Neuron
Article

Visual Image Reconstruction from Human Brain Activity using a Combination of Multiscale Local Image Decoders

Yoichi Miyawaki,[1,2,6] Hajime Uchida,[2,3,6] Okito Yamashita,[2] Masa-aki Sato,[2] Yusuke Morito,[4,5] Hiroki C. Tanabe,[4,5] Norihiro Sadato,[4,5] and Yukiyasu Kamitani[2,3,*]
[1]National Institute of Information and Communications Technology, Kyoto, Japan
[2]ATR Computational Neuroscience Laboratories, Kyoto, Japan
[3]Nara Institute of Science and Technology, Nara, Japan
[4]The Graduate University for Advanced Studies, Kanagawa, Japan
[5]National Institute for Physiological Sciences, Aichi, Japan
[6]These authors contributed equally to this work
*Correspondence: kmtn@atr.jp
DOI 10.1016/j.neuron.2008.11.004

SUMMARY

Perceptual experience consists of an enormous number of possible states. Previous fMRI studies have predicted a perceptual state by classifying brain activity into prespecified categories. Constraint-free visual image reconstruction is more challenging, as it is impractical to specify brain activity for all possible images. In this study, we reconstructed visual images by combining local image bases of multiple scales, whose contrasts were independently decoded from fMRI activity by automatically selecting relevant voxels and exploiting their correlated patterns. Binary-contrast, 10 × 10-patch images (2^{100} possible states) were accurately reconstructed without any image prior on a single trial or volume basis by measuring brain activity only for several hundred random images. Reconstruction was also used to identify the presented image among millions of candidates. The results suggest that our approach provides an effective means to read out complex perceptual states from brain activity while discovering information representation in multivoxel patterns.

INTRODUCTION

Objective assessment of perceptual experience in terms of brain activity represents a major challenge in neuroscience. Previous fMRI studies have shown that visual features, such as orientation and motion direction (Kamitani and Tong, 2005, 2006), and visual object categories (Cox and Savoy, 2003; Haxby et al., 2001) can be decoded from fMRI activity patterns by a statistical "decoder," which learns the mapping between a brain activity pattern and a stimulus category from a training data set. Furthermore, a primitive form of "mind-reading" has been demonstrated by predicting a subjective state under the presentation of an ambiguous stimulus using a decoder trained with unambiguous stimuli (Kamitani and Tong 2005, 2006; Haynes and Rees, 2005). However, such a simple classification approach is insufficient to capture the complexity of perceptual experience, since our perception consists of numerous possible states, and it is impractical to measure brain activity for all the states. A recent study (Kay et al., 2008) has demonstrated that a presented image can be identified among a large number of candidate images using a receptive field model that predicts fMRI activity for visual images (see also Mitchell et al., 2008, for a related approach). But the image identification was still constrained by the candidate image set. Even more challenging is visual image reconstruction, which decodes visual perception into an image, free from the constraint of categories (see Stanley et al., 1999, for reconstruction using LGN spikes).

A possible approach is to utilize the retinotopy in the early visual cortex. The retinotopy associates the specific visual field location to the active cortical location, or voxel, providing a mapping from the visual field to the cortical voxels (Engel et al., 1994; Sereno et al., 1995). Thus, one may predict local contrast information by monitoring the fMRI signals corresponding to the retinotopy map of the target visual field location. The retinotopy can be further elaborated using a voxel receptive-field model. By inverting the receptive-field model, a presented image can be inferred given the brain activity consistent with the retinotopy (Thirion et al., 2006).

However, it may not be optimal to use the retinotopy or the inverse of the receptive field model to predict local contrast in an image. These methods are based on the model of individual voxel responses given a visual stimulus, and multivoxel patterns are not taken into account for the prediction of the visual stimulus. Recent studies have demonstrated the importance of the activity pattern, in particular the correlation among neurons or cortical locations in the decoding of a stimulus (Averbeck et al., 2006; Chen et al., 2006). Since even a localized small visual stimulus elicits spatially spread activity over multiple cortical voxels (Engel et al., 1997; Shmuel et al., 2007), multivoxel patterns may contain information useful for predicting the presented stimulus.

In addition, a visual image is thought to be represented at multiple spatial scales in the visual cortex, which may serve to

retain the visual sensitivity to fine-to-coarse patterns at a single visual field location (Campbell and Robson, 1968; De Valois et al., 1982). The conventional retinotopy, by contrast, does not imply such multiscale representation, as it simply posits a location-to-location mapping. It may be possible to extract multiscale information from fMRI signals and use it to achieve better reconstruction.

Here, we present an approach to visual image reconstruction using multivoxel patterns of fMRI signals and multiscale visual representation (Figure 1A). We assume that an image is represented by a linear combination of local image elements of multiple scales (colored rectangles). The stimulus state at each local element (C_i, C_j, …) is predicted by a decoder using multivoxel patterns (weight set for each decoder, w_i, w_j, …), and then the outputs of all the local decoders are combined in a statistically optimal way (combination coefficient, λ_i, λ_j, …) to reconstruct the presented image. As each local element has fewer possible states than the entire image, the training of local decoders requires only a small number of training samples. Hence, each local decoder serves as a "module" for a simple image component, and the combination of the modular decoders allows us to represent numerous variations of complex images. The decoder uses all the voxels from the early visual areas as the input, while automatically pruning irrelevant voxels. Thus, the decoder is not explicitly informed about the retinotopy mapping.

We applied this approach to the reconstruction of contrast-defined images consisting of 10 × 10 binary patches (Figure 1B). We show that once our model is trained with several hundred random images, it can accurately reconstruct arbitrary images (2^{100} possible images), including geometric and alphabet shapes, on a single trial (6 s/12 s block) or volume (2 s) basis, without any prior information about the image. The reconstruction accuracy is quantified by image identification performance, revealing the ability to identify the presented image among a set of millions of candidate images. Analyses provide evidence that the multivoxel pattern decoder, which exploits voxel correlations especially in V1, and the multiscale reconstruction model both significantly contribute to the high quality of reconstruction.

RESULTS

In the present study, we attempted to reconstruct visual images defined by binary contrast patterns consisting of 10 × 10 square patches (Figure 1). Given fMRI signals **r**, we modeled a reconstruction image $\hat{I}(x|\mathbf{r})$ by a linear combination of local image bases (elements) $\phi_m(x)$ (Olshausen and Field, 1996),

$$\hat{I}(x|\mathbf{r}) = \sum_m \lambda_m C_m(\mathbf{r}) \phi_m(x),$$

where x represents a spatial position in the image, $C_m(\mathbf{r})$ is the contrast of each local image basis predicted from fMRI signals, and λ_m is the combination coefficient of each local image basis.

The local image bases, $\phi_m(x)$, were prefixed such that they redundantly covered the whole image with multiple spatial scales. We used local image bases of four scales: 1 × 1, 1 × 2, 2 × 1, and 2 × 2 patch areas. They were placed at every location in the image with overlaps. Thus, $\phi_m(x)$ served as overcomplete basis functions, the number of which was larger than that of all the patches. Although image elements larger than 2 × 2 or those with nonrectangular shapes could be used, the addition of such elements did not improve the reconstruction performance.

For each local image basis, we trained a "local decoder" that predicted the corresponding contrast using a linearly weighted sum of fMRI signals. The weights of voxels, **w**, were optimized using a statistical learning algorithm described in Experimental Procedures ("sparse logistic regression," Yamashita et al., 2008) to best predict the contrast of the local image element with a training data set. Note that our algorithm automatically selected the relevant voxels for decoding without explicit information about the retinotopy mapping measured by the conventional method.

The combination coefficient, λ_m, was optimized to minimize the errors between presented and reconstructed images in a training data set. This coefficient was necessary because the local image bases were overcomplete and not independent of each other. Trained local decoders and their combination coefficients constituted a reconstruction model.

fMRI signals were measured while subjects viewed a sequence of visual images consisting of binary contrast patches on a 10 × 10 grid. In the "random image session," a random pattern was presented for 6 s followed by a 6 s rest period (Figure 1B). A total of 440 different random images were shown (each presented once). In the "figure image session," an image forming a geometric or alphabet shape was presented for 12 s followed by a 12 s rest period. Five alphabet letters and five geometric shapes were shown six or eight times. We used fMRI signals from areas V1 and V2 for the analysis (unless otherwise stated). The data from the random image session were analyzed by a cross-validation procedure for quantitative evaluation. They were also used to train a model to reconstruct the images presented in the figure image session.

Reconstructed Visual Images

Reconstructed images from all trials of the figure image session are illustrated in Figure 2A. They were reconstructed using the model trained with all data from the random image session. Reconstruction was performed on single-trial, block-averaged data (average of 12 s or six-volume fMRI signals). Note that no postprocessing was applied. Even though the geometric and alphabet shapes were not used for the training of the reconstruction model, the reconstructed images reveal essential features of the original shapes. The spatial correlation between the presented and reconstructed images was 0.68 ± 0.16 (mean ± s.d.) for subject S1 and 0.62 ± 0.09 for S2.

We also found that reconstruction was possible even from 2 s single-volume data without block averaging (Figure 2B). The results show the temporal evolution of volume-by-volume reconstruction including the rest periods. All reconstruction sequences are presented in Movie S1.

Image Identification via Reconstruction

To further quantify reconstruction performance, we conducted image identification analysis (Kay et al., 2008; Thirion et al., 2006) in which the presented image was identified among a

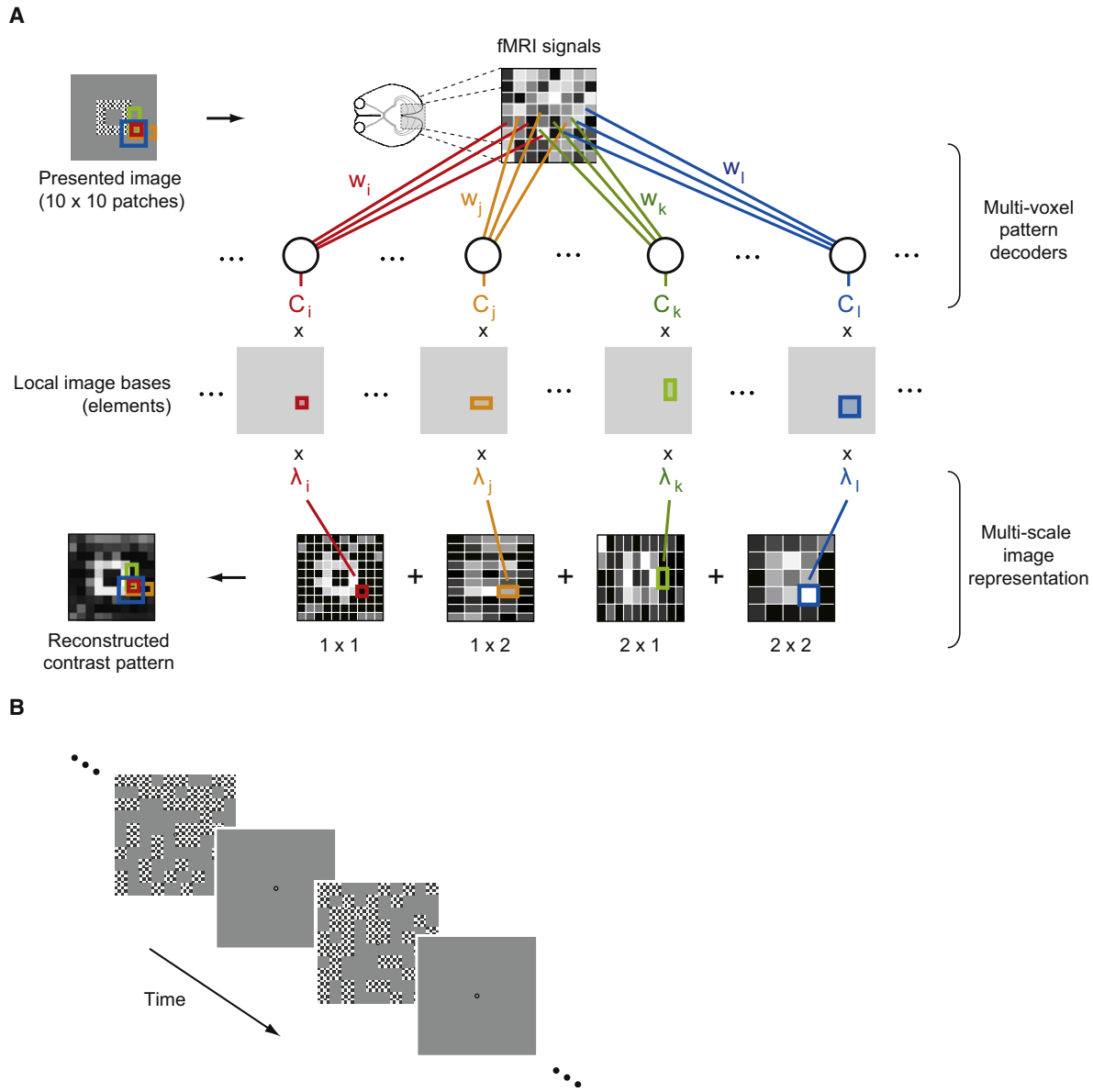


Figure 1. Visual Image Reconstruction

(A) Reconstruction procedure. fMRI activity is measured while a contrast-defined 10 × 10 patch image is presented. "Local decoders" use linearly weighted multivoxel fMRI signals (voxel weights, w_i, w_j, …) to predict the contrasts (contrast values, C_i, C_j, …) of "local image bases" (or elements) of multiple scales (1 × 1, 1 × 2, 2 × 1, and 2 × 2 patch areas, defined by colored rectangles). Local image bases are multiplied by the predicted contrasts and linearly combined using "combination coefficients" (λ_i, λ_j, …) to reconstruct the image. Contrast patterns of the reconstructed images are depicted by a gray scale. Image bases of the same scale (except the 1 × 1 scale) partially overlapped with each other, though the figure displays only nonoverlapping bases for the purpose of illustration.

(B) Sequence of visual stimuli. Stimulus images were composed of 10 × 10 checkerboard patches flickering at 6 Hz (patch size, 1.15° × 1.15°; spatial frequency, 1.74 cycle/°; contrast, 60%). Checkerboard patches constituted random, geometric, or alphabet-letter patterns. Each stimulus block was 6 s (random image) or 12 s (geometric or alphabet shapes) long followed by a rest period (6 or 12 s).

number of candidate images using an fMRI activity pattern (Figure 3A). We generated a candidate image set consisting of an image presented in the random image session and a specified number of random images selected from 2^{100} possible images (combinations of 10 × 10 binary contrasts). Given an fMRI activity pattern, image identification was performed by selecting the image with the smallest mean square difference from the reconstructed image. The rate of correct identification was calculated for a varied number of candidate random images.

In both subjects, image identification performance was far above the chance level, even up to an image set size of 10 million (Figure 3B). The identification performance can be further

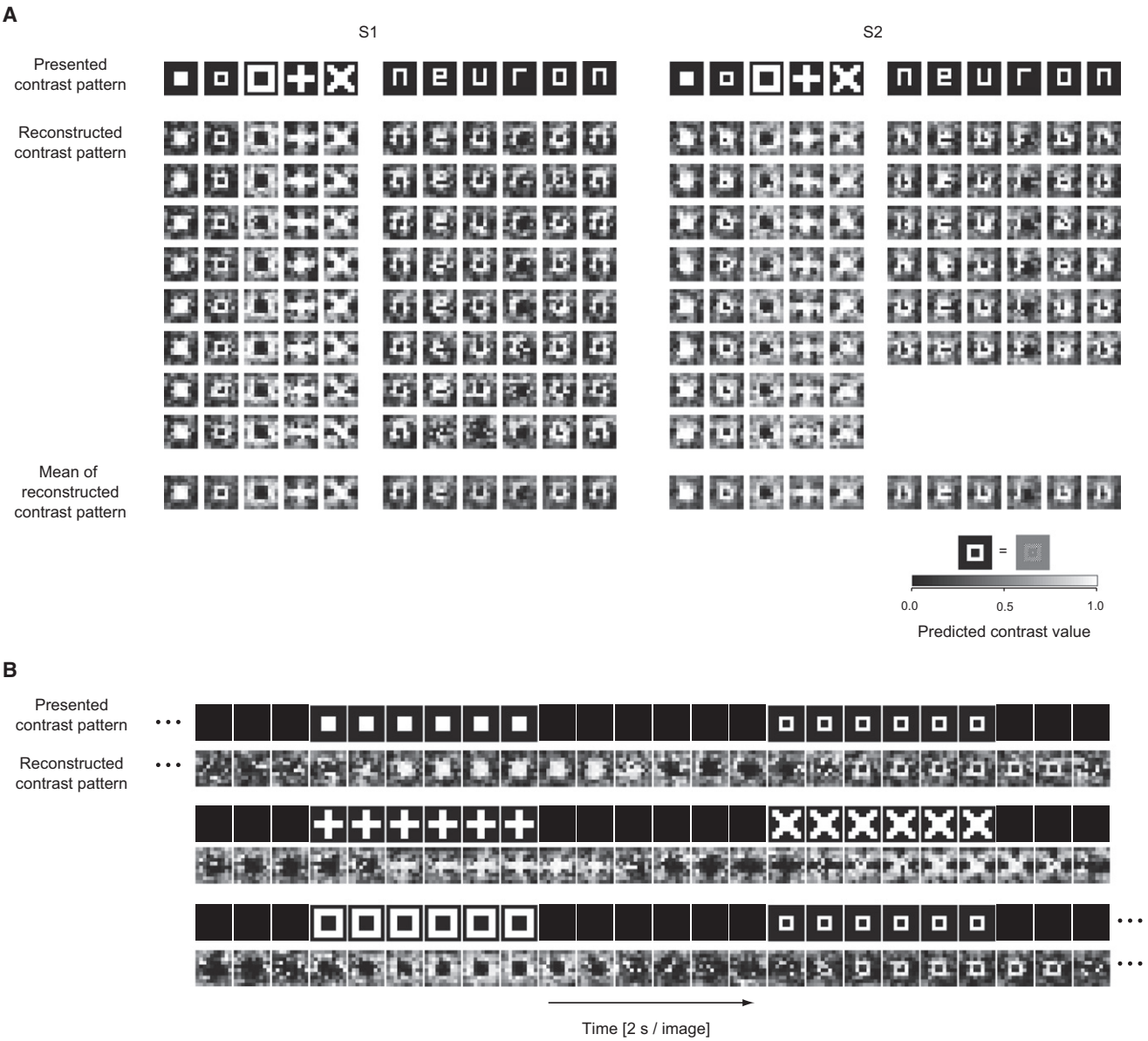


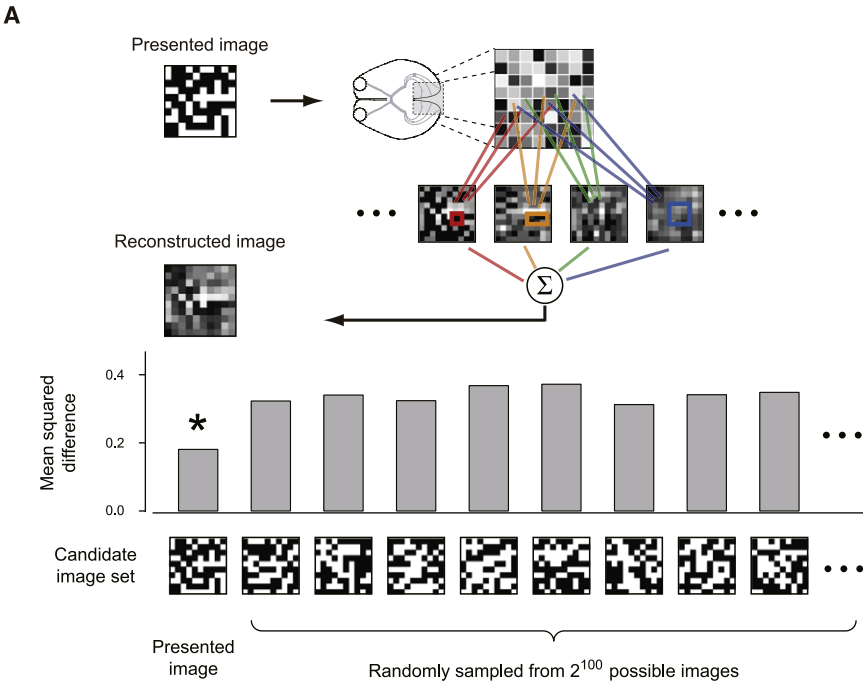
Figure 2. Reconstruction Results

(A) Reconstructed visual images. The reconstruction results of all trials for two subjects are shown with the presented images from the figure image session. The reconstructed images are sorted in ascending order of the mean square error. For the purpose of illustration, each patch is depicted by a homogeneous square, whose intensity represents the contrast of the checkerboard pattern. Each reconstructed image was produced from the data of a single trial, and no postprocessing was applied. The mean images of the reconstructed images are presented at the bottom row. The same images of the alphabet letter "n" are displayed in the rightmost and leftmost columns for each subject.

(B) Visual image reconstruction from a single-volume fMRI activity pattern. Representative reconstruction results are shown with the presented images, including the rest periods (subject S1; 2 s/image). Each reconstructed image was produced from a 2 s single-volume fMRI activity pattern, and no postprocessing was applied. The hemodynamic response delay is not compensated in this display.

extrapolated by fitting the sigmoid function. The extrapolation suggests that performance above 10% correct could be achieved even with image sets of $10^{10.8}$ for S1 and of $10^{7.4}$ for S2, using block-averaged single-trial data. The identification performance with 2 s single-volume data was lower than that of block-averaged data, but was still above the chance level for a large number of candidate images (above 10% correct with image sets of $10^{8.5}$ for S1 and of $10^{5.8}$ for S2).

In the following sections, we examine how multivoxel patterns and multiscale image representation, critical components of our reconstruction model, contributed to the high reconstruction performance.



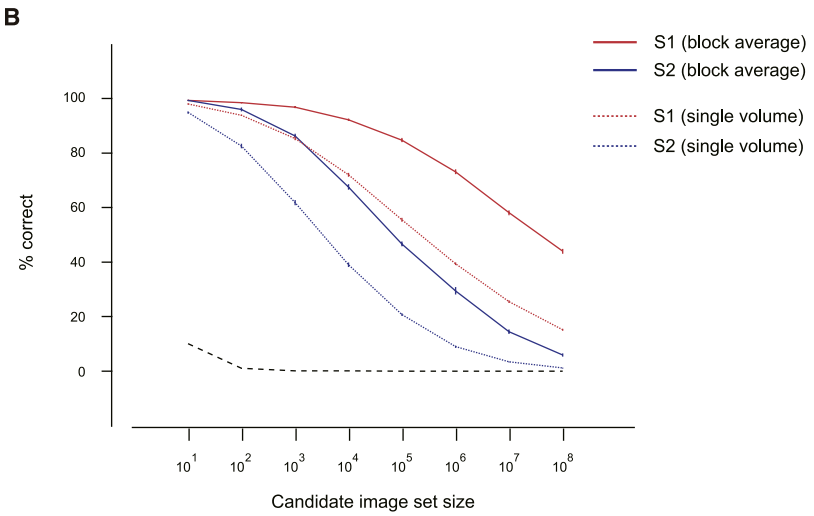


Figure 3. Image Identification using Reconstructed Images

(A) Image identification procedure. The mean square difference is measured between a reconstructed image and each image in the candidate set consisting of the presented image and a specified number of randomly generated images (bar graphs). The one with the smallest difference is identified as the predicted image (marked by an asterisk). The figure depicts an example of correct identification.

(B) Identification performance as a function of image set size. Identification was repeated for 20 candidate sets of randomly generated images for each reconstructed image from the random image session. Reconstructed images were obtained using either block-averaged data (6 s) or single-volume data (2 s). The percentage of correct identification was averaged across the candidate sets (error bars, s.d.; dashed line, chance level).

Weight Distribution on the Cortical Surface

Our algorithm for training local decoders automatically selected relevant voxels and assigned weights, thereby yielding robust classification performance (Yamashita et al., 2008; see Figure S1 for comparison with conventional algorithms without sparse voxel selection). We first examined the distributions of voxel weights of local decoders in comparison with the conventional retinotopy. Cortical surface maps show the distributions of weight magnitudes for a foveal and a peripheral patch (Figure 4A). The largest weight values are found around the cortical locations consistent with the retinotopic representation of the patch locations. The distributions of voxel weights for 1 × 1 decoders are summarized in Figure 4B. Decoders were sorted by the eccentricity and the polar angle of their target locations, and voxels were sorted by their corresponding eccentricity and polar angle identified by the conventional retinotopy mapping from a separate experiment. Large weight values were distributed around the diagonal line, indicating that local decoders mainly used voxels corresponding to the retinotopic locations for their target patches (see Figure S2 for comparison between the conventional retinotopy and the map of the voxels relevant for the decoding). The weight distribution tended to be blurred for peripheral patches, indicating that peripheral decoders failed to select retinotopic voxels (Figure 4B, left). Along the polar angle, the patches around the vertical and horizontal meridians (0/180° and 90°, respectively) showed higher correlation with the retinotopy than those at other angles (Figure 4B, right). This is partly because the locations of the patches arranged on the square grid have anisotropy with respect to the polar angle: the patches around the meridians are located at smaller eccentricity than those at nonmeridian angles on average. When the eccentricity was matched between the meridian and non-meridian patches, the difference became less pronounced.

Advantage of Multivoxel Pattern Decoders

Our local decoders were trained to exploit multivoxel patterns for the prediction of target contrast ("multivoxel pattern decoders"). However, as noted above, the locations of the selected voxels were largely consistent with the conventional retinotopy. Thus, a simple mapping between a cortical location [single or group of voxel(s)] and a stimulus position might be sufficient for the decoding.

To examine whether multivoxel patterns were effectively used for the decoding, we devised other types of local decoders that

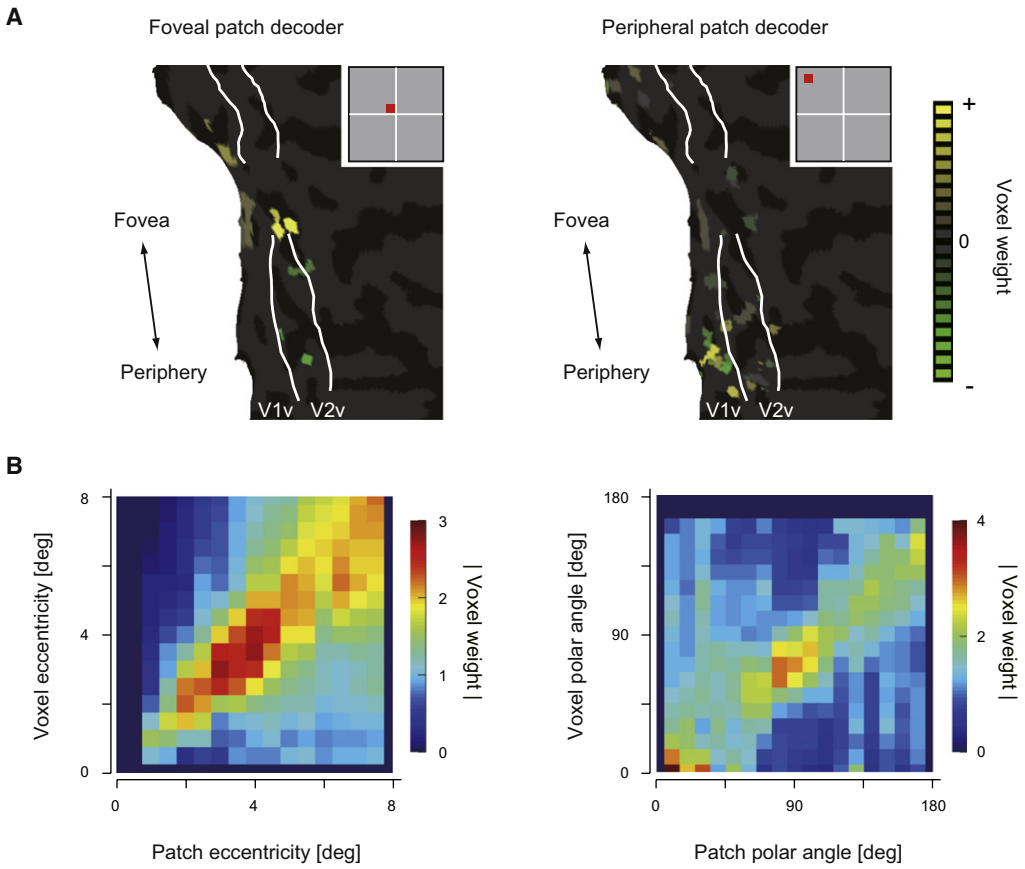


Figure 4. Weight Distribution on the Visual Cortex

(A) Distributions of voxel weights on the flattened cortex for a foveal and a peripheral decoder. Voxel weights are shown on the right visual cortical surface of subject S1. The location of each patch (1 × 1) is indicated in the inset of the top-right corner. The white lines are the V1 and V2 borders.

(B) Summary of voxel weight distribution. Local decoders (1 × 1) for the left visual field were sorted by the eccentricity and the polar angle of their targets (horizontal axis, 0.5° bins for eccentricity and 10° bins for polar angle), and contralateral voxels were sorted by their corresponding eccentricity and polar angle identified by the conventional retinotopy mapping (vertical axis, 0.5° bins for eccentricity and 10° bins for polar angle). The magnitudes of voxel weights were averaged in each target location and cortical location for ten models generated by the cross-validation analysis (two subjects pooled). Similar results were observed for the local decoders for the right visual field.

only used retinotopic voxels ("retinotopic decoders"). By applying the standard general linear model to the data from the random image session, we identified a single voxel with the highest t value, or a group of significantly responsive voxels ($p < 0.05$, false discovery rate [FDR] corrected for multiple comparisons) for each patch (Figure S3). This technique, known as the multifocal retinotopy mapping, gives the equivalent of the conventional phase-encoded retinotopy map (Hansen et al., 2004; Vanni et al., 2005). We used (1) the most responsive voxel, (2) the average of the significantly responsive voxels, or (3) the "t value weighted" average of the significantly responsive voxels as the input. The decoders consisted of the standard univariate logistic regression model. The performance of these decoders was compared with that of the multivoxel pattern decoder.

Cross-validation analysis using the random image trials revealed that the multivoxel pattern decoder achieved significantly higher correct rates than either of the three retinotopic decoders (two-way ANOVA, Bonferroni-corrected $p < 0.05$ for multiple comparisons), while the difference gradually diminished at the periphery approaching the chance level (Figure 5A). Although the figure illustrates the performance only for the 1 × 1 scale, the decoders of other scales showed similar results. The number of the significantly responsive voxels was larger than the number of the voxels selected by the multivoxel pattern decoder for the foveal to middle eccentricity. Since in this range of eccentricity the multivoxel pattern decoder largely outperformed the retinotopic decoders, the higher performance of the multivoxel pattern decoder is not merely due to noise reduction by pooling multivoxel signals. These results indicate that our local decoders did not simply depend on the mapping between a cortical location and a stimulus location, but that they effectively exploited multivoxel patterns.

One of the key features of multivoxel patterns is the correlation between voxels. Our multivoxel pattern decoder takes into account the correlation between voxels in the training data to determine voxel weight parameters, as is the case with other multivariate statistical methods (see Supplemental Data; Yamashita et al., 2008). To examine how voxel correlations contribute

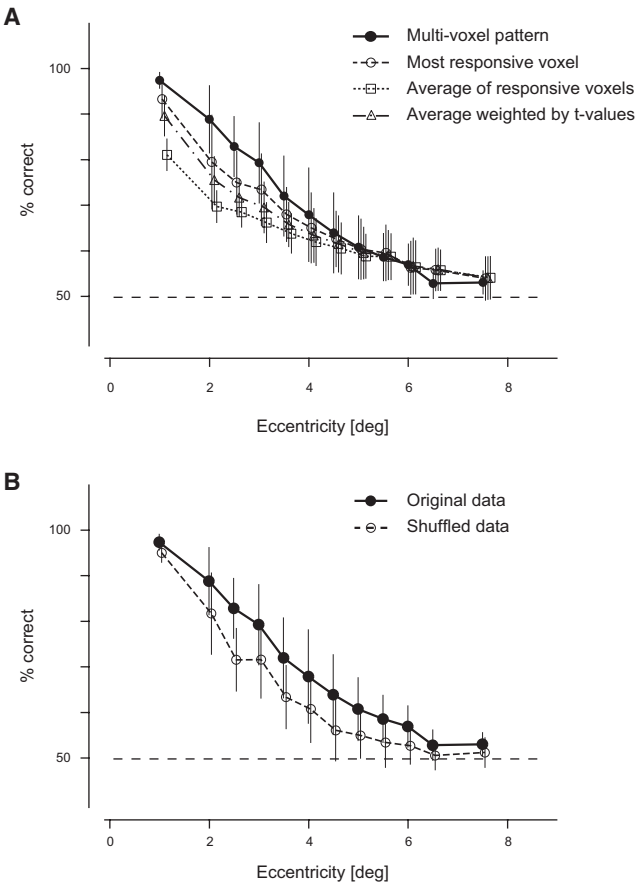


Figure 5. Advantage of Multivoxel Pattern Decoder

(A) Performance of the multivoxel pattern decoder and retinotopic decoders. The binary classification performance for 1 × 1 patches is plotted as a function of eccentricity. Classification was performed using (1) a multivoxel pattern, (2) the most responsive voxel for each patch (with the highest t value), (3) the mean of significantly responsive voxels for each patch ($p < 0.05$, FDR corrected for multiple comparisons), or (4) the mean of the significantly responsive voxels weighted by their t values for each patch. The performance was evaluated by cross-validation using data from the random image session. The average performance was calculated in each 0.5° eccentricity bin (two subjects pooled; error bars, s.d., dashed line, chance level).

(B) Effect of voxel correlation in training data. Performance is compared between the multivoxel pattern decoders trained with the original data and the same decoders trained with "shuffled" data, in which voxel correlations were removed. The results for the multivoxel pattern decoder are the same as those displayed in (A).

to decoding accuracy, we trained the decoder with fMRI data in which voxel correlations were removed and compared the performance with that of the original decoder. The data were created by shuffling the order of the trials with the same stimulus label in each voxel (Averbeck et al., 2006). This shuffling procedure removes voxel correlations that are independent of the stimulus label. Note that since the stimuli were random images, the voxel correlations observed in the original training data do not reflect the correlations between stimulus patches. The trained decoder was tested with independent nonshuffled data.

The performance with shuffled data was significantly lower than that with the original data (two-way ANOVA, $p < 0.05$), particularly at the middle range of eccentricity (Figure 5B). The results suggest that the multivoxel pattern decoder makes effective use of voxel correlation to achieve high decoding performance.

Reconstruction using Individual Visual Areas

We have thus far shown the results obtained using the voxels from V1 and V2 as the input to the decoders. We next compared the reconstruction between individual visual areas by using the voxels in each of V1, V2, and V3 as the input. As illustrated in Figure 6A, reconstruction quality progressively deteriorated along the visual cortical hierarchy. Quantitative comparison was performed by calculating the reconstruction errors for the images from the random image session (squared difference between the presented and the reconstructed contrast in each patch averaged over each entire image). Higher visual areas showed significantly larger errors than V1 (Figure 6B; ANOVA, Bonferroni-corrected $p < 0.05$ for multiple comparisons), indicating that V1 contains most reliable information for reconstructing visual images.

Inspection of the models for these three visual areas revealed the following differences. First, in higher areas, the selected voxels were less localized to the retinotopic locations than in V1 (Figure 6C). Second, the shuffling analysis on the local decoders showed that the performance significantly decreased for all areas when voxel correlations were removed (Figure 6D, 1 × 1 decoders; ANOVA, $p < 0.05$; other scales showed similar results). The performance difference between the original and shuffled data was prominent in V1 but diminished in higher areas, indicating the critical contribution of voxel correlations in V1. Finally, the difference in reconstruction error was also largest in V1 (Figure 6E; ANOVA, Bonferroni-corrected $p < 0.05$ for multiple comparisons), consistent with the performance of the local decoders. These findings suggest that the reliable information available in V1 is represented not only in the ordered retinotopic organization, but also in the correlated voxel patterns.

When we used all voxels from V1 to V4 together as the input to the decoder, most of the nonzero weights were found around the retinotopic voxels in V1, but not in the higher areas (Figure S4). The quality of image reconstruction remained similar to that obtained by V1 and V2 voxels. This preference to V1 voxels may also be accounted for by the fine retinotopic organization and the informative voxel correlations available in V1, from which our decoder can effectively extract information.

Advantage of a Multiscale Reconstruction Model

We then tested the significance of the multiscale representation by comparing the multiscale model with single-scale models that consisted of optimally combined, single-scale image bases (1 × 1, 1 × 2, 2 × 1, or 2 × 2; V1 and V2 voxels used as the input). Representative examples of the reconstructed images obtained from the figure image session are presented in Figure 7A. The reconstructed image from the 1 × 1 scale model showed fine edges but exhibited patchy noise. By contrast, the 2 × 2 scale model produced a spatially blurred image. The images from the 1 × 2 and 2 × 1 scale models contained horizontally and vertically

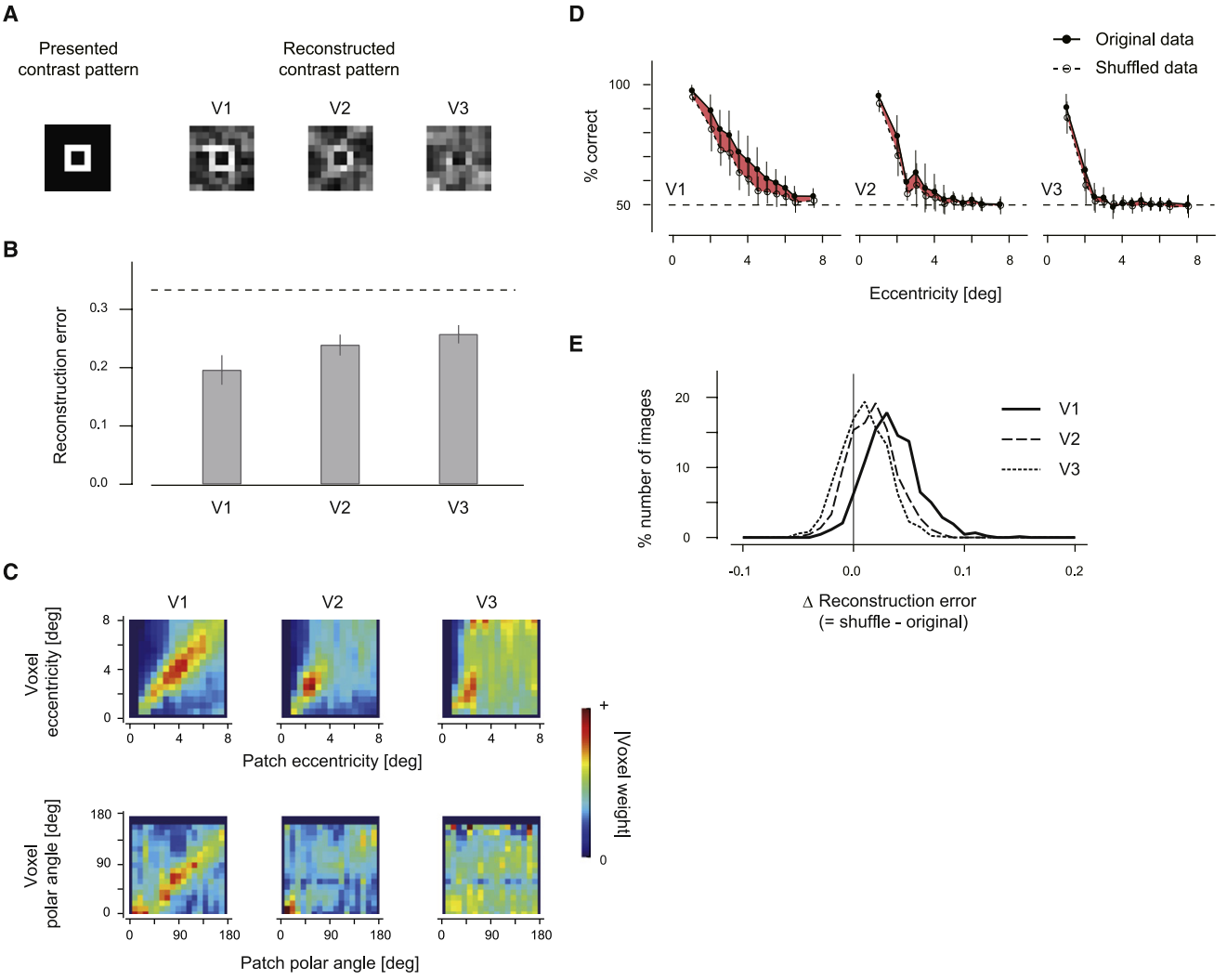


Figure 6. Reconstruction using Individual Visual Areas

(A) Reconstructed images. Examples from the figure image session (S1, "small frame") are shown.

(B) Reconstruction performance with entire images. The bar graph shows reconstruction errors, averaged across all test images in the random image session (two subjects pooled; error bars, s.d.). The dashed line indicates the chance level (1/3), which is achieved when a contrast value for each patch is randomly picked from the uniform distribution of 0 to 1.

(C) Distribution of voxel weights. The results of the local decoders (1 × 1) for the left visual field are displayed as in Figure 4B (color scale normalized for each visual area). Similar results were observed for the local decoders of the right visual field.

(D) Effect of voxel correlations on local decoders. Performance is compared between local decoders (1 × 1) trained with the original data and those trained with "shuffled" data for each visual area. Colored regions show the differences between the original and the shuffled data.

(E) Effect of correlations on reconstructed images. The distribution of difference in reconstruction error (<error with the shuffled> − <error with the original>, for each image) is plotted for each visual area (two subjects pooled).

elongated components. The reconstructed image from the multiscale model appears to have balanced features of these individual scales. The reconstruction error of the multiscale model, calculated with the images from the random image session, was significantly smaller than those of the single-scale models (Figure 7B; ANOVA, Bonferroni-corrected $p < 0.05$ for multiple comparisons).

We also calculated reconstruction errors at each eccentricity (Figure 7C). For all scales, the reconstruction error increased with eccentricity, but the profiles were different. The error sharply increased with eccentricity for the 1 × 1 model, while the profile was rather flat for the 2 × 2 model. As a result, the errors for these models were reversed at the fovea and the periphery. The 1 × 2 and 2 × 1 models showed intermediate profiles. Statistical analysis revealed a significant interaction between scale and eccentricity ($p < 0.05$ for interaction between eccentricity and scale, two-way ANOVA). The multiscale model exhibited an error profile matching the minimum envelope of those for the single-scale models. Thus, the multiscale model appears to optimally find reliable scales at each eccentricity.

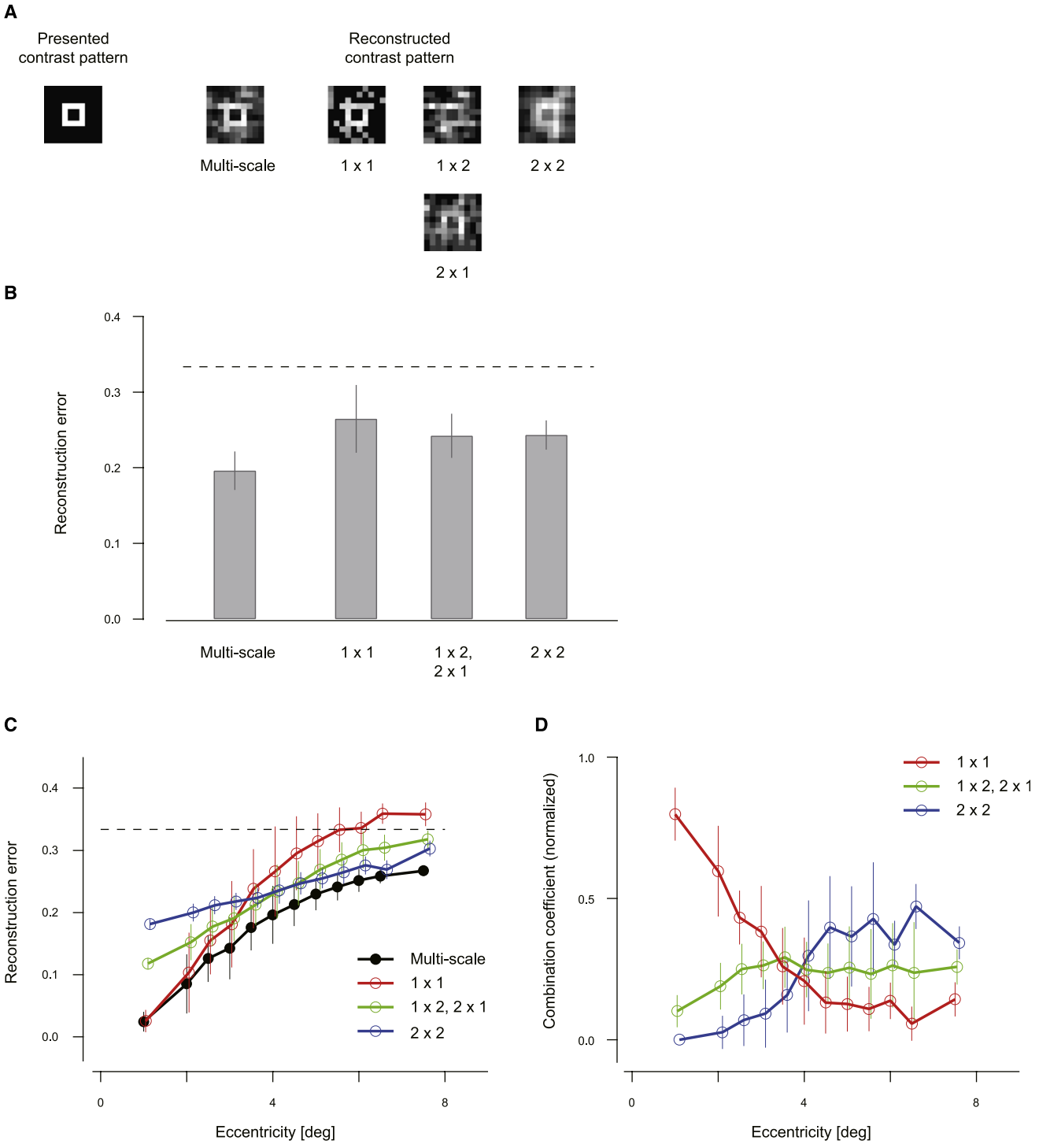


Figure 7. Advantage of Multiscale Reconstruction Model

(A) Reconstructed images. Examples from the figure image session (S1, "small frame") are shown for the multi- and single-scale reconstruction models.

(B) Reconstruction performance with entire images for the multi- and single-scale models. Reconstruction errors were displayed as in Figure 6B. The results for the 1 × 2 and 2 × 1 scales are combined.

(C) Reconstruction performance as a function of eccentricity. Patch-wise errors were averaged across all test images at each eccentricity (two subjects pooled; error bars, s.d.; dashed line, chance level).

(D) Combination coefficients in the multiscale reconstruction model. At each patch, the combination coefficients of the overlapping local decoders were grouped by scale and were normalized by their sum. The normalized combination coefficients were then averaged at each eccentricity for ten models generated by the cross-validation analysis (two subjects pooled; error bars, s.d.).

The combination coefficients of the multiscale model are summarized in Figure 7D. Consistent with the above observation, the model relied on the fine- and coarse-scale decoders for the reconstruction of the foveal and peripheral regions, respectively. These results indicate that the optimization of combination coefficients indeed found reliable local decoders at each visual field location to achieve high reconstruction performance.

Advantage of Overlapping Multiscale Bases

The image bases of different scales were useful since the scale of reliable decoders varied across eccentricity. However, it is not clear whether overlapping multiple scales contributed to the reconstruction. To address this issue, we compared the multiscale model with an "eccentricity-dependent-scale model", in which the region of the same eccentricity was tiled with image bases of a single reliable scale (Figure 8A).

We found that the reconstruction error was significantly smaller with the multiscale model than with the eccentricity-dependent-scale model (ANOVA, $p < 0.05$; Figure 8A). The performance of the multiscale model was particularly better in the middle to peripheral range of eccentricity (Figure 8B). This result is in agreement with the combination coefficients shown in Figure 7D, in which all scales had comparable values at the middle to peripheral eccentricity.

The results shown in Figure 8 were obtained using a particular spatial arrangement of nonoverlapping images bases, whose predicted contrasts were simply combined without optimized combination coefficients. We also tested modified models with slightly different spatial arrangements, and confirmed that the multiscale model outperformed these modified models, too (Figure S5). Our findings indicate that the multiscale representation at single locations indeed contributed to the reconstruction.

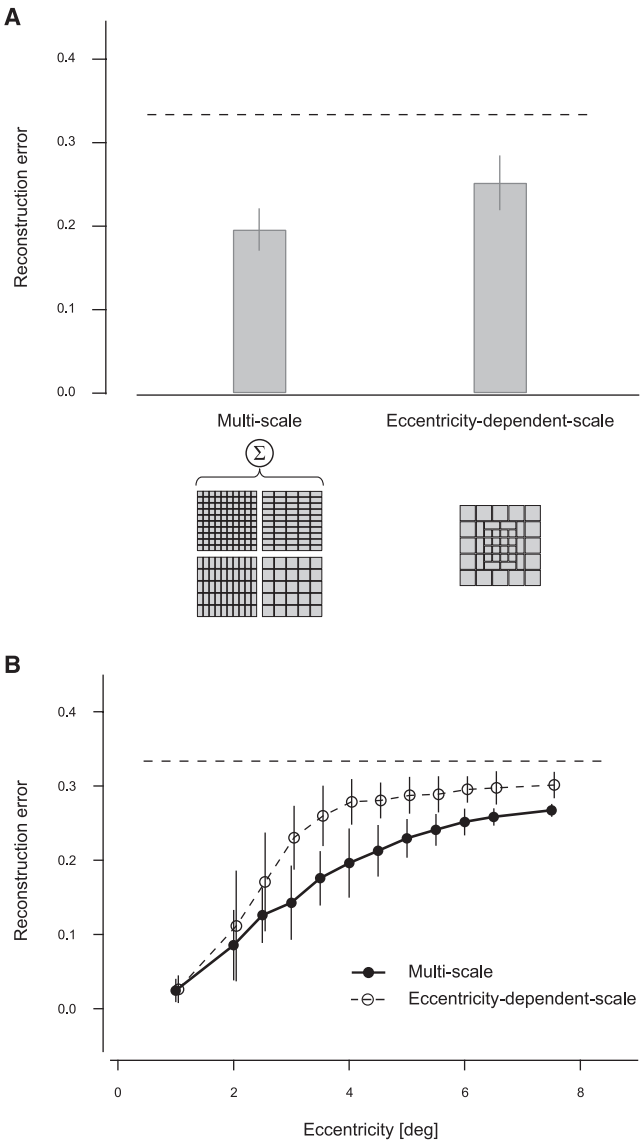


Figure 8. Advantage of Overlapping Multiscale Bases

(A) Comparison of reconstruction performance between the multiscale model and the "eccentricity-dependent-scale" model. Reconstruction errors are shown as in Figures 6B and 7B. The configurations of image bases used for this analysis are illustrated at the bottom. Image bases for the multiscale model overlapped within each scale (except 1 × 1), though the figure displays only nonoverlapping bases.

(B) Patch-wise reconstruction performance as a function of eccentricity. Results are shown as in Figure 7C.

DISCUSSION

We have shown that contrast-defined arbitrary visual images can be reconstructed from fMRI signals of the human visual cortex on a single trial basis. By combining the outputs of local decoders that predicted local contrasts of multiple scales, we were able to reconstruct a large variety of images (2^{100} possible images) using only several hundred random images to train the reconstruction model. Analyses revealed that both the multivoxel and the multiscale aspects of our method were essential to achieve the high accuracy. Our automatic method for identifying relevant neural signals uncovered information represented in correlated activity patterns, going beyond mere exploitation of known functional anatomy.

Although our primary purpose was to reconstruct visual images from brain activity, we also performed image identification analysis to quantify the accuracy (Figure 3). Analysis showed that nearly 100% correct identification was possible with a hundred image candidates and that >10% performance could be achieved even with image sets of $10^{7.4}$–$10^{10.8}$ using 6 s block-averaged data, and with image sets of $10^{5.8}$–$10^{8.5}$ using 2 s single-volume data.

Previous studies have conducted similar image identification analyses. Thirion et al. (2006) reconstructed an image for a 3 × 3 Gabor-patch array based on the retinotopy map and obtained 41%–71% accuracy of image identification with a set of six candidates (we obtained >95% accuracy with a 3 × 3 patch area in the foveal region). Kay et al. (2008) took a different approach to image identification. Instead of performing explicit reconstruction, they constructed a receptive-field model that predicted the brain activity patterns for all candidate images. Then, they identified one image whose predicted brain activity pattern was closest to the measured activity pattern. They estimated

that their model could achieve >10% correct identification with image sets of $10^{7.3}$–$10^{11.3}$ and $10^{3.5}$–$10^{10.8}$, using fMRI responses to 13 repeated presentations (= 52 s) and to a single presentation (= 4 s), respectively. Although a direct comparison with these previous studies is difficult, because of differences in stimuli, the number of trials, scan parameters, etc., the remarkably high identification performance obtained using our model represents the quality of reconstruction.

Decoding from Multivoxel Patterns

A major difference of our approach from the previous ones is that we directly computed the decoding model, instead of elaborating or inverting the encoding model. In our decoding approach, the model is optimized so as to best predict individual stimulus parameters given a multivoxel pattern while taking into account voxel correlations. In contrast, the encoding model is optimized so as to predict individual voxel responses given a stimulus without considering voxel correlations when estimating the model parameters (Kay et al., 2008; Thirion et al., 2006).

Recent imaging studies suggest that there is a better combination of population responses to decode a given visual stimulus than using a signal from the most responsive cortical location or an averaged signal over the responsive cortical locations (Chen et al., 2006; Kamitani and Tong, 2005). In particular, if signals from multiple locations are correlated, a successful decoder should optimally assign various weights, including negative ones, to each location depending on the correlation structure (Averbeck et al., 2006; Chen et al., 2006).

Consistent with this observation, our decoder using a multivoxel pattern outperformed that using a single responsive voxel or an average of responsive voxels (Figure 5A). The shuffling of training data, which removed voxel correlations, impaired the decoding performance, indicating the critical role of voxel correlation for constructing an optimal decoder (Figure 5B). Careful inspection of the weight distributions in Figures 4A and S4A indicates that a decoder trained with the original data uses both positive and negative weights, which are found at nearby locations, particularly at the middle to peripheral range of eccentricity. Additional analyses revealed that the negative weights as well as the positive weights were distributed along the retinotopic voxels (Figure S6A). Further, the magnitudes of negative weights decreased after shuffling the training data (Figure S6B), suggesting that negative weights served to exploit voxel correlation.

Although the study by Chen et al. (2006) suggested that neural activity in V1 contains significant spatial correlations that can be useful for decoding a visual stimulus, it has been unclear whether such informative correlations are present in other areas of the early visual cortex. Our analysis of individual areas (Figure 6) showed that much of the information available in V1 was represented in voxel correlations, while other areas were less dependent on them. Thus, our results suggest that the early visual cortex, particularly V1, represents the visual field not just by its ordered retinotopic mapping, but also by correlated activity patterns.

There are many possible sources of voxel correlation, in addition to stimulus-induced correlated neural activity. As the neural populations in nearby voxels are likely to be synaptically coupled, correlated fMRI signals could be spontaneously induced. Nearby voxels might also show correlations through vascular coupling. A physiological status (e.g., cardiac and respiratory noise) and an fMRI scanner condition (e.g., gradient coil heating) might also cause slow fluctuations correlated among voxels. However, they are unlikely to be major sources of the voxel correlations contributing to the reconstruction because the decoder's performance was not affected by filtering out slow components from the data (Figures S7 and S8). In addition, head motions of a subject and spatial reinterpolation during preprocessing are also unlikely to be the source, since they cannot account for the area-specific effects of the voxel correlations (Figures 6D and 6E). Further analysis will be necessary to understand the sources of voxel correlations and their contribution to the reconstruction.

Multiple Scales of Visual Representation

Our multiscale reconstruction model achieved higher reconstruction accuracy than single-scale models by combining reliable scales at each location. The reliable scales largely depended on eccentricity, which can be related to the receptive field size and the cortical magnification factor. The receptive field size of visual cortical neurons is known to increase with eccentricity (Dumoulin and Wandell, 2008; Kay et al., 2008; Kraft et al., 2005; Smith et al., 2001), and in parallel, the cortical magnification factor decreases with eccentricity (Dougherty et al., 2003; Duncan and Boynton, 2003; Engel et al., 1997). The receptive-field size for the human visual cortex was estimated at about 1°–2° at 7° eccentricity, which is near the most peripheral patch in our stimulus image, while the cortical magnification factor at 7° is about 2–3 mm/°. These estimates suggest that single voxels (3 × 3 × 3 mm) for the peripheral representation carry retinotopic information about more than a single peripheral patch, and thus are not suitable for the decoding of fine-scale (1 × 1) patches, consistent with our reconstruction results (Figure 7C). Such eccentricity-dependent changes in the scale of visual representation may partly account for the superior reconstruction by the multiscale model.

However, it should also be noted that the reconstruction model did not exclusively select a single scale at each eccentricity. At any location except the most foveal region, all scales were effectively combined to improve the reconstruction accuracy (Figure 7D). Previous studies have shown variability in receptive field size among neurons whose receptive fields overlap (De Valois et al., 1982; Hubel and Wiesel, 1968). Even though each fMRI voxel should contain numerous neurons with receptive fields of various sizes, it may be possible to extract scale-specific information by combining many voxels with a weak scale bias in each, analogous to the extraction of orientation information from coarse voxel sampling of cortical columns (Kamitani and Tong, 2005).

The multiscale reconstruction may also be linked with models of multiple spatial frequency channels. Psychophysical evidence has suggested that the human visual system uses multiple narrowly tuned spatial frequency channels to achieve broad-band sensitivity (Campbell and Robson, 1968). Channels tuned to a lower (higher) spatial frequency are assumed to have a larger (smaller) receptive field (De Valois et al., 1982), and recent human fMRI studies have reported supporting results (Dumoulin and Wandell, 2008; Kay et al., 2008; Singh et al., 2000; Smith

et al., 2001). Although our stimulus images had a fixed spatial frequency in the luminance domain (checkerboard pattern), the receptive field size associated with the tuned spatial frequency may underlie the multiscale representation. Other lines of research have suggested multifrequency channel mechanisms for contrast-defined patterns (Arsenault et al., 1999; Landy and Oruc, 2002), which could also be related to the multiple scales in our reconstruction model.

Linearity of Visual Representation

We used sequences of random images to simultaneously present binary contrast at multiple visual field locations (Figure 1B). This procedure is effective in measuring neural responses to localized stimuli if spatial linearity holds between presented visual stimuli and neural responses. The principle is similar to that of the receptive field mapping using white noise stimuli in animal electrophysiology. Previous fMRI studies have used random patterns consisting of flickering patches and showed that the amplitude of the fMRI signal evoked by a combination of the patches equals the sum of those evoked by each individual patch (Hansen et al., 2004). A general linear model can be applied to identify voxels showing significant activity in response to each patch, which provides a cortical activation map equivalent to that obtained by the conventional retinotopy mapping (Vanni et al., 2005). The linearity demonstrated by these studies may underlie the accurate decoding in our study.

However, experiments have shown that neural and behavioral responses to a localized visual stimulus are affected by surrounding stimuli. Such phenomena known as contextual effects (Kapadia et al., 1995; Meng et al., 2005; Sasaki and Watanabe, 2004; Zipser et al., 1996) could compromise the linearity assumption. However, the random patterns rarely contain specific configurations inducing contextual effects enough to bias the training of the local decoders. Thus, the influences from contextual effects may be negligible, and predictions from local decoders are largely based on fMRI signals corresponding to the local state of the visual stimulus.

We also assumed linearity in the image reconstruction model. An entire visual image was represented by a linear superposition of local image bases of multiple scales. Our approach draws an idea from previous theoretical studies modeling visual images by a linear summation of overcomplete basis functions that are spatially localized with various scales (Olshausen and Field, 1996). Our successful reconstruction supports the linear representation model of visual images, though elaborate models with non-linearity might further improve the reconstruction performance.

Modular Decoding and Its Applications

Our approach provides a general procedure to deal with complex perceptual experience consisting of numerous possible states by using multiple decoders as modules. If a perceptual state can be expressed by a combination of elemental features, a modular decoder can be trained for each feature with a small number of data, but their combination could predict numerous states including those that have never been experienced. Similar modular methods have been proposed for constructing ''encoding'' models that predict brain activity induced by complex stimuli or mental states, too (Kay et al. 2008; Mitchell et al., 2008).

Although we focused here on the reconstruction of contrast patterns, our approach could be extended to reconstruct visual images defined by other features, such as color, motion, texture, and binocular disparity. Likewise, motor functions may also be dealt with using our approach. A large variety of motor actions could be described by a combination of putative modules (Poggio and Bizzi, 2004). Thus, the modular decoding approach may greatly improve the flexibility of prediction, which could also expand the capacity of neural prosthetics or brain-machine interfaces (Donoghue, 2002; Wolpaw and McFarland, 2004).

More interesting are attempts to reconstruct subjective states that are elicited without sensory stimulation, such as visual imagery, illusions, and dreams. Several studies have suggested that these subjective percepts occur in the early visual cortex (Kosslyn et al., 1995), consistent with the retinotopy map (Meng et al., 2005; Sasaki and Watanabe, 2004; Thirion et al., 2006). Of particular interest is to examine if such subjective percepts share the same representation as stimulus-evoked percepts (Kamitani and Tong, 2005, 2006; Haynes and Rees, 2005). One could address this issue by attempting to reconstruct a subjective state using a reconstruction model trained with physical stimuli. The combination of elemental decoders could even reveal subjective states that have never been experienced with sensory stimulation. Reconstruction performance can also be compared among cortical areas and reconstruction models. Thus, our approach could provide valuable insights into the complexity of perceptual experience and its neural substrates.

EXPERIMENTAL PROCEDURES

Subjects

We first screened four subjects for head motion in preliminary scans, and two of them (male adults with normal or corrected-to-normal visual acuity) who showed the least head motion underwent the full experimental procedure. The subjects gave written informed consent. The study was approved by the Ethics Committee of ATR and National Institute for Physiological Sciences.

Visual Stimulus and Experimental Design

Visual stimuli were rear-projected onto a screen placed in the scanner bore using a gamma-corrected LCD projector.

We had three types of experimental session to measure the fMRI responses of the visual cortex: (1) the random image session, (2) the figure image session, and (3) the conventional retinotopy mapping session.

In the random image session, each run contained 22 stimulus blocks. Each stimulus block was 6 s long followed by a 6 s intervening rest period. Extra rest periods were added at the beginning (28 s) and at the end (12 s) of each run. In each stimulus block, an image consisting of 12 × 12 small square patches (1.15° × 1.15° each) was presented on a gray background with a fixation spot. Each patch was either a flickering checkerboard (spatial frequency, 1.74 cycles/°; temporal frequency, 6 Hz) or a homogeneous gray area, with equal probability. Each stimulus block had a different spatial arrangement of random patches. To avoid the effects of the stimulus frame, the central 10 × 10 area was used for analysis. Twenty runs were repeated, and a total of 440 different random patterns were presented to each subject.

In the figure image session, each run had ten stimulus blocks. Each stimulus block was 12 s long followed by a 12 s intervening rest period. Extra rest periods were included, as in the random image session. Stimulus images consisted of flickering checkerboard patches as in the random image session, but formed geometric shapes (“square,” “small frame,” “large frame,” “plus,” and “X”) or alphabet letters (“n,” “e,” “u,” “r,” and “o”). In each run, five geometric

shapes or five alphabets were presented, and each image was repeated twice. Subject S1 performed four geometric-shape runs and four alphabet runs, while S2 performed four geometric-shape runs and three alphabet runs.

In these sessions, subjects viewed the stimulus sequence while maintaining fixation. To help subjects suppress eye blinks and firmly fixate the eyes, the color of the fixation spot changed from white to red 2 s before each stimulus block started. To ensure alertness, subjects were instructed to detect the color change of the fixation (red to green, 100 ms) that occurred after a random interval of 3–5 s from the beginning of each stimulus block.

The retinotopy mapping session followed the conventional procedure (Engel et al., 1994; Sereno et al., 1995) using a rotating wedge and an expanding ring of flickering checkerboard. The data were used to delineate the borders between visual cortical areas, and to identify the retinotopy map on the flattened cortical surfaces. Note that the retinotopic mapping was only used to relate the conventional retinotopy and the location of voxels selected by our method.

MRI Acquisition

Preliminary experiments were performed using 3.0-Tesla Siemens MAGNETOM Allegra located at National Institute for Physiological Sciences. MRI data for the presented results were all obtained using a 3.0-Tesla Siemens MAGNETOM Trio A Tim scanner located at the ATR Brain Activity Imaging Center. An interleaved T2*-weighted gradient-echo echo-planar imaging (EPI) scan was performed to acquire functional images to cover the entire occipital lobe (TR, 2000 ms; TE, 30 ms; flip angle, 80°; FOV, 192 × 192 mm; voxel size, 3 × 3 x 3 mm; slice gap, 0 mm; number of slices, 30). T2-weighted turbo spin echo images were scanned to acquire high-resolution anatomical images of the same slices used for the EPI (TR, 6000 ms; TE, 57 ms; flip angle, 90°; FOV, 192 × 192 mm; voxel size, 0.75 × 0.75 × 3.0 mm). T1-weighted magnetization-prepared rapid-acquisition gradient-echo (MP-RAGE) fine-structural images of the whole-head were also acquired (TR, 2250 ms; TE, 2.98 or 3.06 ms; TI, 900 ms; flip angle, 9°; field of view, 256 × 256 mm; voxel size, 1.0 × 1.0 × 1.0 mm).

MRI Data Preprocessing

The first 8 s scans of each run were discarded to avoid instability of the MRI scanner. The acquired fMRI data underwent slice-timing correction and three-dimensional motion correction by SPM2 (http://www.fil.ion.ucl.ac.uk/spm). The data were then coregistered to the within-session high-resolution anatomical image of the same slices used for EPI and subsequently to the whole-head high-resolution anatomical image. The coregistered data were then reinterpolated by 3 × 3 × 3 mm voxels. The retinotopy session data were transformed to the Talairach coordinates and the visual cortical borders were delineated on the flattened cortical surfaces using Brain Voyager 2000 (http://www.brainvoyager.com). The voxel coordinates around the gray-white matter boundary in V1–V4 were identified and transformed back into the original coordinates of the EPI images. After voxels of extremely low signal amplitudes were removed, ~3000 voxels were selected in V1–V4 (subject S1, 3003 voxels; S2, 3258 voxels). Most of the reconstruction analyses were done using V1 and V2 voxels (S1, 1396 voxels; S2, 1550 voxels). For the analysis of individual areas, the following numbers of voxels were identified: V1, 797; V2, 820; V3, 779 voxels for S1, and V1, 903; V2, 902; V3, 913 voxel for S2. Voxels near the area border were included in both areas.

The fMRI data then underwent linear trend removal within each run. Amplitude normalization relative to the mean amplitude of the first 20 s rest period in each run was performed to minimize the baseline difference across runs. The fMRI signals of each voxel were averaged within each stimulus block after shifting the data by 4 s to compensate for hemodynamic delays.

Labeling of fMRI Data

Each fMRI data sample was labeled by the mean contrast values of local image elements in the corresponding stimulus image. Local image elements were 1 × 1, 1 × 2, 2 × 1, and 2 × 2 patch areas covering the entire 10 × 10 patch area with overlaps (a total of 361 elements; 1 × 1, 100; 1 × 2, 90; 2 × 1, 90; 2 × 2, 81). The mean contrast value of each local image element was defined as the number of flickering patches divided by the total number of patches (1 × 1, [0 or 1]; 1 × 2 and 2 × 1, [0, 0.5, or 1]; 2 × 2, [0, 0.25, 0.5, 0.75, or 1]).

Training of Local Decoders

Local decoders were defined to predict the mean contrast of each local image element. They were individually trained with fMRI data and the corresponding class labels representing the mean contrast values. Each local decoder consisted of a multi-class classifier, which classified fMRI data samples into the classes defined by the mean contrast values. We could use a regression model that gives a continuous output, but we chose to use the classification model simply because our preliminary study showed better performance with classification than with regression.

Our classification model is based on multinomial logistic regression (Bishop, 2006), in which each contrast class has a linear discriminant function that calculates the weighted sum of the inputs (voxel values). Its output is then transformed into the probability for the contrast class given the inputs. The discriminant function for contrast class k in a local decoder is expressed as,

$$y_{\mathbf{w}_k}(\mathbf{r}) = \sum_{d}^{D} w_k^d r^d + w_k^0,$$

where w_k^d is a weight parameter for voxel d and contrast class k, r^d is the fMRI signal of voxel d, w_k^0 is the bias, and D is the number of voxels. The probability that an fMRI signal pattern $\mathbf{r} = [r^1, r^2, r^D]^T$ (T, transpose) belongs to the contrast class k is defined using the softmax function,

$$p_{\mathbf{w}}(k|\mathbf{r}) = \frac{\exp\left[y_{\mathbf{w}_k}(\mathbf{r})\right]}{\sum_{j}^{K} \exp\left[y_{\mathbf{w}_j}(\mathbf{r})\right]},$$

where K is the number of the contrast classes. The predicted contrast class for mth local image element, $C_m(\mathbf{r})$ is chosen as the contrast class with the highest probability. Note that although the statistics terminology calls this type of model multinomial logistic "regression", it performs classification rather than regression in the sense that the output is a categorical variable.

In conventional multinomial logistic regression, the weight parameters are determined by finding the values that maximize the likelihood function of the weight parameters given a training data set,

$$p_{\mathbf{w}}(\mathbf{S}|\mathbf{w}_1, \cdots, \mathbf{w}_k) = \prod_{n}^{N} \prod_{k}^{K} p_{\mathbf{w}}(k|\mathbf{r}_n)^{s_{nk}},$$

where $\mathbf{S}$ represents a class label matrix whose element s_{nk} is 1 if the trial n corresponds to the contrast class k otherwise 0, $\mathbf{w}_k$ is the weight vector for contrast class k including the bias term ((D + 1) × vector), and N is the number of trials.

In this study, we adopted a full-Bayesian approach to the estimation of weight parameters ("sparse logistic regression," Yamashita et al., 2008). The above likelihood function was combined with a prior distribution for each weight to obtain the posterior distribution. Weight parameters were estimated by taking the expectation of the posterior distribution for each weight.

The prior distribution of a weight parameter is described by a zero-mean normal distribution with a variance, whose inverse is treated as a hyperparameter,

$$p\left(w_k^d|\alpha_k^d\right) = \mathrm{N}\left(0, \frac{1}{\alpha_k^d}\right),$$

where N represents a normal distribution, and α_k^d is the hyperparameter denoting the inverse of the variance, or precision, of the weight value for voxel d and contrast class k. The hyperparameter α_k^d is also treated as a random variable, whose distribution is defined by,

$$p\left(\alpha_k^d\right) = \frac{1}{\alpha_k^d}.$$

These prior distributions are known to lead to "sparse estimation" in which only a small number of parameters have nonzero values and the remaining parameters are estimated to be zero (Tipping, 2001). Thus, the prior distributions implement the assumption that only a small number of voxels are relevant for the decoding of each local image element. This sparseness assumption may be validated by the fact that a spatially localized visual stimulus gives rise to neural activity only in small regions of the early visual cortex. The sparse

parameter estimation could avoid overfitting to noisy training data by pruning irrelevant voxels (Bishop, 2006), and thereby help to achieve high generalization (test) performance (Yamashita et al., 2008). The number of remaining nonzero voxels is shown in Figure S9.

Since the direct evaluation of the posterior distribution is analytically intractable, we used a variational Bayesian method to approximate the distribution. The algorithm for the parameter estimation is described in Supplemental Data.

Combination of Local Decoders

The outputs of the local decoders were combined by a linear model of the corresponding local image elements,

$$\hat{I}(x|\mathbf{r}) = \sum_{m}^{M} \lambda_m C_m(\mathbf{r})\phi_m(x),$$

where $\phi_m(x)$ represents a local image element, or a basis, ($\phi_m(x) = 1$ if location x is contained in the area of the local image element, otherwise $\phi_m(x) = 0$), $C_m(\mathbf{r})$ is the predicted contrast, and λ_m is the combination coefficient.

Combination coefficients, λ_m, were determined by the least square method using a training data set. We divided training data into subgroups, and the local contrasts for each subgroup were predicted by the decoders trained with the other subgroups. After calculating the local contrasts, $C_m(\mathbf{r})$, for all training samples, optimal combination coefficients were obtained by finding the non-negative values that minimize the sum of the square errors between the presented and the reconstructed images. The final reconstruction model was obtained by integrating the combination coefficients and the local decoders that were retrained using all training samples.

Evaluation of Performance

The trained reconstruction model was tested with independent samples. We performed two types of reconstruction tests. First, to obtain a quantitative and unbiased evaluation, we conducted cross-validation analysis using the samples in the random image session. Second, to illustrate the quality of reconstructed images, the model obtained from the random image session was used to reconstruct the images presented in the figure image session.

In the cross-validation analysis, the 20 runs in the random image session were divided into ten groups (two runs per group), and the reconstruction model was trained with nine groups and tested with the remaining group. This procedure was repeated until all groups were tested (10-fold cross-validation). In each step of cross-validation, the training data set (nine groups, or 18 runs) was divided into one versus eight subgroups to obtain combination coefficients as described above. The combination coefficients and the local decoders that were retrained using all nine groups were integrated into a reconstruction model.

For the reconstruction of the images in the figure image session, all 20 runs in the random image session were used as a training data set. They were divided into one versus nine subgroups to obtain combination coefficients as described above. The combination coefficients and the local decoders that were retrained using all ten groups were integrated into a reconstruction model.

SUPPLEMENTAL DATA

The Supplemental Data can be found with this article online at http://www.neuron.org/supplemental/S0896-6273(08)00958-6.

ACKNOWLEDGMENTS

The authors thank M. Kawato and K. Toyama for helpful comments; A. Harner and S. Murata for technical assistance; and T. Beck and Y. Yamada for manuscript editing. This research was supported in part by the SRPBS, MEXT, the NICT-KARC, the Nissan Science Foundation, and the SCOPE, SOUMU.

Accepted: November 4, 2008
Published: December 10, 2008

REFERENCES

Arsenault, A.S., Wilkinson, F., and Kingdom, F.A. (1999). Modulation frequency and orientation tuning of second-order texture mechanisms. J. Opt. Soc. Am. A Opt. Image Sci. Vis. *16*, 427–435.

Averbeck, B.B., Latham, P.E., and Pouget, A. (2006). Neural correlations, population coding and computation. Nat. Rev. Neurosci. *7*, 358–366.

Bishop, C.M. (2006). Pattern Recognition and Machine Learning (New York: Springer).

Campbell, F.W., and Robson, J.G. (1968). Application of Fourier analysis to the visibility of gratings. J. Physiol. *197*, 551–566.

Chen, Y., Geisler, W.S., and Seidemann, E. (2006). Optimal decoding of correlated neural population responses in the primate visual cortex. Nat. Neurosci. *9*, 1412–1420.

Cox, D.D., and Savoy, R.L. (2003). Functional magnetic resonance imaging (fMRI) ''brain reading'': detecting and classifying distributed patterns of fMRI activity in human visual cortex. Neuroimage *19*, 261–270.

De Valois, R.L., Albrecht, D.G., and Thorell, L.G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. Vision Res. *22*, 545–559.

Donoghue, J.P. (2002). Connecting cortex to machines: recent advances in brain interfaces. Nat. Neurosci. *5*, 1085–1088.

Dougherty, R.F., Koch, V.M., Brewer, A.A., Fischer, B., Modersitzki, J., and Wandell, B.A. (2003). Visual field representations and locations of visual areas V1/2/3 in human visual cortex. J. Vis. *3*, 586–598.

Dumoulin, S.O., and Wandell, B.A. (2008). Population receptive field estimates in human visual cortex. Neuroimage *39*, 647–660.

Duncan, R.O., and Boynton, G.M. (2003). Cortical magnification within human primary visual cortex correlates with acuity thresholds. Neuron *38*, 659–671.

Engel, S.A., Rumelhart, D.E., Wandell, B.A., Lee, A.T., Glover, G.H., Chichilnisky, E.J., and Shadlen, M.N. (1994). fMRI of human visual cortex. Nature *369*, 525.

Engel, S.A., Glover, G.H., and Wandell, B.A. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. Cereb. Cortex *7*, 181–192.

Hansen, K.A., David, S.V., and Gallant, J.L. (2004). Parametric reverse correlation reveals spatial linearity of retinotopic human V1 BOLD response. Neuroimage *23*, 233–241.

Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., and Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science *293*, 2425–2430.

Haynes, J.D., and Rees, G. (2005). Predicting the orientation of invisible stimuli from activity in human primary visual cortex. Nat. Neurosci. *8*, 686–691.

Hubel, D.H., and Wiesel, T.N. (1968). Receptive fields and functional architecture of monkey striate cortex. J. Physiol. *195*, 215–243.

Kamitani, Y., and Tong, F. (2005). Decoding the visual and subjective contents of the human brain. Nat. Neurosci. *8*, 679–685.

Kamitani, Y., and Tong, F. (2006). Decoding seen and attended motion directions from activity in the human visual cortex. Curr. Biol. *16*, 1096–1102.

Kapadia, M.K., Ito, M., Gilbert, C.D., and Westheimer, G. (1995). Improvement in visual sensitivity by changes in local context: parallel studies in human observers and in V1 of alert monkeys. Neuron *15*, 843–856.

Kay, K.N., Naselaris, T., Prenger, R.J., and Gallant, J.L. (2008). Identifying natural images from human brain activity. Nature *452*, 352–355.

Kosslyn, S.M., Thompson, W.L., Kim, I.J., and Alpert, N.M. (1995). Topographical representations of mental images in primary visual cortex. Nature *378*, 496–498.

Kraft, A., Schira, M.M., Hagendorf, H., Schmidt, S., Olma, M., and Brandt, S.A. (2005). fMRI localizer technique: efficient acquisition and functional properties of single retinotopic positions in the human visual cortex. Neuroimage *28*, 453–463.

Landy, M.S., and Oruc, I. (2002). Properties of second-order spatial frequency channels. Vision Res. *42*, 2311–2329.

Meng, M., Remus, D.A., and Tong, F. (2005). Filling-in of visual phantoms in the human brain. Nat. Neurosci. *8*, 1248–1254.

Mitchell, T.M., Shinkareva, S.V., Carlson, A., Chang, K.M., Malave, V.L., Mason, R.A., and Just, M.A. (2008). Predicting human brain activity associated with the meanings of nouns. Science *320*, 1191–1195.

Olshausen, B.A., and Field, D.J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature *381*, 607–609.

Poggio, T., and Bizzi, E. (2004). Generalization in vision and motor control. Nature *431*, 768–774.

Sasaki, Y., and Watanabe, T. (2004). The primary visual cortex fills in color. Proc. Natl. Acad. Sci. USA *101*, 18251–18256.

Sereno, M.I., Dale, A.M., Reppas, J.B., Kwong, K.K., Belliveau, J.W., Brady, T.J., Rosen, B.R., and Tootell, R.B. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science *268*, 889–893.

Shmuel, A., Yacoub, E., Chaimow, D., Logothetis, N.K., and Ugurbil, K. (2007). Spatio-temporal point-spread function of fMRI signal in human gray matter at 7 Tesla. Neuroimage *35*, 539–552.

Singh, K.D., Smith, A.T., and Greenlee, M.W. (2000). Spatiotemporal frequency and direction sensitivities of human visual areas measured using fMRI. Neuroimage *12*, 550–564.

Smith, A.T., Singh, K.D., Williams, A.L., and Greenlee, M.W. (2001). Estimating receptive field size from fMRI data in human striate and extrastriate visual cortex. Cereb. Cortex *11*, 1182–1190.

Stanley, G.B., Li, F.F., and Dan, Y. (1999). Reconstruction of natural scenes from ensemble responses in the lateral geniculate nucleus. J. Neurosci. *19*, 8036–8042.

Thirion, B., Duchesnay, E., Hubbard, E., Dubois, J., Poline, J.B., Lebihan, D., and Dehaene, S. (2006). Inverse retinotopy: inferring the visual content of images from brain activation patterns. Neuroimage *33*, 1104–1116.

Tipping, M. (2001). Sparse Bayesian learning and the relevance vector machine. J. Mach. Learn. Res. *1*, 211–244.

Vanni, S., Henriksson, L., and James, A.C. (2005). Multifocal fMRI mapping of visual cortical areas. Neuroimage *27*, 95–105.

Wolpaw, J.R., and McFarland, D.J. (2004). Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. Proc. Natl. Acad. Sci. USA *101*, 17849–17854.

Yamashita, O., Sato, M.A., Yoshioka, T., Tong, F., and Kamitani, Y. (2008). Sparse estimation automatically selects voxels relevant for the decoding of fMRI activity patterns. Neuroimage *42*, 1414–1429.

Zipser, K., Lamme, V.A., and Schiller, P.H. (1996). Contextual modulation in primary visual cortex. J. Neurosci. *16*, 7376–7389.