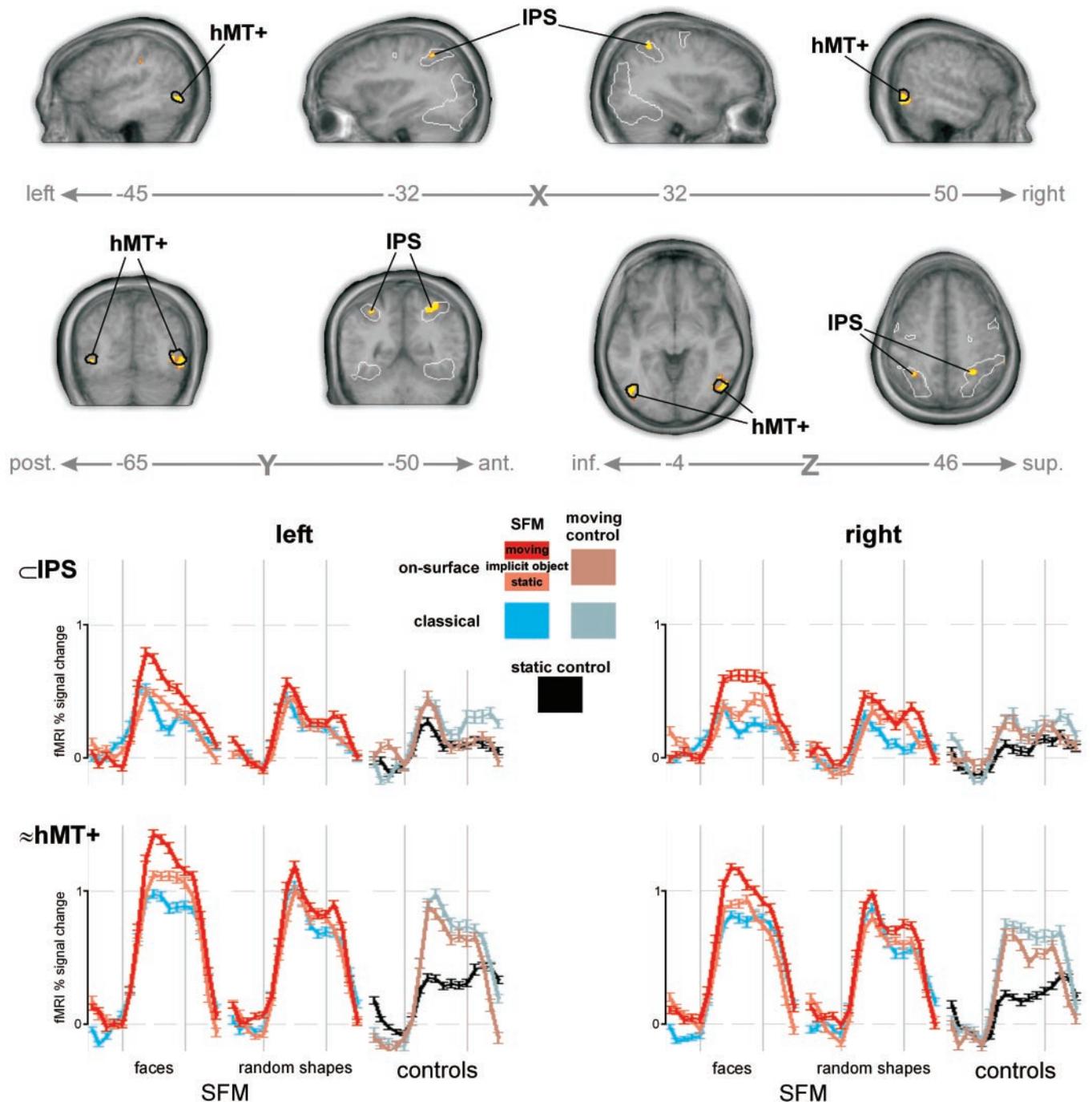


Figure 4. Regions reflecting motion of the SFM-implicit object (group analysis). In on-surface SFM, the dots move on the surface of an object, whereas the implicit object itself can be either stationary or moving. *Top*, To detect regions reflecting motion of the implicit object independent of the surface-defining 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 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 Fig. 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 is outlined in white (threshold as in the bottom row of Fig. 2). Group analysis is based on Talairach-space correspondence. *ant.*, Anterior; *post.*, posterior; *inf.*, inferior; *sup.*, superior. *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 SEM. The color coding is defined in the visual legend. For statistical details, see Statistical analysis in Materials and Methods.

with the Talairach-space group analysis (Fig. 2), the individually localized left hMT+ responded more strongly to on-surface than to classical SFM stimuli (Fig. 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 the 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.

LOC

LOC responded with approximately equal moderate activity to both moving and static random-dot control stimuli. This re-

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 ± 5	-68 ± 4	-3 ± 2	46 ± 3	-62 ± 3	0 ± 3
LOC	-40 ± 6	-71 ± 5	-11 ± 3	40 ± 3	-71 ± 3	-8 ± 6
FFA	-39 ± 4	-49 ± 4	-20 ± 3	36 ± 2	-49 ± 6	-18 ± 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).

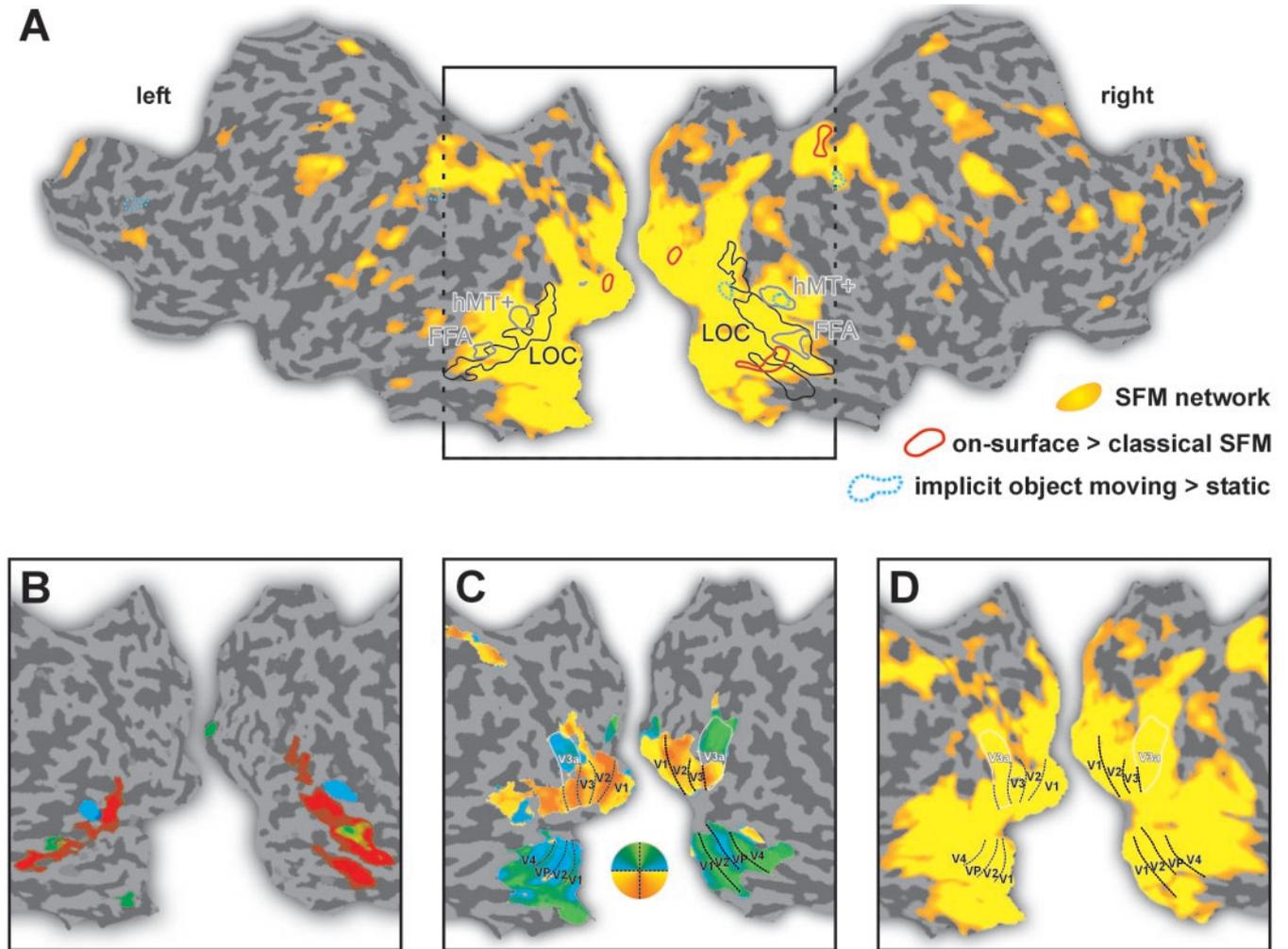


Figure 5. SFM-object-recognition network (single-subject results). Results for subject J.S. presented on flat maps of the cortical hemispheres. *Dark- and light-gray regions* approximately correspond to sulci and gyri, respectively (*dark gray* indicates concave, and *light gray* indicates convex shape of the cortical surface in its original folded state). Thresholded statistical maps are superimposed in color. *A* closely parallels Figure 2 (group results). *Orange to yellow regions* were significantly active during SFM object recognition ($p < 0.05$; corrected). For abbreviations, see Figure 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. *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). *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 *D*, the boundaries of the early visual areas 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).

sponse 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. Thus, the response of LOC to our SFM stimuli was object-selective but also (and more pronouncedly) face-selective. This is consistent

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

FFA

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 (Fig. 2, red outlines) is also reflected in the response profiles of the key regions. 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 6 (Talairach coordinates in Table 2). The peak SFM face selectivity approximately 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 cannot explain the symmetry of the shifts with respect to the

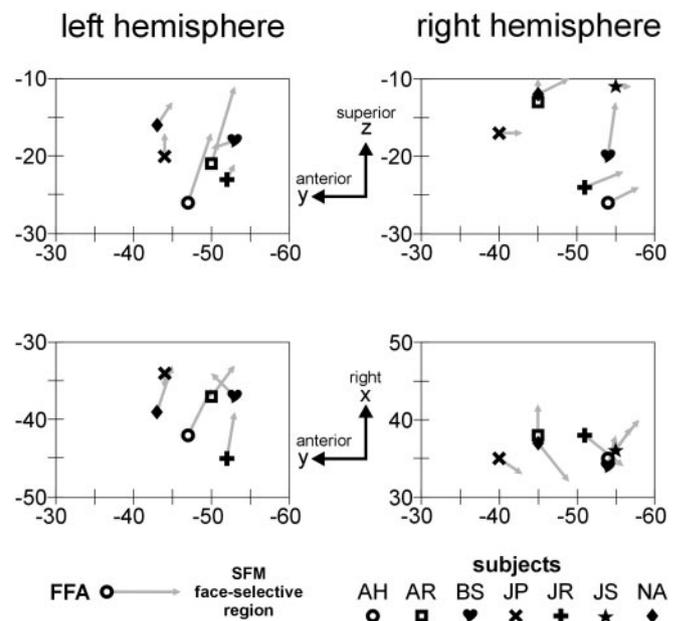


Figure 6. Talairach locations of FFA and SFM face-selective regions in each subject. FFA as localized with photo stimuli is not identical to the peak face-selective region determined with SFM stimuli. Shift vectors point from the photo face-selectivity peak (i.e., FFA) to the face-selectivity peak obtained 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.

medial plane or 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.

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	←dist→	x	y	z	rad	←dist→	x	y	z	rad	←dist→					
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)										FFA (Photo experiment)					Face-selective region (SFM experiment)					FFA (Photo experiment)				
x	y	z	rad	←dist→	x	y	z	rad	←dist→	x	y	z	rad	←dist→	x	y	z	rad	←dist→					
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	-54	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 millimeters) 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, we have specified the radius (rad, in millimeters) of a sphere of the same volume as the region of interest. (The volume of the region of interest, thus, is $v = \frac{3}{4}\pi r^3$.) The distance and volume measurements were performed individually for each subject, but in Talairach space. Although 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. N.I. (not localizable) indicates that the region could not be localized (effect absent or insignificant).

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. The significance level of the map as a whole is 5% (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). Not localizable indicates that the region could not be localized and may not be present.

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 three-dimensional 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 determine 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 (Fig. 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 (Fig. 4). This implicit-object-motion effect is markedly stronger in both the IPS region and hMT+ for faces than for random shapes (Fig. 4, *bottom*). There was no region responding more strongly to stationary than to moving implicit objects in on-surface SFM.

Discussion

SFM 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 also 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 (Tanaka, 2000) and its putative human counterpart, ventral temporal cortex (Haxby et al., 2001), contain complex object representations that are somewhat cue-invariant (Sáry 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 three-dimensional 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 rather than 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. The crucial factor may be that faces are more frequent and behaviorally relevant than random shapes in natural vision. Visual objects may be represented by a basis system of complex shape templates optimized for the representation of natural shapes, including faces. For the ventral-stream regions, this bottom-up explanation is in line with the dominant view in the literature and appears compelling.

However, a top-down mechanism with differential effects during face and random-shape conditions is also plausible, especially because the task was face–nonface categorization. Below we argue that prior knowledge about the shape of faces may be used to disambiguate surface representations in hMT+ through feedback.

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 Fig. 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. 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, because in natural vision on-surface motion (e.g., water or light–shadow contours moving across a surface) is a rather rare phenomenon, whereas classical SFM processing contributes to perception whenever there is relative motion between observer and object. 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.

Mental effort is a vague notion, merely suggesting greater computational vigor. Computing surface structure from on-surface motion, however, probably requires not just greater computational vigor but altogether different computations, which may be less efficiently organized in the visual system.

An IPS region may represent implicit-object motion

We found a region in the IPS that directly reflects implicit object motion during on-surface SFM perception (Fig. 4). Braddick et al. (2000) found a similarly located IPS region active during coherent motion stimulation. Orban et al. (1999) found IPS regions active during classical SFM perception. Although these findings are consistent with ours, they do not suggest an account of the implicit-object-motion effect. Low-level differences between the conditions are minimal and cannot explain this effect, suggesting that implicit-object motion is really the crucial factor.

A related possibility is that the IPS region serves a function that depends on implicit-object motion, for example 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 and Haxby (1999) describe an IPS region involved in the control of smooth-pursuit eye movements. They identify this region as belonging to the parietal eye field. Because 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 and Haxby (1999) is clearly removed from ours in Talairach space (14 and 23 mm for left and right subregions, respectively; these are significant distances given intersubject variability). Although 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.

HMT+: motion flowfield and surface depth map

The function most frequently attributed to hMT+ is the representation of the visual motion flowfield. This view is supported by numerous studies in monkeys (Maunsell and Van Essen, 1983a; Pack and Born, 2001) as well as humans (Tootell et al., 1995; Goebel et al., 1998; for 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, because it may be a part of it (Singh et al., 2000) and we do not have the data to distinguish the two. 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 electrophysiological results. Monkey MT cells have been shown to carry information not merely about motion direction and velocity but also about binocular-disparity (Maunsell and Van Essen, 1983b) and motion-defined surface orientation (Xiao et al., 1997). There is evidence that the activity of these cells determines the visual depth percept (DeAngelis and Newsome, 1999). Bradley et al. (1998) have shown how MT cells with near and far depth selectivity reflect the monkey's interpretation of an ambiguous SFM cylinder 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).

Together, the evidence suggests a dynamic model of SFM, 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.

This model also explains why hMT+ activity reflects implicit-object motion. 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 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.

Is the representational machinery in hMT+ specialized for the structure of natural objects (cf. Kourtzi et al., 2002)? HMT+ activity reflected the difference between faces and random shapes, but this effect was 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.

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 (Table 2, Fig. 6). This suggests that the ventral-stream representation 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. It must be noted that these conclusions are tentative, because 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 use SFM stimuli of natural objects or shape-from-shading representations of faces and random shapes (Fig. 1C) to further explore these issues.

References

- Albright TD (1984) Direction and orientation selectivity of neurons in visual area MT of the macaque. *J Neurophysiol* 52:1106–1130.
- Allison T, Ginter H, McCarthy G, Nobre AC, Puce A, Luby M, Spencer DD (1994) Face recognition in human extrastriate cortex. *J Neurophysiol* 71:821–825.
- Amedi A, Malach R, Hendler T, Peled S, Zohary E (2001) Visuo-haptic object-related activation in the ventral visual pathway. *Nat Neurosci* 4:324–330.
- Andersen RA, Bradley DC (1998) Perception of three-dimensional structure from motion. *Trends Cogn Sci* 2:222–228.
- Boynton GM, Engel SA, Glover GH, Heeger DJ (1996) Linear systems analysis of functional magnetic resonance imaging in human V1. *J Neurosci* 16:4207–4221.
- Braddick OJ, O'Brien JM, Wattam-Bell J, Atkinson J, Turner R (2000) Form and motion coherence activate independent, but not dorsal/ventral segregated, networks in the human brain. *Curr Biol* 10:731–734.

- Bradley DC, Chang GC, Andersen RA (1998) Encoding of three-dimensional structure-from-motion by primate area MT neurons. *Nature* 392:714–717.
- Corbetta M, Shulman GL, Miezin FM, Petersen SE (1995) Superior parietal cortex activation during spatial attention shifts and visual feature conjunction. *Science* 270:802–805.
- Corbetta M, Akbudak E, Conturo TE, Snyder AZ, Ollinger JM, Drury HA, Linenweber MR, Petersen SE, Raichle ME, Van Essen DC, Shulman GL (1998) A common network of functional areas for attention and eye movements. *Neuron* 21:761–773.
- Culham JC, Brandt SA, Cavanagh P, Kanwisher NG, Dale AM, Tootell RB (1998) Cortical fMRI activation produced by attentive tracking of moving targets. *J Neurophysiol* 80:2657–2670.
- Culham JC, He S, Dukelow S, Verstraten FA (2001) Visual motion and the human brain: what has neuroimaging told us? *Acta Psychol (Amst)* 107:69–94.
- DeAngelis GC, Newsome WT (1999) Organization of disparity-selective neurons in macaque area MT. *J Neurosci* 19:1398–1415.
- Goebel R, Khorram-Sefat D, Muckli L, Hacker H, Singer W (1998) The constructive nature of vision: direct evidence from functional magnetic resonance imaging studies of apparent motion and motion imagery. *Eur J Neurosci* 10:1563–1573.
- Grill-Spector K, Kushnir T, Edelman S, Avidan G, Itzhak Y, Malach R (1999) Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron* 24:187–203.
- Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293:2425–2430.
- Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci* 17:4302–4311.
- Kanwisher N, Tong F, Nakayama K (1998) The effect of face inversion on the human fusiform face area. *Cognition* 68:B1–B11.
- Kanwisher N, Stanley D, Harris A (1999) The fusiform face area is selective for faces, not animals. *NeuroReport* 10:183–187.
- Kourtzi Z, Kanwisher N (2001) Representation of perceived object shape by the human lateral occipital complex. *Science* 293:1506–1509.
- Kourtzi Z, Bühlhoff HH, Erb M, Grodd W (2002) Object-selective responses in the human motion area MT/MST. *Nat Neurosci* 5:17–18.
- Levy I, Hasson U, Avidan G, Hendler T, Malach R (2001) Center-periphery organization of human object areas. *Nat Neurosci* 4:533–539.
- Malach R, Reppas JB, Benson RR, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady TJ, Rosen BR, Tootell RB (1995) Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc Natl Acad Sci USA* 92:8135–8139.
- Malach R, Levy I, Hasson U (2002) The topography of high-order human object areas. *Trends Cogn Sci* 6:176–184.
- Maunsell JH, Van Essen DC (1983a) Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. *J Neurophysiol* 49:1127–1147.
- Maunsell JH, Van Essen DC (1983b) Functional properties of neurons in middle temporal visual area of the macaque monkey. II. Binocular interactions and sensitivity to binocular disparity. *J Neurophysiol* 49:1148–1167.
- Orban GA, Dupont P, De Bruyn B, Vogels R, Vandenberghe R, Mortelmans L (1995) A motion area in human visual cortex. *Proc Natl Acad Sci USA* 92:993–997.
- Orban GA, Sunaert S, Todd JT, Van Hecke P, Marchal G (1999) Human cortical regions involved in extracting depth from motion. *Neuron* 24:929–940.
- Pack CC, Born RT (2001) Temporal dynamics of a neural solution to the aperture problem in visual area MT of macaque brain. *Nature* 409:1040–1042.
- Paradis AL, Cornilleau-Peres V, Droulez J, Van De Moortele PF, Lobel E, Berthoz A, Le Bihan D, Poline JB (2000) Visual perception of motion and 3-D structure from motion: an fMRI study. *Cereb Cortex* 10:772–783.
- Perrett DI, Rolls ET, Caan W (1982) Visual neurones responsive to faces in the monkey temporal cortex. *Exp Brain Res* 47:329–342.
- Petit L, Haxby JV (1999) Functional anatomy of pursuit eye movements in humans as revealed by fMRI. *J Neurophysiol* 81:463–471.
- Petit L, Dubois S, Tzourio N, DeJardin S, Crivello F, Michel C, Etard O, Denise P, Roucoux A, Mazoyer B (1999) PET study of the human foveal fixation system. *Hum Brain Mapp* 8:28–43.
- Puce A, Allison T, Gore JC, McCarthy G (1995) Face-sensitive regions in human extrastriate cortex studied by functional MRI. *J Neurophysiol* 74:1192–1199.
- Sáry G, Vogels R, Orban GA (1993) Cue-invariant shape selectivity of macaque inferior temporal neurons. *Science* 260:995–997.
- Sereno ME, Trinath T, Augath M, Logothetis NK (2002) Three-dimensional shape representation in monkey cortex. *Neuron* 33:635–652.
- Sereno MI, Dale AM, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, Rosen BR, Tootell RB (1995) Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science* 268:889–893.
- Singh KD, Smith AT, Greenlee MW (2000) Spatiotemporal frequency and direction sensitivities of human visual areas measured using fMRI. *NeuroImage* 12:550–564.
- Smith AT, Greenlee MW, Singh KD, Kraemer FM, Hennig J (1998) The processing of first- and second-order motion in human visual cortex assessed by functional magnetic resonance imaging (fMRI). *J Neurosci* 18:3816–3830.
- Tanaka K (2000) Mechanisms of visual object recognition studied in monkeys. *Spat Vis* 13:147–163.
- Tootell RB, Reppas JB, Kwong KK, Malach R, Born RT, Brady TJ, Rosen BR, Belliveau JW (1995) Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *J Neurosci* 15:3215–3230.
- Tootell RB, Mendola JD, Hadjikhani NK, Ledden PJ, Liu AK, Reppas JB, Sereno MI, Dale AM (1997) Functional analysis of V3A and related areas in human visual cortex. *J Neurosci* 17:7060–7078.
- Treue S, Husain M, Andersen RA (1991) Human perception of structure from motion. *Vision Res* 31:59–75.
- Ungerleider LG, Mishkin M (1982) Two cortical visual systems. In: *Analysis of visual behavior* (Ingle DJ, ed), pp 549–586. Cambridge, MA: MIT.
- Van Oostende S, Sunaert S, Van Hecke P, Marchal G, Orban GA (1997) The kinetic occipital (KO) region in man: an fMRI study. *Cereb Cortex* 7:690–701.
- Wallach H, O'Connell DN (1953) The kinetic depth effect. *J Exp Psychol* 45:205–217.
- Xiao DK, Marcar VL, Raiguel SE, Orban GA (1997) Selectivity of macaque MT/V5 neurons for surface orientation in depth specified by motion. *Eur J Neurosci* 9:956–964.
- Zeki S, Watson JD, Lueck CJ, Friston KJ, Kennard C, Frackowiak RS (1991) A direct demonstration of functional specialization in human visual cortex. *J Neurosci* 11:641–649.