Transformation from image-based to perceptual representation of materials along the human ventral visual pathway

Chihiro Hiramatsu,1,2 Naokazu Goda, a,b,⁎, Hidehiko Komatsu a,b

a Division of Sensory and Cognitive Information, National Institute for Physiological Sciences, Okazaki 444-8585, Japan
b Department of Physiological Sciences, The Graduate University for Advanced Studies (SOKENDAI), Okazaki 444-8585, Japan

ARTICLE INFO

Article history:
Received 5 January 2011
Revised 21 April 2011
Accepted 26 April 2011
Available online xxxx

Keywords:
Material perception
fMRI
Multivoxel pattern analysis
Image statistics
Texture
Visual cortex

ABSTRACT

Every object in the world has its own surface quality that is a reflection of the material from which the object is made. We can easily identify and categorize materials (wood, metal, fabric etc.) at a glance, and this ability enables us to decide how to interact appropriately with these objects. Little is known, however, about how materials are represented in the brain, or how that representation is related to material perception or the physical properties of material surface. By combining multivoxel pattern analysis of functional magnetic resonance imaging data with perceptual and image-based physical measures of material properties, we found that the way visual information about materials is coded gradually changes from an image-based representation in early visual areas to a perceptual representation in the ventral higher-order visual areas. We suggest that meaningful information about multimodal aspects of real-world materials reside in the ventral cortex around the fusiform gyrus, where it can be utilized for categorization of materials.

© 2011 Elsevier Inc. All rights reserved.

Introduction

Identifying the materials from which an object is made based on its surface properties helps us to understand the quality or function of the object (Adelson, 2001). In our daily life, we use information about materials to decide how to interact with objects. For example, we know that shirts made from fabric are pliable enough to be squeezed and that whereas metallic spoons are hard and tough, glass plates are hard but fragile and need to be handled carefully. Perceiving materials visually also evokes feelings related to other sensory modalities and even aesthetic and affective feelings. We feel furry objects are soft, stones are heavy, and seeing a log cabin may make us feel relaxed.

Psychophysical studies have demonstrated that our ability to visually identify and categorize real-world materials such as metal, wood, fabric etc. from images is quick and seemingly effortless (Sharan, 2009). This ability, together with empirical knowledge about material properties, enables us to interact appropriately with objects (Buckingham et al., 2009). Moreover, the remarkable ease with which we identify and categorize objects implies the presence of neural circuitry that efficiently extracts information about materials.

Despite its importance, little attention has been paid to the neural mechanisms underlying material perception. Perceiving materials requires comprehensive analysis of the color, texture and light reflection/transmission properties (e.g., glossiness and translucency) of a surface. Although a number of electrophysiological studies in monkeys and functional magnetic resonance imaging (fMRI) studies in humans have focused on these individual aspects of surface properties (Arcizet et al., 2008; Bartels and Zeki, 2000; Dojat et al., 2006; Edwards et al., 2003; Hanazawa and Komatsu, 2001; Komatsu, 1998; Koteles et al., 2008; Newman et al., 2005; Peuskens et al., 2004; Stilla and Sathian, 2008), how information about these properties is integrated in the brain for identification and categorization of materials is largely unknown. Recently, a series of fMRI studies by Cant and Goodale (2007), Cant et al. (2005) and by Cavina-Pratesi et al. (2010a,b) identified multiple cortical regions selective for individual features of surface properties (color and texture) around the collateral sulcus and the fusiform gyrus, in the medial portion of the occipitotemporal cortex. They suggested that a constellation of these regions is involved in inferring surface materials (Cavina-Pratesi et al., 2010b), though how information about such a wide variety of materials is encoded in these regions, and how it is related to material perception, have not yet been evaluated.

Our aim in the present study was to reveal how information about materials is encoded and transformed into perceptual representations in the brain. To this end, we used pattern-information analysis to investigate the representational content of a region by assessing the information carried in that region’s pattern of activity (Edelman et al., 1998; Kriegeskorte et al., 2008; Mur et al., 2009). A number of studies have demonstrated that various basic categories of objects (e.g., Cox...
Savoy, 2003; Haxby et al., 2001), scenes (Peelen et al., 2009; Walther et al., 2009), facial expressions (Said et al., 2010), vocal emotions (Ethofer et al., 2009), odors (Howard et al., 2009) etc., as well as simple visual features like orientation (e.g., Haynes and Rees, 2005; Kamitani and Tong, 2005) and color (Brouwer and Heeger, 2009; Parkes et al., 2009), can be decoded from the pattern of brain activity. Moreover, by analyzing the patterns of activity in various brain regions, recent studies have been able to show that higher-order visual areas represent objects categorically and hierarchically (Kriegeskorte et al., 2008) and represent objects within categories or with particular shapes in perceptually relevant ways (Hausofer et al., 2008; Op de Beeck et al., 2008, 2010; Weber et al., 2009). We therefore believed that this type of analysis had the potential to read out categorical representations of materials in the brain.

We conducted fMRI experiments using images of materials from 9 basic real-world categories (metal, ceramic, glass, stone, bark, wood, leather, fabric, and fur). Each material category was represented by eight different realistic, synthesized exemplar images with controlled 3D shapes (Fig. 1, Supplementary Fig. 1), and we analyzed the patterns of fMRI activity while subjects viewed these images. We first explored whether material categories could be decoded from the pattern of local brain activity in specific cortical regions. If this were the case, we would next investigate whether the way material information was encoded was related to image-based or perceptual representations. To examine that, we compared neural dissimilarities, image-based dissimilarities and perceptual dissimilarities across material categories. Neural dissimilarities were measured based on differences in the patterns of brain activity, image-based dissimilarities were based on differences in image statistics, and perceptual dissimilarities were based on differences in perceptual impressions characterized using the semantic differential (SD) method (Osgood et al., 1957). We computed the correlations between neural dissimilarities and image-based or perceptual dissimilarities, and then evaluated what kind of information was represented in various cortical regions.

We found that material information is represented differently across different visual areas and that there is a gradual shift from image-based to perceptual representation along the ventral visual pathway. The early retinotopic areas mainly represent the image properties of the materials, while the higher areas around the fusiform gyrus represent perceptual high-level categories.

Materials and methods

Subjects

Five healthy Japanese subjects (2 males and 3 females, age 21–33 years) who were naive to the purpose of this study participated in experiments after providing written informed consent. The study was approved by the ethics committee of National Institute for Physiological Sciences. All subjects were right-handed and had normal or corrected-to-normal vision.

Visual stimuli

The virtual 3D images classified into 9 material categories (metal, ceramic, glass, stone, bark, wood, leather, fabric and fur) were created using NewTek LightWave 3D graphic software. Each category contained 8 exemplars with different surfaces and slightly different nonsense shapes (Fig. 1, Supplementary Fig. 1). The different surfaces were rendered by varying the surface reflectance parameters and mapping synthesized textures or photographs of real materials that were taken by the authors or obtained from a royalty-free texture book (ASCII Media Works, Tokyo). All images were rendered as 600×600 pixels with a light gray surround (194 cd/m²).

The images were centered on a uniformly gray background (145 cd/m², 1280×1024 pixels) and displayed with a Victor DLA-M200L projector projecting onto a half-transparent screen in a MRI scanner. Subjects viewed the screen via a mirror. Neurobehavioral Systems Presentation software was used for the stimulus presentation. The display system was calibrated by measuring the red, green and blue spectral power distributions using a PhotoResearch PR-650 SpectraScan spectrophotometer.

Image analysis

We measured 20 low-level image statistics to assess image properties of the material images used in our experiment. The center region of each material image (256×256 pixels), within the object contour, was extracted and downsampled to 128×128 pixels. A steerable wavelet pyramid transform (Heeger and Bergen, 1995; Portilla and Simoncelli, 2000) was then applied to the image luminance to decompose the image into 12 subband images (3 scales, 4 orientations). Twelve subband statistics (log of mean magnitude of each subband) were derived from the subbands. Additionally, 8 moments of the CIELAB coordinates (mean and standard deviation of L*, a* and b* coordinate, and skew and kurtosis of L*) were derived from pixel color histograms. These 20 low-level image statistics were obtained for all images and z-scored separately for each statistic. To visualize the relationship of the 72 images in terms of these image statistics in a low-dimensional space, we computed Euclidean distances between images in the 20 dimensional space representing the 20 z-scored low-level image statistics and applied a classical multidimensional scaling (MDS) to the Euclidean distances between images.

---

Please cite this article as: Hiramatsu, C., et al., Transformation from image-based to perceptual representation of materials along the human ventral visual pathway, NeuroImage (2011), doi:10.1016/j.neuroimage.2011.04.056
Psychological experiments

We used the SD method (Osgood et al., 1957) to extract the subjects’ impressions when they viewed the materials. We first collected visual and non-visual adjectives according to previous studies (Hollins et al., 1993; Rao and Lohse, 1996; Picard et al., 2003) and free description of impressions towards our stimuli. We then selected 12 adjective pairs, which seem to be suited to characterize 9 material categories used in the present study; matte–glossy, opaque–transparent, simple–complex, regular–irregular, colorful–colorless, smooth–rough, dry–wet, cold–warm, soft–hard, light–heavy, elastic–inelastic, and natural–artificial. All adjectives were given in Japanese. The images were centrally presented (7.5° × 7.5°), for 500 ms), and the subjects were asked to verbally rate each image using each adjective pair on a 5-level scale. The ratings for different adjective pairs were conducted in separate blocks, in each of which the subject rated all images presented in random order using one adjective pair (once each). All ratings were performed in the fMRI scanner over 2–3 sessions conducted after the fMRI experiment. The mean rating for each adjective pair across the 5 subjects was then calculated. To assess the perceptual relationships of the 72 images, we applied a classical MDS to the Euclidean distances between images in the 12 dimensional space representing 12 mean ratings. A dendrogram was also constructed from a hierarchical cluster analysis of the Euclidean distances between images (Ward method).

In a separate experiment, we tested whether the subjects could classify the material images into 9 categories. The images were centrally presented (7.5° × 7.5°, for 500 ms), and the subjects were asked to verbally select a category of the material image from 9 listed categories (3 times each). No feedback was given. This experiment was conducted before the fMRI experiment.

Main fMRI experiment

Five subjects participated in the fMRI experiment to measure brain activity while passively viewing 9 material categories. Each run consisted of 9 category blocks (16 s each) interleaved with 16-s fixation-only blocks. In each category block, 8 exemplar images were centrally presented (7.5° × 7.5°, for 500 ms with 500 ms intervals) in random order (twice each). A uniform gray image was presented when the material images were not presented. To control fixation and attention of the subjects, subjects were required to gaze at a central fixation spot (0.13° × 0.13°) throughout a run and indicate when the color changed briefly to green (2 times/block on average) by pressing a button with their right hand. Each subject performed 14 runs over 2–3 sessions.

Localizer fMRI experiments

Subjects also participated in separate conventional retinotopic mapping and localizer experiments. The retinotopic mapping consisted of 4 runs of eccentricity mapping using an expanding/contracting checkerboard ring and 4 runs of polar mapping using clockwise/counter-clockwise rotating checkerboard wedges (32 s/cycle, 8.5 cycles). The localizer experiment consisted of 8–10 runs, in which chromatic natural images of faces, scenes, objects, grid-scrambled objects and their achromatic versions were presented in separate 16 s-blocks (7.5° × 7.5°, each image for 500 ms with 500 ms intervals) interleaved with 16-s fixation-only blocks. In both of these experiments, the subjects performed the same task used in the main fMRI experiment.

MRI acquisition and preprocessing

Functional images (ascending T2*-weighted gradient-echo echo-planar sequence; TR, 2000 ms; TE, 30 ms, flip angle, 75°; FOV, 192 × 192 mm; voxel size, 3 × 3 × 3 mm; slice gap, 0.3 mm; 35 slices) and anatomical images (magnetization-prepared rapid-acquisition gradient-echo (MPRAGE); TR, 1500 ms; TE, 4.38 ms; flip angle, 8°; voxel size, 0.75 × 0.75 × 3.0 mm) were acquired using a Siemens Allegra 3T scanner. The images covered the entire occipital lobe and parts of the temporal, parietal and frontal lobes. Oblique scanning was used to exclude the eyeballs from the images. All functional images were motion-corrected, registered with anatomical images, and spatially normalized to Montreal Neurological Institute (MNI) space (3-mm isotropic voxels) using SPM2 (http://www.fil.ion.ucl.ac.uk/spm). With the exception of the retinotopic mapping data, the images were then spatially smoothed using a 4-mm full-width at half-maximum (FWHM) Gaussian kernel.

Whole-head anatomical images (MPRAGE; TR, 2500 ms; TE, 4.38 ms; flip angle, 8°; voxel size, 0.9 × 0.9 × 1.0 mm) were also acquired. The images were normalized to MNI space (1-mm isotropic voxels) using SPM2, and nonuniformity was corrected using MIPAV (http://mipav.cit.nih.gov), after which the images were employed to reconstruct cortical surfaces using CARET (http://www.nitrc.org/projects/caret/). Thereafter, an anatomical mask representing the gray matter voxels in the whole visual cortex (visual cortex mask) was created by selecting voxels lying within 1.5 mm of the occipital, temporal and parietal lobes of the cortical surface of each subject. The functional voxels within this visual cortex mask were used for subsequent analysis.

Multivoxel pattern analysis (MVPA)

A Princeton MVPA toolbox (http://www.pni.princeton.edu/mvpa/) was used to perform MVPA. To remove drifting signals and improve the signal-to-noise ratio in each run, functional images preprocessed as above were temporally z-scored, quadratically detrended and temporally smoothed (3-point moving average). Six volumes were then extracted 8–20 s after onset for each category in each run as the patterns of activity for the respective categories.

Searchlight analysis

A spherical searchlight analysis (Kriegeskorte et al., 2006) was carried out to examine how accurately material categories were decoded from local patterns of activity in various regions using linear support vector machine (SVM). For each voxel in the visual cortex mask, the accuracy of the 9-category classification was computed based on pattern of activity within a sphere (radius, 12-mm) centered at that voxel. The SVM was trained to classify the pattern of activity into 9 categories using data from 12 runs and tested on 1 of the remaining 2 runs to obtain the accuracy of the 9-category classification (LIBLINEAR; http://www.csie.ntu.edu.tw/~cjlin/liblinear/; Crammer-Singer multiclass classification method, chance level 1/9). This cross-validation procedure was iterated 14 times while changing the training and test runs. In each iteration process, 1 run was not used in either the training or testing steps and kept for subsequent analysis to avoid the “peeking” problem (Pereira et al., 2009). The accuracy of the 9-category classification was averaged across the 14-fold cross-validations and assigned to the center voxel of the sphere to obtain a voxel map of the 9-category classification accuracy. This map of the classification accuracy was obtained for each subject and spatially smoothed (8-mm FWHM). A voxel-wise random-effect group analysis (one-tailed t test) was then performed using SPM2 to assess whether the mean accuracy across subjects was significantly above the chance level (1/9) for each voxel. The t-value obtained for each voxel was thresholded at t < 0.01 and corrected for multiple comparisons at the cluster-level (p < 0.05, minimum cluster size 80 voxels). This group analysis was constrained on the voxels within the union of the visual cortex masks for the 5 subjects. The group result was projected onto the cortical surface of a representative subject using CARET (average

Please cite this article as: Hiramatsu, C., et al., Transformation from image-based to perceptual representation of materials along the human ventral visual pathway, NeuroImage (2011), doi:10.1016/j.neuroimage.2011.04.056
volumetric data. An “average fiducial mapping” method (Van Essen, 2005) was used with the average cortical surfaces from the 5 subjects to improve registration between the group results and the cortical surface.

Region of interest (ROI) definition

We defined 5 bilateral ROIs on the individual cortical surfaces: V1/V2, V3/V4, ventral higher-order area FG/CoS, lateral higher-order area LOs/PIfS, and dorsal higher-order area V3AB/IPS. Retinotopic areas, V1, V2, V3, V4, V3A, and V3B, were delineated based on retinotopy obtained from the retinotopic mapping experiment (Wandell et al., 2007), and neighboring areas were combined because of the relatively small size of each area for the subsequent analyses. The retinotopy (eccentricity and polar angle) was derived from the phase of the significant stimulus frequency component of the voxel time-series (Fourier F test, p < 0.001 uncorrected). V4 was defined as a region with continuous upper and lower visual field representation in the ventral visual cortex (hV4; Wandell et al., 2007). FG/CoS was defined as the region around the fusiform gyrus (FG), anterior to V3/V4, extending into the collateral sulcus (CoS) medially and the occipitotemporal sulcus (OTS) laterally. LOs/PIfS was a region anterior to V3/V4 and lateral to V3A/V3B (V3AB), covering the lateral occipital sulcus (LOS) and posterior inferotemporal sulcus (pITS). IPS was a region in the intraparietal sulcus. FC, CoS, OTS, LOS, pITS and IPS were defined anatomically based on the sulcal probability map in the CARET PALS atlas that was registered with individual cortical surface (spherical morphing method; Van Essen, 2005). Each ROI consisted of voxels not overlapping those of other ROIs. To illustrate the approximate borders of the early visual areas in the 5 subjects, we created a visual area probability map for the 5 subjects on the atlas surface by assigning each surface node to the most likely area.

In each of the ROIs, the same numbers of voxels were selected for each subject based on visual responsiveness determined by voxel-wise general linear model (GLM) analysis on data from the last 2 of the 14 runs using SPM2. The visual responsiveness was defined as the t-value obtained by contrasting all categories vs. fixation-only baseline activity. The voxels were regarded as visually-responsive if the t-values were greater than 0.0. Unless otherwise stated, we selected 500 most visually-responsive voxels (e.g., 500 voxels with highest t-values) in each ROI (nearly maximal number of the voxels attainable in some ROIs/subjects). We also examined the effect of the number of voxels per ROI by nearly maximal number of the voxels attainable in some ROIs/subjects (e.g., 500 voxels with highest t-values) in each ROI (nearly maximal number of the voxels attainable in some ROIs/subjects). We also examined the effect of the number of voxels per ROI by changing it from 100 to 1000 in each subject. Voxel values were also selected in each ROI or across the whole visual cortex mask based on color, object, face or place selectivity determined based on the GLM analysis on the localizer experiment data. These selectivities (t-values) were derived for each subject by contrasting the chromatic vs. achromatic, object vs. scrambled object, face vs. object and place, place vs. object and face, respectively. The voxels were regarded as selective only if the t-values were greater than 2.33 (p < 0.01, uncorrected).

ROI analysis: 9-category classification

We computed accuracies of the 9-category classification based on the patterns of activity in the ROIs. The classification accuracy was derived using linear SVM based on the data from the first 12 runs (11 runs for training and 1 run for testing), which were not used for the voxel selection (see above). The 12-fold cross-validation gave the 9-category classification accuracy for each subject, and the mean accuracy across subjects was calculated. One-tailed t-tests were used to compare the mean accuracies across subjects to chance (1/9) for each of ROIs/conditions, and one-way ANOVA and post-hoc Tukey-Kramer tests were used to compare the mean accuracies across subjects among ROIs.

We also computed the 9-category classification accuracy using the voxels that showed the highest accuracy in the searchlight analysis. In this case, SVM was trained for 12 runs and tested on 1 run that was not used during the searchlight analysis (see searchlight analysis). This cross-validation procedure was iterated 14 times while changing the training and test runs (14-fold cross-validation). One-tailed t tests were used to compare the mean accuracies across subjects to chance (1/9) as above.

Additionally, the 9-category classification accuracy was obtained by combining the voxels across subjects for each of the 12 runs before classification. This method has been shown to improve decoding accuracy (Brouwer and Heeger, 2009). In this case, the significance of the accuracy was assessed by comparing the accuracy with a null distribution of accuracy obtained using data with randomly shuffled category labels (1000 times, one-tailed random permutation test).

ROI analysis: dissimilarities between categories

We computed neural dissimilarities between category-related patterns of activity in each ROI and compared them with image-based and perceptual dissimilarity between categories. Neural dissimilarity was quantified using two methods, which were complementary to each other. One was the Euclidean distance between mean patterns of activities across all volumes elicited by each category (Edelman et al., 1998; Mur et al., 2009). Another was the pairwise category classification accuracy (Said et al., 2010; Weber et al., 2009) obtained using linear SVM. We call these two measures “Euclidean-distance-based neural dissimilarity” (END) and “classification-based neural dissimilarity” (CND), respectively. END weighs all voxels equally, whereas CND weights informative voxels more and has greater power to detect differences between categories. In both cases, the voxels in each ROI were selected as in the 9-category classification, and the dissimilarity was computed using data from the first 12 runs only (see above). The mean neural dissimilarity across subjects was obtained and displayed in percentiles as a matrix (Kriegeskorte et al., 2008). The image-based and perceptual dissimilarities between categories were calculated from Euclidean distances between centroids of each category (mean across eight exemplars) in the multivariate spaces of 20 low-level image statistics and 12 SD ratings, respectively.

The relationships between the neural, image-based and perceptual dissimilarities were assessed using Pearson simple and partial correlation coefficients and visualized in 3D space using non-metric MDS (Kruskal’s normalized stress criterion). We assessed the high-level categorical structure of the material representation using hierarchical cluster analysis of the dissimilarities between categories (Ward method) and superordinate category index (SCI). The 9 categories were grouped into 3 superordinate categories (SCs) based on the dendrogram of perceptual dissimilarity. SCIs were defined to quantify these superordinate categorical structures as follows:

\[ SCI = \frac{mBSCD - mWSCD}{sdAD} \]

where \( mBSCD \) indicates mean pairwise dissimilarity between SCs, \( mWSCD \) indicates mean pairwise dissimilarity within the same SC, and \( sdAD \) indicates the standard deviation of all pairwise dissimilarities. The significance of correlation coefficient or SCI for each of ROIs was determined using a random permutation test. A null distribution of correlation coefficient or SCI was generated in each ROI by computing correlation coefficients or SCI using dissimilarity data in which labels of category-pairs were randomly shuffled (10000 times). The significance was determined by comparing the actual correlation coefficient or SCI with the null distribution obtained from the shuffled data. The significance of the difference in correlation coefficients between ROIs was also determined using the permutation-based method, in which the actual correlation difference was compared with the null distribution of the correlation difference obtained using the shuffled data. All results were considered significant at \( p < 0.05 \).

Please cite this article as: Hiramatsu, C., et al., Transformation from image-based to perceptual representation of materials along the human ventral visual pathway, NeuroImage (2011), doi:10.1016/j.neuroimage.2011.04.056
Results

Image and perceptual property of materials

We first examined the image and perceptual properties of 72 material images comprising the 9 basic real-world categories (metal, ceramic, glass, stone, bark, wood, leather, fabric and fur; Fig. 1, Supplementary Fig. 1). The image properties were assessed by analyzing 20 low-level image statistics, including a set of mean magnitudes of wavelet subbands (log mean magnitudes of three spatial frequency × four orientations bands) and CIELAB histograms (mean and standard deviation of L*, a*, and b* coordinates, and skew and kurtosis of L*) that are frequently used for modeling V1 activity (Kay et al., 2008; Naselaris et al., 2009) and for synthesizing stochastic natural textures in computational studies (Heeger and Bergen, 1995; Portilla and Simoncelli, 2000). Perceptual properties were assessed in a psychological experiment using the SD method, in which subjects were asked to rate each image using 12 bipolar adjective pairs; some of the adjective pairs visually characterized the materials (e.g., matte–glossy), while others were related to non-visual perception (e.g., soft–hard) or an abstract concept (e.g., natural–artificial).

Fig. 2 shows the scatter plots for the material images in image-feature space (Fig. 2A) and perceptual space (Fig. 2B), which were derived by applying classical MDS analyses to the 20 low-level image statistics and the mean ratings from the 5 subjects obtained using the SD method, respectively. The images that belonged to the same basic categories were widely distributed, though they overlapped, in image-feature space, indicating that images from different categories share low-level image features. By contrast, images from the same category clearly clustered in perceptual space. A hierarchical cluster analysis of the SD ratings indicated that each of the 9 basic categories clustered nearly perfectly (Fig. 2C). We also confirmed that subjects could accurately classify the images into the 9 categories in a separate behavioral experiment conducted before the fMRI experiment (mean accuracy across subjects = 0.43 for glass, 0.83–0.97 for others; mean across 9 categories = 0.84). Taken together, these findings indicate that the subjects categorically perceived the material images.

Notably, 3 large perception clusters emerged from among the 9 basic categories (Fig. 2B, C). We call these large clusters “superordinate categories” referring to the categorical hierarchy in object recognition (Mervis and Rosch, 1981). The first superordinate category (SC1) was composed of metal, ceramic and glass (Fig. 2B, C, blue); in the SD method, adjectives such as simple, smooth, hard, cold, artificial and glossy were given high ratings in this category. The second superordinate category (SC2) was composed of stone, bark and wood (Fig. 2B, C, green); adjectives such as complex, rough, irregular and natural were given high ratings in this category. The third superordinate category (SC3) was composed of leather, fabric and fur (Fig. 2B, C, red); adjectives such as soft, elastic, warm and light were given high ratings in this category. At the top of the dendrogram (Fig. 2C), the SC1 is first separated from the other two superordinate categories, indicating that the SC2 and SC3 are more closely related to each other than to the SC1.
Distribution of material information in the visual cortex

We measured brain activity while subjects viewed images from the 9 material categories. Each subject performed 14 runs in which 8 exemplar images from each category were presented in separate blocks, interleaved with fixation-only blocks. The subjects were asked to gaze at a central fixation spot throughout a run and respond when the color of the fixation spot occasionally changed.

If a brain region carried information about materials and this information is spatially clustered, it would be possible to decode the material categories from the patterns of fMRI activity in that region elicited by the various materials. Based on this idea, we first examined how information about materials is distributed in the visual cortex by measuring how accurately the patterns of activity in various regions classify the 9 categories. Specifically, we conducted a spherical searchlight analysis (Kriegeskorte et al., 2006), in which the accuracy of the 9-category classification was computed using linear SVM applied to local patterns of activity within spheres (radius, 12 mm) centered in various regions (see Materials and methods). Group analysis of the accuracy maps from individual subjects revealed that several brain regions yielded accuracies significantly above chance (one-tailed t test, \( p < 0.05 \), corrected for multiple comparisons at the cluster-level, minimum cluster size 80 voxels, chance level = 1/9; Fig. 3A). These regions were distributed in early retinotopic regions (V1, V2, V3 and V4) and around the lateral occipital sulcus in both hemispheres, and extended into areas around the fusiform gyrus, posterior collateral sulcus and anterior lingual gyrus in the right ventral/medial occipital cortex. Similar results were also obtained when we used a non-parametric statistical test based on permutation method instead of t test (see Supplementary methods and Supplementary Fig.2). The results of searchlight analysis suggest that visual information about material categories is scattered across large regions, including both the lower and higher visual areas.

Classification accuracy of material categories in various brain regions

We next asked whether there are differences in the classification accuracy among the various cortical regions. We divided the visual cortex into 5 ROIs, V1/V2, V3/V4, ventral higher-order area FG/CoS, lateral higher-order area LOS/pITS and dorsal higher-order area V3AB/IPS, based on retinotopy and anatomical landmarks (see Materials and methods). Neighboring areas were combined for ROI analysis to obtain voxel numbers large enough for subsequent analyses. We selected the same numbers of voxels from each ROI in

---

Fig. 3. Nine-category classification in various cortical regions. (A) Regions with accuracies significantly above the chance level (1/9) for the 5 subjects in a searchlight analysis (one-tailed t test, \( p < 0.05 \), corrected for multiple comparisons at the cluster-level). The t-values were mapped onto flattened cortical surfaces of a representative subject using a pseudocolor scale. Insets show the ventral view of inflated cortical surfaces. CoS: collateral sulcus, FG: fusiform gyrus, IPS: intraparietal sulcus, LOS: lateral occipital sulcus, OTS: occipitotemporal sulcus, L: lateral, P: posterior. (B) Accuracy of the 9-category classification in each ROI obtained with the 500 most visually-responsive voxels from each subject. The right most bars (“whole visual cortex”) indicated accuracies obtained with the 500 voxels that yielded the highest accuracy per subject in individual searchlight analyses. Black bars show the mean accuracies for the 5 subjects and white bars show accuracies obtained by combining voxels across the 5 subjects (a total of 2500 voxels) before classification. Error bars indicate±s.e.m. The chance level (1/9) is indicated by a horizontal dotted line. * \( p < 0.05 \), ** \( p < 0.001 \) (one-tailed t test for black bars; one-tailed permutation test for white bars). Same pattern of results was obtained for black bars using permutation tests (\( p < 0.05 \) except for V3AB/IPS, see Supplementary methods). (C) Distributions of the 500 most visually-responsive voxels used in the ROI analysis. The color scale denotes the number of overlaps across subjects. For clarity, only voxels overlapping across at least 2 subjects are shown. In (A) and (C), white dotted lines indicate approximate borders of the early visual areas (V1, V2, V3, V4, and V3AB), which were drawn based on a visual area probability map for the 5 subjects.
each subject based on their visual responsiveness measured using data from the last 2 runs and calculated the accuracy of the 9-category classification using data from the first 12 runs. When 500 most visually-responsive voxels per ROI per subject were selected (Fig. 3C; nearly maximal number of visually-responsive voxels available in V1/V2 and V3/V4), the mean classification accuracies across subjects (Fig. 3B, black bars) were significantly above chance (one-tailed t test, \( p \leq 0.0099 \)) in all areas except V3AB/IPS (\( p = 0.12 \)). Accuracies tended to be higher in early retinotopic areas, and this tendency remained stable when the number of voxels per ROI per subject was varied (Supplementary Fig. 3A). The accuracies were significantly different among 5 ROIs for 100–400 voxels/ROI in separate one-way ANOVA for each voxel number (\( F(4,20) \geq 2.9, p \leq 0.048 \) for 100–400 voxels/ROI, \( F(4,20) = 2.8, p = 0.053 \) for 500 voxels/ROI), and V3AB/IPS (for 100–500 voxels/ROI) and LOS/pITs (for 200 voxels/ROI) were significantly different from V1/V2 in post hoc Tukey–Kramer tests (\( p < 0.05 \); Supplementary Fig. 3A).

The classification accuracy was improved by gathering the most informative voxels from across the whole visual cortex (Fig. 3B rightmost bar). These voxels were selected based on the results of the searchlight analysis (500 voxels that yielded the highest accuracies in the individual searchlight analyses). This is consistent with the idea that the information distinguishing material categories is scattered across large regions in the visual cortex. We further tested whether the classification accuracy was improved by combining voxels across the 5 subjects before classification (Brouwer and Heeger, 2009). The classification accuracy obtained with this method was greatly improved (Fig. 3B, white bars), and the accuracies were significantly above chance, even in the V3AB/IPS (one-tailed permutation test, \( p < 0.001 \)). The classification accuracy exceeded 0.45 if most informative voxels were combined across subjects before classification. This accuracy level was, however, still relatively low compared with perceptual classification accuracy (mean accuracy = 0.84). We confirmed that this was not because SVM misclassified a specific category routinely: accuracies of 8-category classification computed after excluding one category in each ROI did not change depending on the excluded category (one-way ANOVA for each ROI, \( F(8,36) \leq 0.35, p \geq 0.94 \)). The reason for relatively low accuracy might be because the signal-to-noise ratio of the measurements was not high enough or because selectivity to materials was distributed largely uniformly, yielding relatively small differences in spatial pattern of activities across different materials.

Classification accuracy of material categories by color/object/face/place-selective voxels

It has been well documented that the extrastriate cortex contains subregions encoding specific visual stimulus features/categories such as color, object, face and place. We therefore examined whether color-, object-, face- and place-selective voxels, determined in a separate localization experiment, contributed differently to classification of material categories (Supplementary Fig. 4). Classification accuracies for the 9 material categories based on the 300 most color-, object-, face- or place-selective voxels (nearly maximal number attainable for color-selective voxels in some subjects) across the whole visual cortex of each subject were significantly above chance in all features/categories (one-tailed t test, \( p \leq 0.027 \); Supplementary Fig. 4A, B). No significant difference was observed among features/categories (one-way ANOVA, \( F(3,16) = 1.8, p = 0.20 \); see also Supplementary Fig. 3B for different voxel number). On the other hand, classification accuracy did not significantly change, even when color-, object-, face- or place-selective voxels were each separately excluded from each ROI (one-way ANOVA for each ROI, \( F(4,20) \leq 0.27, p \geq 0.89 \); Supplementary Fig. 4C). This means that the activities of voxels selective for other visual stimulus features/categories, such as color, object, face or place, are not particularly important for classification of material categories; instead, information distinguishing material categories is distributed across voxels that are selective or nonselective for these features/categories.

Representation of materials in various brain regions

The results described so far indicate that material information is distributed across a wide region, extending from lower to higher visual areas. Two important questions are: how is material information encoded in these regions and how is it related to material perception? To address these questions, we examined the relationship between neural dissimilarities and image-based or perceptual dissimilarities across material categories. We initially calculated the mean activity pattern for each category and then computed the Euclidean distance between each category pair. This was regarded as the neural dissimilarity between the category pairs (Edelman et al., 1998; Mur et al., 2009), and we call this measure of neural dissimilarity “Euclidean-distance-based neural dissimilarity” (END). The neural dissimilarity was calculated for each subject with data from the first 12 runs using the 500 most visually-responsive voxels per ROI per subject and averaged across subjects. Image-based and perceptual dissimilarities between categories were also calculated from the Euclidean distances between centroids (mean across eight exemplars) in the 20-dimensional space comprised of the 20 low-level image statistics and in the 12-dimensional space comprised of the 12 ratings obtained using the SD method. The obtained dissimilarities were displayed in percentiles as a 2D matrix format (dissimilarity matrix; Kriegeskorte et al., 2008). We found that the pattern of the image-based dissimilarity matrix was remarkably similar to the neural dissimilarity matrix of V1/V2, whereas the perceptual dissimilarity matrix was more similar to the neural dissimilarity matrix of FG/CoS than to that of V1/V2 (Fig. 4; Supplementary Fig. 5 for all ROIs).

To quantitatively compare the neural representation of material categories in each ROI with the image-based or perceptual representations, we computed the correlations between the neural dissimilarities and image-based and perceptual dissimilarities. The correlation between the neural and image-based dissimilarities was stronger in V1/V2 (one-tailed permutation test, \( p < 0.0001 \)) than in FG/CoS (\( p = 0.022 \); Fig. 5A; see Supplementary Fig. 5 for all ROIs), whereas the correlation between the neural and perceptual dissimilarities was strong in FG/CoS (\( p < 0.0001 \)) and was not significant in V1/V2 (\( p = 0.065 \); Fig. 5B). Because there was a correlation between image-based and perceptual dissimilarity, though that correlation was weak (\( r = 0.25, p = 0.063 \)), we calculated coefficients of neural-image partial correlation and neural-perceptual partial correlation for each ROI, while excluding the image-perceptual correlation. The neural-image partial correlation was significant in V1/V2 (one-tailed permutation test, \( p < 0.0001 \)) and V3/V4 (\( p = 0.0047 \)), and not in higher areas (FG/CoS: \( p = 0.082 \), LOS/pITs: \( p = 0.75 \), V3AB/IPS: \( p = 0.60 \); Fig. 5C). By contrast, the neural–perceptual partial correlation was not significant in V1/V2 (\( p = 0.22 \)), and significant in other 4 ROIs (V3AB/IPS: \( p = 0.014 \); others: \( p < 0.0001 \); Fig. 5D). No significant difference among the latter 4 ROIs was observed in pairwise comparisons (two-tailed permutation test, \( p = 0.05 \) Bonferroni-corrected). Similar results were obtained using different numbers of voxels per ROI per subject (Supplementary Fig. 7A, B). These results suggest that activities in V1/V2 discriminate material categories based on image-based dissimilarity, whereas activities in higher visual areas are based on perceptual dissimilarity.

We also conducted a series of analyses using another measure of neural dissimilarities. We computed the classification accuracy between all pairs in the 9 categories in each ROI using linear SVM and regarded the pairwise classification accuracies as neural dissimilarities between the pairs of categories (Said et al., 2010; Weber et al., 2009). This measure weighs more on the informative voxels, and
would be more powerful to detect difference between categories compared with END when the dimension of data (i.e., number of voxels) is large enough. We call this “classification-based neural dissimilarity” (CND). Consistent with analyses employing END, analyses with this measure demonstrated that the correlation between the neural and image-based dissimilarities was stronger in V1/V2 (one-tailed permutation test, $p=0.0001$) than in FG/CoS ($p=0.029$; Fig. 6A, see Supplementary Fig. 6 for all ROIs), whereas the correlation between the neural and perceptual dissimilarities was stronger in FG/CoS ($p=0.0001$; Fig. 6B) than in V1/V2 ($p=0.016$; Fig. 6B). The neural–image partial correlation was significant in V1/V2 and V3/V4 (one-tailed permutation test, $p \leq 0.0002$; Fig. 6C), whereas the neural–perceptual partial correlation was significant in FG/CoS ($p=0.0002$) and V3AB/IPS ($p=0.010$; Fig. 6D). Unlike the case with END, the neural–perceptual partial correlation was not significant in V3/V4 ($p=0.076$) and LOS/pITS ($p=0.054$), and tended to depend more on voxel number (Supplementary Fig. 7E). But, importantly, FG/CoS and V3AB/IPS consistently showed high and stable neural–perceptual partial correlation for 400 or more voxels per ROI.

The measures of neural dissimilarities we employed reflected not only the difference in the spatial pattern of activity among categories but also the overall difference in activities (difference in mean activity across voxels) in the ROIs. To examine whether the image-based or perceptual dissimilarity reflects the spatial pattern of activity, we computed neural dissimilarities END after subtracting spatial mean from the pattern of activity (mean-subtracted END) and conducted correlational analyses using this dissimilarity. We found that the results obtained with this measure resembled those obtained without subtracting mean activity, although the degree of correlation tended to decrease (Supplementary Fig. 8). The mean-subtracted END was significantly correlated with the image-based dissimilarity in V1/V2 and V3/V4 (one-tailed permutation test, $p<0.0001$), and with perceptual dissimilarity in FG/CoS ($p=0.0001$), LOS/pITS ($p=0.032$) and V3AB/IPS ($p=0.0027$) but not in V3/V4 ($p=0.067$). This pattern of results suggests that image-based and perceptual dissimilarities reflect, at least partly, the spatial pattern of activity. The decreased correlation, especially in V3/V4, suggests that mean activity would contribute as well.

The correlational analyses with these measures of dissimilarities between categories also suggest that the transition from image-based to perceptual representation would occur gradually along the visual area hierarchy. This transition can be well visualized by applying MDS to the image-based, perceptual and neural dissimilarities (Fig. 7, Supplementary movie 1). The MDS-derived space illustrates that the representations in V1/V2 and FG/CoS are closest to (i.e., most correlated with) the image-based and perceptual representations, respectively. Additionally, the neighboring areas are located more closely than distant areas and the overall configurations of neural representations resemble the area hierarchy in the visual cortex.

Hierarchical category structure in the representation of materials

One notable characteristic observed in the SD ratings of our stimuli was that the 9 categories were grouped into 3 superordinate categories. Remarkably, hierarchical cluster analysis of the CND in FG/CoS, though not of the END, revealed that the 9 material categories form the 3 perceptual superordinate categories and the topology of the dendrogram is identical to that in perception: SC2 (green) and SC3 (red) clustered first and SC1 (blue) became an out-group (Fig. 8A). This topology was observed only in the activity of FG/CoS and only when the 500–600 voxels/ROI were used. When we changed the number of voxels, the topology was degraded and became inconsistent with perception, which suggests selection of the appropriate voxels is very important to properly read out the superordinate categorical structure from brain activity, as was seen in previous studies on decoding high-level categories (Ethofer et al., 2009; Kriegeskorte et al., 2008). To quantify the degree to which neural representation is consistent with the superordinate categorical structure, we calculated a SCI (Fig. 8B, Supplementary Fig. 7C, F). The SCI is the difference between the mean dissimilarity for pairs of basic categories that belong to different superordinate categories and the mean dissimilarity for pairs of categories that belong to the same superordinate category, normalized by the standard deviation of the dissimilarities for all category pairs. A positive SCI indicates that the dissimilarity between different superordinate categories is greater than that within the same superordinate category. Regardless of the measures of neural dissimilarity, the SCI in FG/CoS was highest next to perception and was significantly positive (one-tailed permutation test, $p=0.0079$, 0.0013, with END and CND, respectively). Other ROIs showed no significantly positive SCIs (V1/V2: $p=0.54$, 0.27, V3/V4: $p=0.088$, 0.15, LOS/pITS: $p=0.16$, 0.082, V3AB/IPS: $p=0.19$, 0.42 with END and CND, respectively).

The role of object selective voxels on material representation

Because FG/CoS contained a large number of object-selective voxels, it was of interest to see whether object-selective and nonselective voxels in this region encode information about materials differently. To examine this, we divided FG/CoS into object-selective

Fig. 4. Dissimilarity matrices for low-level image-statistics, perception and neural activities (V1/V2 and FG/CoS). Measures of neural dissimilarity between categories were obtained from the Euclidean distance between mean patterns of activity for each category (Euclidean distance-based neural dissimilarities). In each ROI, the 500 most visually-responsive voxels in each subject were selected. The image-based and perceptual dissimilarities between categories were obtained from Euclidean distances between centroids (mean of 8 exemplars) in the 20-dimensional space comprised of low-level image statistics and the 12-dimensional space comprised of SD ratings, respectively. The color scale indicates the dissimilarity between category pairs in percentiles. Note that the dissimilarity matrices are symmetrical along the diagonal line. Ba: bark, Ce: ceramic, Fa: fabric, Fu: fur, GI: glass, Le: leather, Me: metal, St: stone, Wo: wood.
and nonselective subdivisions based on object-selectivity and analyzed neural dissimilarities of the subdivisions separately (Fig. 9). When 300 most visually-responsive voxels (maximally attainable number for some subjects) were selected from each subdivision, the activities in both subdivisions correlated significantly with perception, regardless of the measures of neural dissimilarity (one-tailed permutation test, object selective: \( p < 0.0001, p = 0.011 \), nonselective: \( p < 0.0001, p = 0.017 \), with END and CND, respectively), and not with image statistics (object selective: \( p \geq 0.11 \), nonselective: \( p \geq 0.14 \)). SCIs were significantly positive in both subdivisions (one-tailed permutation test, \( p = 0.0052, 0.039 \), nonselective: \( p = 0.016, 0.027 \), with END and CND, respectively). These results suggest that both object-selective and non-selective voxels in FG/CoS contributed in the representation of perceptual material categories in this region.

Fig. 5. Relationship between Euclidean distance-based neural dissimilarities (END) and image-based or perceptual dissimilarities. (A) Image-based dissimilarities plotted against neural dissimilarities for all category pairs in V1/V2 and FG/CoS. Horizontal axis represents Euclidean distance between categories in the 500-dimensional space representing activity of the 500 most visually-responsive voxels in each ROI. Vertical axis represents Euclidean distance between categories in the 20-dimensional space representing z-scored 20 low-level image statistics. The Pearson correlation coefficients are shown in inset. (B) Perceptual dissimilarities plotted against neural dissimilarities for all category pairs in V1/V2 and FG/CoS. Horizontal axis is the same as that in (A), Vertical axis represents Euclidean distance between categories in the 12-dimensional space representing z-scored 12 SD ratings. (C) Partial correlation coefficients between image-based dissimilarities and neural dissimilarities in each ROI. (D) Partial correlation coefficients between perceptual dissimilarities and neural dissimilarities in each ROI. * \( p < 0.05 \), ** \( p < 0.001 \), *** \( p < 0.0001 \) (one-tailed permutation test).

Fig. 6. Relationship between classification-based neural dissimilarities (CND) and image-based or perceptual dissimilarities. (A–B) Image-based dissimilarities (A) and perceptual dissimilarities (B) plotted against neural dissimilarities for all category pairs in V1/V2 and FG/CoS. Horizontal axes in (A–B) represent pairwise category classification accuracy obtained by using SVM. The 500 most visually-responsive voxels per ROI was used. Vertical axes in (A–B) are the same as those in Fig. 5A, B. Pearson correlation coefficients are shown in inset. (C) Partial correlation coefficients between image-based dissimilarities and neural dissimilarities in each ROI. (D) Partial correlation coefficients between perceptual dissimilarities and neural dissimilarities in each ROI. * \( p < 0.05 \), ** \( p < 0.001 \), *** \( p < 0.0001 \) (one-tailed permutation test).
The adjective pairs used in the SD method included visual and non-visual ones. Thus the extracted perceptual dissimilarity might reflect multimodal aspects of materials and the semantic knowledge of materials. Are the non-visual aspects of the materials actually driving the neural-perceptual correlation in FG/CoS? To examine this, we computed separately “visual dissimilarity” from the ratings by 5 adjective pairs that mainly characterize visual aspects (matte–glossy, opaque–transparent, simple–complex, regular–irregular, colorful–colorless) and “non-visual dissimilarity” from the ratings by 7 non-visual adjective pairs (smooth–rough, dry–wet, cold–warm, soft–hard, light–heavy, elastic–inelastic, natural–artificial; Fig. 10A). These visual and non-visual adjective pairs both highly correlated with the original perceptual dissimilarity (visual: \( r = 0.94 \), non-visual: \( r = 0.95 \)), suggesting that both visual and non-visual aspects contributed to the original perceptual dissimilarity to a similar degree. Partial correlation analyses revealed that neural dissimilarity in FG/CoS significantly correlated with both visual and non-visual dissimilarities (one-tailed permutation test, visual: \( p < 0.0001 \), \( p = 0.0004 \), non-visual: \( p < 0.0001 \), \( p = 0.0023 \), with END and CND, respectively; Fig. 10B). These results suggest that the activity in FG/CoS reflects not only visual but also non-visual aspects of materials.

**Discussion**

We have demonstrated that the patterns of activity in the early and higher visual cortex contain information distinguishing material categories. These areas are in good agreement with regions previously shown to be sensitive to surface properties such as color and texture (Bartels and Zeki, 2000; Cant et al., 2009; Cant and Goodale, 2007; Cavina-Pratesi et al., 2010a,b; Newman et al., 2005; Peuskens et al., 2004; Stilla and Sathian, 2008); however, the present study revealed profound differences in the way material information is represented across brain regions. By integrating analyses of image statistics and the perceptual properties of different materials, we found that neural representation gradually shifts from image-based representation in early areas to perceptual category representation in higher areas around the fusiform gyrus, along the ventral visual pathway.

Simple low-level image features, such as spatial frequency differed among the 9 material categories in this study. For instance, glossy materials like metal and ceramic tended to have a high standard deviation of luminance, a high degree of luminance skew, and a low spatial frequency reflecting their smooth mesostructure. Motoyoshi et al. (2007) and Sharan et al. (2008) demonstrated that low-level image statistics like luminance skew can explain the lightness and glossiness of a surface when it is perceived as a 3D shape. They hypothesized that early visual areas may have the potential to extract such information. Consistent with this idea, our analysis indicated that activity in V1/V2 contained information to distinguish material categories from one another and correlated well with the low-level image statistics, including standard deviation and skew of \( L^* \). This likely reflects the process of extracting such low-level image statistics. The remarkably high neural-image correlation in V1/V2 (\( r = 0.63, 0.81 \) with END and CND, respectively) also indicates that the low-level image statistics used in this study successfully characterized the visual information for the materials processed at the V1/V2 level.

Although low-level image features can be sources to distinguish some material categories, by themselves they would not be sufficient for material differentiation.
for perceptual identification and categorization of materials. The physical appearance of real-world materials (or natural texture) is intricate and their properties cannot be described using only a few parameters. Computational studies have shown that several hundred parameters are necessary to synthesize various classes of realistic, natural textures (Heeger and Bergen, 1995; Portilla and Simoncelli, 2000) and to classify natural images of objects into material categories that the objects are made from using a machine learning technique (Liu et al., 2010). Our visual system would be evolutionarily tuned to extract features useful for identification and categorization of materials. It is generally believed that gradually more complex visual features are extracted along the ventral visual pathway, and electrophysiological studies have shown that some neurons in monkey V4 (Arcizet et al., 2008; Hanazawa and Komatsu, 2001) as well as in the monkey inferior temporal cortex (Koteles et al., 2008) exhibit complex response selectivity for some classes of artificial and natural textures. Thus neurons in V3/V4 presumably represent mid-to-high-level features for material identification and categorization.

Our present findings are consistent with that idea, since the low-level image statistics that explained activity in V1/V2 quite well explained activities in V3/V4 only to a lesser degree. The results of MDS analysis (Fig. 7) also support the idea that V3/V4 is situated midway between V1/V2 and FG/CoS. Thus image features more complex than those analyzed in this study would be necessary to explain activities in areas other than V1/V2 and perceptual categorization of materials. To understand how our visual system efficiently extracts information about materials, it will be important to know what types of features explain the neural representations in higher areas and the perceptual representation.

We found that activity in FG/CoS, a higher-order area in the ventral visual cortex, reflected the perceptual relationship between material categories. The neural–perceptual correlation in FG/CoS as well as V3AB/IPS was robustly observed using different measures of neural dissimilarities, whereas those in other regions were not consistently observed. Importantly, a perceptual high-level categorical structure resided only in FG/CoS (Fig. 8). The SCI, which characterizes a degree to which neural representation reflects perceptual high-level categorical structure, was significant only in FG/CoS, regardless of the measures of neural dissimilarity. Furthermore, the perfect topology of perceptual superordinate categories could be reproduced from the activity in FG/CoS, when CND was used as a measure of neural dissimilarity and 500–600 voxels were selected from each subject. Although this number of voxels seems to be large compared with the...
size of single area in the ventral visual cortex, this is consistent with the view that information distinguishing some categories is not localized but spatially distributed in wide regions of the brain (e.g., Hanson et al., 2004; Haxby et al., 2001). Some MVPA studies have also used large number of voxels (400 or more) to read out high-level categories with SVM (Ethofer et al., 2009; Said et al., 2010). Taken together, we suggest that relatively large number of voxels is necessary because information about high-level material category would be relatively widely distributed in the ventral visual cortex and partly because of the methodological limitations.

In the present study, we employed two measures of neural dissimilarity between material categories, namely END and CND. The results obtained by the two measures were similar but there were also differences. On the one hand, END tended to show relatively high correlation with perception (Figs. 5 and 6) and the correlation was more stable than CND when voxel number per ROI was small (Supplementary Fig. 7). On the other, CND tended to show relatively high SCI in some ROIs when voxel number was large (Fig. 8, Supplementary Fig. 7). We note that the pattern of the results obtained using END after subtracting mean activity from each pattern of activity resembled those obtained with CND (Supplementary Fig. 8). This suggests that the inconsistencies between the two measures of neural dissimilarity are partly explained by the differences in the contribution of overall activity in the ROI; the effect of the overall activity on END would be relatively large, since all voxels equally contribute to END, whereas more informative voxels would make larger contribution in CND. Nonetheless, mean-subtracted END emphasized the difference in the spatial pattern of activity and yielded results more similar to those obtained by CND. Importantly, the neural-image correlation in V1/V2 and V3/V4, and the neural–perceptual correlation as well as SCI in FG/CoS were always significant regardless of the neural dissimilarity measures when 500 voxels/ROI were used. Thus the choice of the neural dissimilarity measure does not affect our main conclusion that the activities in V1/V2 and V3/V4 reflected low-level image statistics and that activity in FG/CoS reflected perceptual high-level categories of materials.

Our measure of the perceptual relationship between material categories is based on ratings of impressions characterized using the SD method. We confirmed that this measure of perceptual dissimilarity was stable even when one or two of the 12 adjective pairs were excluded (correlation with original dissimilarity was larger than 0.96). Furthermore, in a separate experiment, we confirmed that the perceptual dissimilarity used in the present study was not biased to the specific method employed, by testing perceptual dissimilarity by direct judgments of material similarity between a pair of images from different categories. The perceptual dissimilarity obtained by the similarity judgment was highly correlated with the perceptual dissimilarity obtained by the SD method (r = 0.81, see Supplementary methods and Supplementary Fig. 9) and was also significantly correlated with neural dissimilarity in FG/CoS (neural–perceptual partial correlation = 0.57, 0.37, p = 0.0002, 0.017 with END and CND, respectively).

Some of the adjectives used to measure the perceptual dissimilarity reflect non-visual impressions of materials, and we showed that the activity in FG/CoS did correlate with dissimilarity in such non-visual as well as visual aspects of materials (Fig. 10). Interestingly, it has been reported that ventral cortical regions around the fusiform gyrus are responsive to tactile and auditory material stimuli (Arnott et al., 2008; Pietrini et al., 2004). Presumably, the ventral higher-order area integrates information about materials across modalities and constructs supramodal representations of materials. Such information could be matched with stored memories about the functions of objects, and could be used to adequately perform tasks with objects made from various materials.

One may argue that activity in FG/CoS reflects not the representation of material category per se but the representation of object category associated with the materials. In previous studies, it has been shown that higher-order area likely overlapping with FG/CoS carries information about object category (e.g., Haxby et al., 2001; Kriegeskorte et al., 2008), and is activated during the tasks requiring object knowledge (Chao et al., 1999; Goldberg et al., 2006). In line with this, one possible interpretation is that the structure of superordinate categories in FG/CoS could be related to specific object category; for example, proximity of activities in FG/CoS to leather, fabric, and fur that belong to the same superordinate perceptual category in FG/CoS could be explained by the close relation of these materials to object category such as “clothing” or “things that are draped over the bodies of various animals”. However, we showed that the neural representation in FG/CoS was still matched with perception even when the analysis was restricted in the object-nonscopic subdivision within FG/CoS (Fig. 9). Thus we may reasonably suggest that the activity in this region reflects not the object category but the material categories as mentioned above.

The representation in the LOS/pITs, located lateral to the FG/CoS, matched with perception in a lesser degree than that in FG/CoS. Although the lateral region carried information about materials (Fig. 3), this region did not exhibit consistently significant correlation with perceptual impressions (Figs. 5D and 6D), and showed low SCIs (Fig. 8B). This suggests a reduced role of lateral higher-order area in material perception, consistent with recent fMRI results obtained with selective attention and fMRI adaptation paradigms in that the lateral and medial portions of the occipital cortex are involved in coding object shape and surface-related information, respectively (Cant et al., 2009; Cant and Goodale, 2007; Cavina-Pratesi et al., 2010a,b).

Additionally, we found that V3A/IPS, located in the dorsal/parietal visual cortex, does not contain information that reliably distinguishes material categories (Fig. 3), but it correlates with perceptual impressions, marginally (Figs. 5D and 6D). Since the dorsal/parietal region plays important roles in visually-guided action (Goodale and Milner, 1992), it is possible that activity in this region exhibits correlation with judgments more related to action on the materials. Our present results may also be affected by the experimental conditions under which the subjects viewed materials while attending to the fixation spot. It would be interesting to know whether this region contributes to the material judgments when subjects are engaged in active tasks that require recognizing and/or acting on the materials.

Overall, our results show that material information is transformed into perceptual representations reflecting multimodal impressions and high-level categories in the ventral higher-order area. This suggests a role for the frontal cortex in the development or do they have an evolutionary origin? From this perspective, it will be important to examine material perception in humans and animals. In the present study, we focused on the brain activity and perception in adult humans who have been exposed to many real-world materials, which raised the question: are high-level categories constructed during development or do they have an evolutionary origin? From this perspective, it will be important to examine material perception in human infants and animals and relate their representations to those in human adults. The method used here could be extended for that purpose. The combination of pattern-information analysis of brain imaging data and psychological experiments like those used in the present study will become a powerful tool with which to address questions that were difficult to address in the past about the way sensory information is represented in the brain.

Acknowledgments

We thank Hiroki Tanabe and Norihiro Sadato for help conducting fMRI experiments. This study was supported by Grant-in-Aid for Scientific Research on Priority Areas — System study on higher-order brain functions (17022040) and Grant-in-Aid for Scientific Research on Innovative Areas (22135007) from the Ministry of Education, Culture, Sports, Science and Technology of Japan to H.K., Grants-in-Aid for Research on Priority Areas

Please cite this article as: Hiramatsu, C., et al., Transformation from image-based to perceptual representation of materials along the human ventral visual pathway, NeuroImage (2011), doi:10.1016/j.neuroimage.2011.04.056

C. Hiramatsu et al. / NeuroImage xxx (2011) xxx–xxx 12

ARTICLE IN PRESS