

# Timing, Timing, Timing: Fast Decoding of Object Information from Intracranial Field Potentials in Human Visual Cortex

Hesheng Liu,<sup>1,2</sup> Yigal Agam,<sup>1</sup> Joseph R. Madsen,<sup>3</sup> and Gabriel Kreiman<sup>1,4,5,\*</sup>

<sup>1</sup>Department of Neuroscience and Ophthalmology, Children's Hospital Boston, Harvard Medical School, Boston, MA 02115, USA

<sup>2</sup>Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Boston, MA 02115, USA

<sup>3</sup>Department of Neurosurgery, Children's Hospital Boston, Harvard Medical School, Boston, MA 02129, USA

<sup>4</sup>Center for Brain Science, Harvard University, Boston, MA 02115, USA

<sup>5</sup>Swartz Center for Theoretical Neuroscience, Harvard University, Boston, MA 02115, USA

\*Correspondence: [gabriel.kreiman@tch.harvard.edu](mailto:gabriel.kreiman@tch.harvard.edu)

DOI 10.1016/j.neuron.2009.02.025

## SUMMARY

The difficulty of visual recognition stems from the need to achieve high selectivity while maintaining robustness to object transformations within hundreds of milliseconds. Theories of visual recognition differ in whether the neuronal circuits invoke recurrent feedback connections or not. The timing of neurophysiological responses in visual cortex plays a key role in distinguishing between bottom-up and top-down theories. Here, we quantified at millisecond resolution the amount of visual information conveyed by intracranial field potentials from 912 electrodes in 11 human subjects. We could decode object category information from human visual cortex in single trials as early as 100 ms poststimulus. Decoding performance was robust to depth rotation and scale changes. The results suggest that physiological activity in the temporal lobe can account for key properties of visual recognition. The fast decoding in single trials is compatible with feedforward theories and provides strong constraints for computational models of human vision.

## INTRODUCTION

Humans and other primates can recognize objects very rapidly, even when these objects are presented in complex scenes and have undergone transformations in position, size, color, illumination, and rotation (Potter and Levy, 1969). The robustness of visual perception to image changes is thought to arise gradually within the hierarchical architecture of the ventral visual cortex, in which object representations become increasingly discriminative and invariant (Logothetis and Sheinberg, 1996; Rolls, 1991; Tanaka, 1996). The highest stages of the temporal lobe show key selectivity and invariance properties that could underlie perceptual recognition as shown by convergent evidence from monkey electrophysiology (Desimone et al., 1984; Hung et al., 2005; Logothetis and Sheinberg, 1996; Perrett et al., 1992; Tanaka, 1996; Tsao

et al., 2006), monkey functional imaging (Tsao et al., 2006), lesion studies in monkeys (Dean, 1976; Holmes and Gross, 1984), human functional imaging (Grill-Spector and Malach, 2004; Haxby et al., 2001; Kanwisher et al., 1997), human electrophysiology (Kraskov et al., 2007; Kreiman et al., 2000; McCarthy et al., 1999; Quiroga et al., 2005), and neurological studies in humans (Damasio et al., 1990; De Renzi, 2000; Tippett et al., 1996).

Theories of visual recognition can be largely divided into two main groups: (1) theories advocating top-down influences and long recurrent feedback connections (e.g., Bullier, 2001; Garrido et al., 2007; Hochstein and Ahissar, 2002) and (2) theories relying on a bottom-up approach and largely dependent on feedforward processing steps (e.g., Fukushima, 1980; Riesenhuber and Poggio, 1999; Rolls, 1991; VanRullen and Thorpe, 2002). Although ultimately both bottom-up and top-down influences are likely to play significant roles in visual object recognition, the relative contribution of bottom-up versus top-down processes in different aspects of visual object recognition remains unclear. The location and timing of responses elicited in visual cortex constitute important constraints to quantify the role of feedforward and feedback signals.

In macaque monkeys, selective responses can be elicited in inferior temporal cortex ~100 ms after stimulus onset (e.g., Hung et al., 2005; Keyser et al., 2001; Perrett et al., 1992; Richmond et al., 1983). Less is known about the latency and location of responses in human visual cortex. Scalp electroencephalographic (EEG) recordings suggest that complex visual categorization tasks (distinguishing the presence of an animal in a natural scene) can trigger differential signals ~150 ms after stimulus onset (Thorpe et al., 1996), but the localization of these signals in the brain is not clear (see also Johnson and Olshausen, 2005). Face-selective responses at ~200 ms after stimulus onset have been reported in field potentials recorded from the human temporal lobe averaged over tens to hundreds of trials (McCarthy et al., 1999). Single-neuron recordings in the human medial temporal lobe (areas that are not exclusively visual including the hippocampus, entorhinal cortex, amygdala, and parahippocampal gyrus) yield latencies ranging from 200 to 500 ms (Kreiman et al., 2000; Mormann et al., 2008).

Although there is no hard threshold to separate bottom-up from top-down signals, several pieces of evidence suggest that

after ~200 ms there is likely to be prominent impact from back-projections (Garrido et al., 2007; Schmolesky et al., 1998; Tomita et al., 1999). In order to unravel the neuronal circuits and mechanisms that give rise to selective, robust, and fast visual pattern recognition, it is important to combine location information with a characterization of the dynamics of the physiological events involved in object processing in single trials. In this study, we used neurophysiological recordings to directly quantify the degree of selectivity, invariance, and speed of visual processing in the human visual cortex. We recorded intracranial field potentials from 912 subdural electrodes implanted in 11 subjects for evaluation of surgical approaches to alleviate resilient forms of epilepsy (Engel et al., 2005; Kreiman, 2007; Ojemann, 1997). We observed that physiological responses in the occipital and temporal lobes, most prominently in the inferior occipital gyrus, inferior temporal cortex, fusiform gyrus, and parahippocampal gyrus, showed strong selectivity to complex objects while maintaining robustness to object transformations. The robust and selective responses were clear even in single trials and within ~100 ms after stimulus onset.

## RESULTS

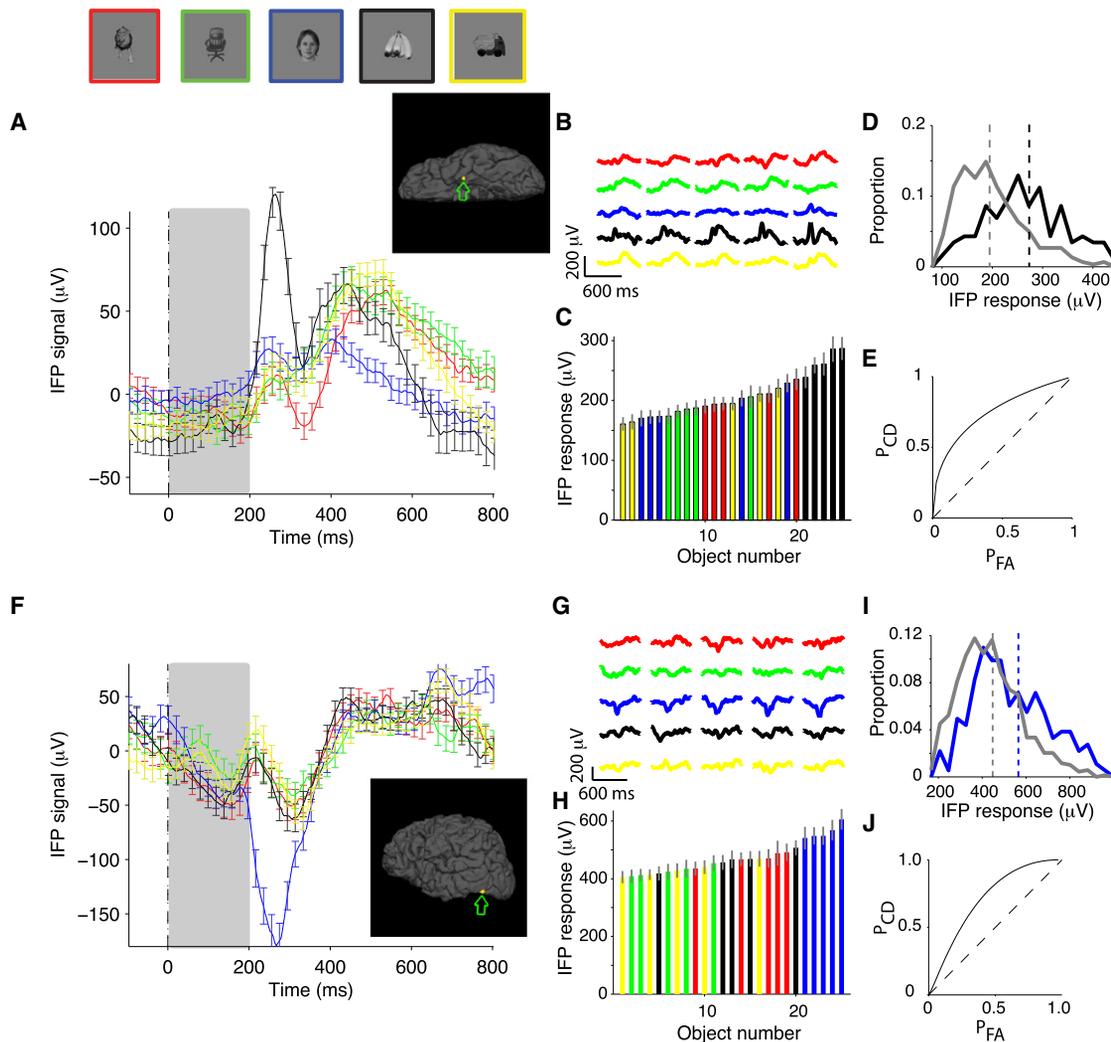
### Object Selectivity in Human Visual Cortex

Eleven subjects (6 male, 9 right-handed, 12–34 years old) were presented with digital photographs of objects belonging to five categories (“animals,” “chairs,” “human faces,” “fruits,” and “vehicles”) while recording electrophysiological activity from 48 to 126 intracranial electrodes implanted to localize epileptic seizure foci (Experimental Procedures). There were five different exemplars per category (see Figure S1B available online). Images were grayscale and contrast normalized, and were presented for 200 ms in pseudorandom order, with an interval between images of 600 ms (Figure S1A), while subjects performed a one-back task. The physiological responses from two example recording sites are shown in Figure 1. An electrode located in the left temporal pole (Talairach coordinates: –27.0, –11.8, –35.3; area number 43 in Table S2) showed differential activity across object categories, with an enhanced response to images from the fruits category (Figure 1A, black trace). To quantify the degree of selectivity, we defined the intracranial field potential (IFP) response as the range of the IFP signal, that is,  $\max(\text{IFP}) - \min(\text{IFP})$ , in a time window from 50 to 300 ms after stimulus onset. The time window from 50 to 300 ms was chosen so as to account for the approximate latency of the neural response (Richmond et al., 1983; Schmolesky et al., 1998) while remaining within the average saccade time (Rayner, 1998). We note that in many cases the selectivity extended well past the 300 ms boundary (e.g., see Figures S1E and S1F for two particularly clear examples). Throughout the manuscript, we focus on the early part of the physiological responses (see Discussion). We also quantitatively compared different time windows and other possible definitions of the IFP response (Figure S3; Supplemental Data). In the example in Figure 1A, a one-way analysis of variance (ANOVA) across categories on the IFP responses yielded  $p < 10^{-6}$  (Experimental Procedures). The selective response was consistent across repetitions ( $z$  score = 3.2; Supplemental Data) and also across the different exemplars for each category (Figure 1B).

When we ranked the objects according to the IFP response, the five best objects were the five exemplars from the fruits category (Figure 1C; bootstrap  $p < 10^{-5}$ ; Experimental Procedures). Therefore, the category selectivity in this electrode cannot be ascribed to responses to only one or a few particular exemplar objects. Another example electrode is shown in Figures 1F–1J. This electrode, located in the left inferior occipital gyrus (Talairach coordinates: –48.8, –69.1, –11.8; area number 14 in Table S2) in a different subject, showed stronger responses to images from the human faces category (Figure 1F, blue trace). More examples are shown in Figures S1D–S1G.

Although many studies have focused on analyzing the average responses across multiple repetitions, the brain needs to discriminate object information in real time. This imposes a strict criterion for selectivity. We used a statistical learning approach (Bishop, 1995; Hung et al., 2005) to quantify whether the information from IFPs available in single trials was sufficient to distinguish one category from the other categories (Experimental Procedures). Figure 1D shows the distribution of the IFP responses for the trials where a fruit was presented (black trace) versus all the other trials (gray trace), and Figure 1E shows the corresponding receiver operating characteristic (ROC) curve (Green and Swets, 1966; Kreiman et al., 2000). We used a support vector machine (SVM) classifier with a linear kernel to compute the proportion of single trials in which we could detect the preferred category (in this example, fruits). Other ways of defining selectivity, including a one-way ANOVA on the IFP responses (Kreiman et al., 2000), a point-by-point ANOVA (Thorpe et al., 1996), and the use of other statistical classifiers (Bishop, 1995), yielded similar results (Supplemental Data). We separated the repetitions used to train the classifier and those used to test its performance to avoid overfitting. The output of the classifier is a performance value (labeled “classification performance” throughout the text) that achieves 100% if all test trials are correctly classified and 50% for chance levels. In the examples in Figures 1A and 1F, we could decode the preferred object categories (fruits and human faces, respectively) from the IFP responses in single trials with performance levels of  $69\% \pm 6\%$  and  $65\% \pm 5\%$ , respectively.

We recorded data from 912 electrodes in 11 subjects (see Table S1 for the recording site location distribution). To determine whether a given classification performance value was significantly different from chance levels, we used a permutation procedure whereby we randomly shuffled the object category labels (Supplemental Data). Based on this bootstrap procedure, we used a threshold of 3 standard deviations from the mean based on the null hypothesis (Figure S2A). We observed selective responses in at least one electrode in all the subjects. Most of the selective electrodes (67%) responded preferentially to one rather than multiple object categories (Figure S2B). Across all the electrodes, we observed selective responses to the five different object categories that were tested (Figure S2C), expanding on previous IFP reports describing mostly selectivity to faces (McCarthy et al., 1999). The distribution of selective responses across categories was not uniform: there were more electrodes responding to the human faces category than to any of the other four categories (Figure S2C). We note that the observation of an electrode that shows selectivity for a particular object



**Figure 1. Selectivity in Intracranial Field Potential Recordings from the Human Cortex**

(A) Example showing the responses of an electrode in the left temporal pole (Talairach coordinates:  $-27.0, -11.8, -35.3$ ; see inset depicting the electrode position). The plot shows the average IFP response to each of five object categories: “animals” (red), “chairs” (green), “human faces” (blue), “fruits” (black), and “vehicles” (yellow). Above the plot, we show one exemplar object from each category (but the data correspond to the average over five exemplars; see Figure S1B for images of the other object exemplars). The stimulus was presented for 200 ms (gray rectangle) and the responses are aligned to stimulus onset at  $t = 0$ . The error bars denote SEM and are only shown in one of every ten time points for clarity ( $n = 158, 150, 165, 149, 159$ ).

(B) Average IFP signal (initial 600 ms) for each of the 25 exemplar objects. Responses are colored based on object category (see [A]) and are aligned to stimulus onset.

(C) The IFP response was defined as the range of the signal ( $\max(\text{IFP}) - \min(\text{IFP})$ ) in the 50–300 ms interval post-stimulus onset (see Figure S3 and Supplemental Data for other IFP response definitions). A one-way ANOVA on the IFP response yielded  $p < 10^{-5}$ . Average IFP response ( $\pm$ SEM) to each of the 25 exemplar objects. The responses are color coded based on object category and the responses are sorted by their magnitude. Note that the five images containing “fruits” yielded the top five responses (bootstrap  $p < 10^{-5}$ ; Experimental Procedures).

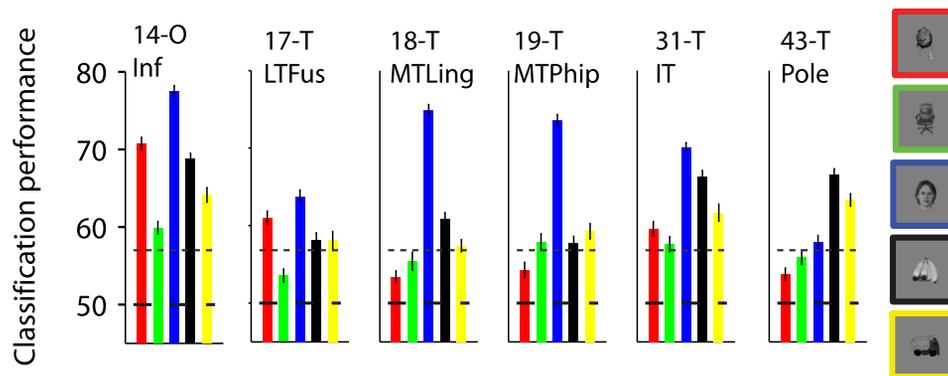
(D) Distribution of IFP responses for objects in the preferred category (“fruits,” black) and all the other objects (gray). The vertical dashed lines denote the mean of the distributions.

(E) ROC curve indicating the probability of correctly identifying the object category *in single trials* (Experimental Procedures). The departure from the diagonal (dashed line) indicates increased correct detection ( $P_{CD}$ ) for a given probability of false alarms ( $P_{FA}$ ). The classification performance for this electrode was  $69\% \pm 6\%$  (see Hung et al., 2005 and Experimental Procedures).

(F–J) Another example showing the responses of an electrode located in the left inferior occipital gyrus in a different subject (Talairach coordinates:  $-48.8, -69.1, -11.8$ ; see inset depicting the electrode position). The format and conventions are the same as in (A)–(E) (here  $n = 189, 167, 202, 194, 198$ ). The classification performance for this electrode was  $65\% \pm 5\%$ .

category does not necessarily imply the existence of a “brain area” devoted exclusively to that category (see Discussion). The classification performance values ranged from 58% to

82% (Figure S2D;  $61\% \pm 4\%$ , mean  $\pm$  SD). Other statistical classifiers and other definitions of IFP response yielded similar conclusions (Figure S3; Supplemental Data).



**Figure 2. Locations Showing Strongest Selectivity**

Classification performance using neural ensembles from different brain locations. We randomly selected ten electrodes from the pool of all the electrodes in each location. The subplots show the classification performance using the responses from these neural ensembles for the six brain regions that yielded the highest selectivity (see Figure S5 for the ensemble performance values for all other locations). The location codes and names are shown in Table S2. The bar colors denote the object categories (see exemplars on the right). The horizontal dashed line indicates the chance level (50%), and the dotted line shows the statistical significance threshold (Figure S2A). Error bars indicate 1 SD.

As illustrated in the two examples in Figure 1, the category selectivity could not be explained by preferences for individual exemplars. Very few electrodes showed statistically significant performance levels when the classifier was trained to discriminate among individual exemplars instead of across categories (Supplemental Data). Additionally, the classification performance also dropped significantly when each category was arbitrarily defined by randomly drawing 5 exemplars out of the total of 25 exemplars (Supplemental Data). We asked whether the physiological differences across object categories could be explained by low-level image properties shared by the exemplars of a given category. To address this question, we considered 15 different image properties including contrast, number of bright or dark boxes, and so forth and computed the correlation coefficient between the IFP responses and these image properties for each electrode. Although in a few electrodes we observed significant correlations, on average, the low-level image properties that yielded the strongest correlations could not account for more than 10% of the variance in the IFP responses (Supplemental Data).

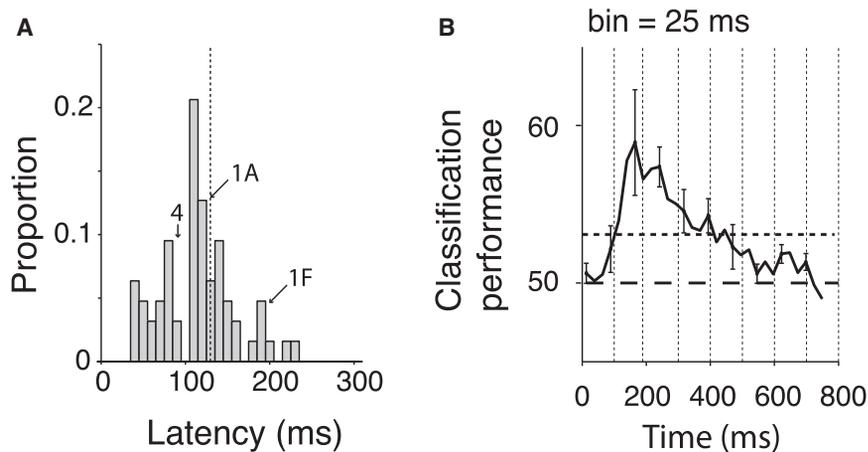
Of the 912 electrodes, 111 (12%) showed selectivity. These electrodes were nonuniformly distributed across the different lobes (occipital: 35%; temporal: 14%; frontal: 7%; parietal 4%). To determine the location of the electrodes, we coregistered the preoperative magnetic resonance images with the postoperative computer tomography images and mapped each electrode onto one of 80 possible brain locations (Table S2). The areas with the highest proportions of selective electrodes were the inferior occipital gyrus (86%; area 14 in Table S2), fusiform gyrus (38%; area 17 in Table S2), parahippocampal portion of the medial temporal lobe (26%; area 19 in Table S2), midoccipital gyrus (22%; area 15 in Table S2), lingual gyrus in the medial temporal lobe (21%; area 18 in Table S2), inferior temporal cortex (18%; area 31 in Table S2), and the temporal pole (14%; area 43 in Table S2). The number of selective electrodes in all the areas that we studied is shown in Table S1.

The analyses reported thus far describe the selectivity of individual electrodes, but the brain has access to information from

many different locations. In order to quantify how well we could decode information by considering the ensemble activity across electrodes or across brain regions, we constructed a vector containing the responses of multiple electrodes and used the same statistical classifier approach described above (see Hung et al., 2005 and Experimental Procedures). Data from multiple electrodes were concatenated assuming that the activity was independent (Supplemental Data). Figure S4A shows one way to build such an ensemble vector by selecting electrodes based on the  $r_v$  value, that is, the ratio of the variance across object categories to the variance within object categories. The electrode selection was performed using only the training data. Using such an ensemble of 11 electrodes, the overall classification performance ranged from 74% to 95%. These values correspond to binary classification for each object category where chance performance is 50%; the values for multiclass classification where chance performance is 20% are shown in Figure S4B. These values provide only a lower bound to the information available in the neural ensemble, because the addition of more electrodes, the use of more complex nonlinear classifiers, and the incorporation of interactions across areas could enhance the classification performance even more. In order to further evaluate the spatial specificity of the selective responses, we built neural ensembles by drawing electrodes based on their location (Table S1). The regions that yielded the highest classification performance levels are shown in Figure 2, and included the inferior occipital gyrus (area 14 in Table S2), fusiform gyrus (area 17 in Table S2), lingual gyrus (area 18 in Table S2), parahippocampal gyrus (area 19 in Table S2), inferior temporal cortex (area 31 in Table S2), and the temporal pole (area 43 in Table S2). The classification performance values for all areas are shown in Figure S5. The ensemble results in Figure 2 (and Figure S5) are consistent with the location results presented for individual electrodes above.

#### Timing of Selective Responses

The intracranial field potentials show a temporal resolution of milliseconds. We therefore used the detailed time course of the



**Figure 3. Fast Decoding of Object Category**

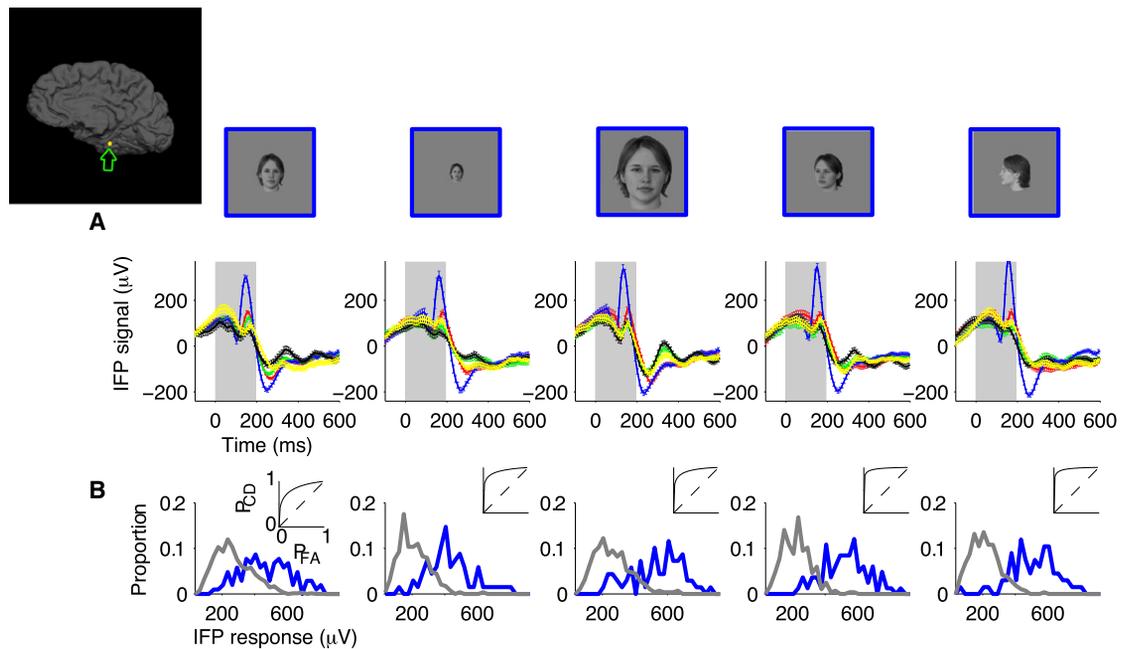
(A) Distribution of IFP latencies for all the electrodes that showed selective responses ( $n = 94$ ; see text). The latency was defined as the first time point where ten consecutive points yielded  $p < 0.01$  in a one-way ANOVA across object categories (see [Experimental Procedures](#) and [Figure S6A](#)). The dashed line shows the mean of the distribution, and the arrows show the position of the example electrodes in [Figures 1A](#), [1F](#), and [4](#). (B) Average classification performance as a function of time from stimulus onset averaged over all five categories (see [Figures S7A–S7C](#) for the classification performance for each category and for other bin sizes). The horizontal dashed line indicates chance classification performance, and the dotted line indicates the statistical significance threshold based on randomly shuffling the category labels. The vertical dashed lines are spaced every 100 ms to aid visualization of the dynamics of the response. Error bars indicate 1 SD.

[Experimental Procedures](#) and [Hung et al., 2005](#)). This figure shows classification performance as a function of time from stimulus onset averaged over all five categories (see [Figures S7A–S7C](#) for the classification performance for each category and for other bin sizes). The horizontal dashed line indicates chance classification performance, and the dotted line indicates the statistical significance threshold based on randomly shuffling the category labels. The vertical dashed lines are spaced every 100 ms to aid visualization of the dynamics of the response. Error bars indicate 1 SD.

IFP to estimate the latency of the selective responses. We followed the procedure in [Thorpe et al. \(1996\)](#): (1) we defined an electrode to be “selective” if there were 25 consecutive bins (bin size = 2 ms at Brigham and Women’s Hospital [BWH] and 3.9 ms at Children’s Hospital Boston [CHB]) where a one-way ANOVA on the IFP responses across object categories yielded  $p < 0.01$ ; and (2) for the selective electrodes, we defined the latency as the earliest time point where 10 consecutive bins yielded  $p < 0.01$  in the same analysis ([Figure S6A](#)). The latencies depend on the threshold parameters; this parameter dependence is described in [Figures S6B](#) and [S6C](#). The parameters 25, 10, and 0.01 represent conservative cutoff values that yield a low probability of false alarm. This selectivity criterion yielded 94 selective electrodes (cf. 111 selective electrodes reported above); the number of electrodes that passed the selectivity criterion in 1000 shuffles of the category labels was 0. The distribution of latencies is shown in [Figure 3A](#); the mean latency across all the electrodes was  $115 \pm 44$  ms. The latencies for those electrodes in the inferior occipital gyrus (area 14 in [Table S2](#)) were shorter than the ones for electrodes in temporal lobe locations ([Figure S6D](#)). In order to evaluate the earliest time point at which we could reliably decode category information in single trials, we binned the IFP responses using windows of 25 ms and computed the IFP range in each bin. We used an SVM classifier to decode category information in each time bin ([Figure 3B](#)). This is a demanding task for the classifier, because we only used a very small amount of data (25 ms) to train the classifier and to test its performance (see [Figures S7A–S7C](#) for the results for each object category and using bin sizes of 25, 50, and 100 ms). In spite of the task difficulty, we could still discriminate category information as early as  $\sim 100$  ms after stimulus onset. The latencies and decoding results are compatible with single-neuron studies in macaque monkey inferior temporal cortex ([Hung et al., 2005](#); [Keyser et al., 2001](#)), with psychophysical observations in humans ([Potter and Levy, 1969](#)) and also with scalp EEG recordings in humans ([Thorpe et al., 1996](#)).

### Invariance to Scale and Viewpoint

A key aspect of visual recognition involves not only being able to discriminate among different objects but also achieving invariance to object transformations ([Desimone et al., 1984](#); [Hung et al., 2005](#); [Logothetis et al., 1995](#); [Riesenhuber and Poggio, 1999](#); [Rolls, 1991](#)). We considered two important and typical object transformations: scaling and depth rotation. Each exemplar object was presented in a “default” rotation and size ( $3^\circ$  of visual angle) and also two additional rotations ( $\sim 45^\circ$  and  $\sim 90^\circ$ ) and two additional scales ( $1.5^\circ$  and  $6^\circ$  of visual angle; [Figure S1C](#)). Except in the case of human faces, the definition of a default viewpoint for the other object categories is arbitrary (these default viewpoints are shown in [Figure S1B](#)). We asked whether the selective responses, such as the ones illustrated in [Figure 1](#), would be maintained in spite of these object transformations. An example of a recording site that showed strong robustness to scale and depth rotation transformations is shown in [Figure 4A](#). This electrode, located in the parahippocampal part of the right medio-temporal-gyrus (Talairach coordinates: 32.6,  $-34.8$ ,  $-13.6$ ), showed an enhanced response to human faces at all the scales and depth rotations that we tested. The selective responses were consistent across repetitions, exemplars, and object transformations ([Figure S8](#)). The distribution of IFP responses for the preferred and nonpreferred categories and the ROC analyses ([Figure 4B](#)) suggested that it would be possible to discriminate the preferred object across object transformations in single trials. To evaluate the degree of extrapolation across changes in scale and depth rotation, we trained the classifier using *only* the IFP responses to the default scale and rotation and tested its performance using the IFP responses to the transformed objects ([Experimental Procedures](#)). Other training set/test set combinations yielded similar results ([Supplemental Data](#)). The decoder showed invariance, with an average classification performance across the four conditions in [Figure 4](#) of  $73\% \pm 8\%$ . Overall, invariance to scale and depth rotation was observed in at least one electrode in 10 of the 11 subjects.



**Figure 4. Robustness to Scale and Depth Rotation**

(A) Responses of an electrode located in the parahippocampal part of the right medio-temporal-gyrus (Talairach coordinates: 32.6, -34.8, -13.6; see inset depicting the electrode position). Each column depicts the IFP responses to images where the objects were shown at different scales or depth rotations. Each object was presented in a default viewpoint (front view in the case of human faces; see Figure S1B for the default viewpoint for the other categories) at a scale of 3° (column 1, total  $n = 329$ ), with the same default viewpoint at a scale of 1.5° (column 2, total  $n = 359$ ), at a scale of 6° (column 3, total  $n = 363$ ), at a scale of 3° and ~45° depth rotation (column 4, total  $n = 370$ ), and at a scale of 3° and ~90° depth rotation (column 5, total  $n = 355$ ). Above the physiological responses, we show an example object to illustrate the changes in scale and viewpoint, but the data correspond to the average over all five exemplars for each category. The format for the response subplots is the same as in Figure 1A. Error bars indicate SEM. The gray rectangle denotes the stimulus presentation time. Figure S8 shows the responses to each one of the exemplars for this electrode.

(B) Distribution of IFP responses (IFP range in the 50–300 ms window) for “human faces” (blue) versus other object categories (gray). The insets show the corresponding ROC curves.

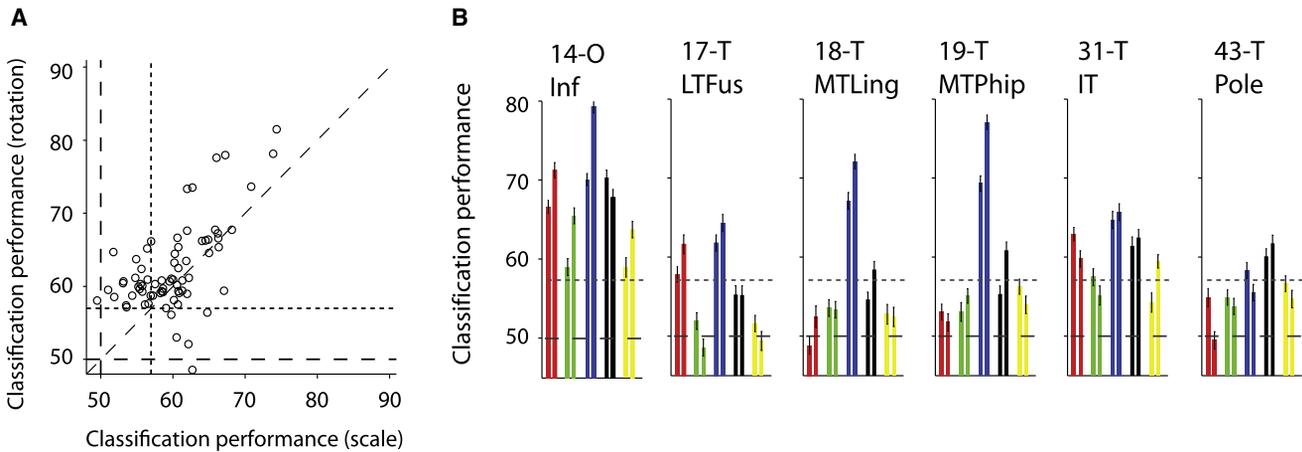
Forty-nine electrodes (44% of the selective electrodes) showed invariance to scale; the classification performance ranged from 60% to 82% (Figure S2F;  $64\% \pm 4\%$ , mean  $\pm$  SD). Sixty-four electrodes (58% of the selective electrodes) showed invariance to depth rotation; the classification performance ranged from 60% to 81% (Figure S2E;  $65\% \pm 4\%$ , mean  $\pm$  SD). Most of the electrodes that showed invariance to scale changes also showed invariance to depth rotation changes and vice versa. There was a strong correlation between the classification performance values for these two transformations (Figure 5A; Pearson correlation coefficient = 0.70). The majority of the electrodes that showed invariance to scale or depth rotation were located in the occipital lobe (inferior occipital gyrus; area 14 in Table S2) and in the temporal lobe (parahippocampal gyrus [area code 19], inferior temporal cortex [area code 31], fusiform gyrus [area code 17], medial temporal lobe [area code 18], and temporal pole [area code 43]). The number of electrodes in each location that showed invariance to scale and depth rotation is shown in Table S1. The classification performance values for a neural ensemble of 11 electrodes ranged from 58% to above 90% (Figures S4C and S4D). The neural ensembles formed by electrodes in the inferior occipital gyrus or electrodes in the temporal lobe showed the strongest degree of robustness to scale and

depth rotation changes (Figure 5B; see Figure S5 for the classification performance values for ensembles from all brain regions).

We estimated the latency of the IFP responses as described in the previous section (Figure S6A). Latencies were computed separately for each one of the five combinations of scale and depth rotation. There was no significant difference among the latencies for the different object transformations (Figure 6A). Robust and selective responses were apparent within ~100 ms of processing, suggesting that the neural responses in the temporal lobe may not require additional time to analyze different scales or depth rotations. It should be noted that this statement refers to categorizing the images based on the physiological responses, and may not extrapolate to identifying individual exemplars. Consistent with the short latencies, we could also decode the object category as early as ~100 ms after stimulus onset, even when extrapolating across scaled and rotated versions of the objects (Figure 6B; see also Figures S7D–S7F).

## DISCUSSION

Visual object recognition is challenging because the same object can cast an infinite number of patterns onto the retina. The difficulty of the problem is illustrated by the observation that, in spite



**Figure 5. Locations Showing Strongest Robustness to Object Changes**

(A) Classification performance for scale (x axis) versus rotation (y axis) invariance. In both cases, the classifier was trained using the IFP responses to the default scale and rotation, and its performance was evaluated using the IFP responses to the scaled objects (x axis) or rotated objects (y axis) (see [Experimental Procedures](#)). The horizontal and vertical dashed lines indicate chance performance, and the dotted lines indicate the statistical significance threshold. The diagonal dashed line is the identity line.

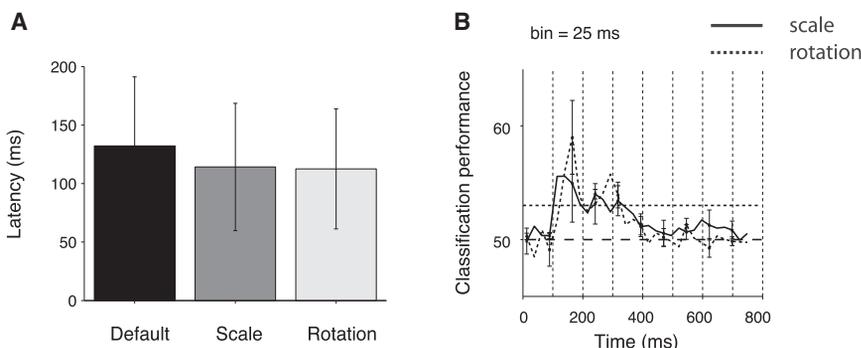
(B) Locations that showed strongest robustness to scale and rotation changes (the format and conventions are the same as in [Figure 2](#)). In contrast to [Figure 2](#), here there are two bars for each category: the first one corresponds to rotation invariance (average over the two rotations) and the second one corresponds to scale invariance (average over the two scales). Error bars indicate 1 SD. The corresponding results for all locations are shown in [Figure S5](#).

of major advances in computational power, human recognition performance currently outperforms computer vision algorithms (e.g., [Serre et al., 2007a, 2007b](#)). The speed of visual recognition imposes a strong constraint on theories of visual recognition by limiting the plausible number of computational steps ([Deco and Rolls, 2004](#); [Keyser et al., 2001](#); [Riesenhuber and Poggio, 1999](#); [Serre et al., 2007a](#); [Thorpe et al., 1996](#); [VanRullen and Thorpe, 2002](#)). Our results show that neural activity in the human occipital and temporal cortex ([Figures 2 and 5](#)) supports a representation that is consistent with the basic properties of recognition at the behavioral level in terms of selectivity ([Figures 1 and 2](#)), robustness to object transformations ([Figures 4 and 5](#)), and speed ([Figures 3 and 6](#)).

It may be tempting to describe the selectivity in one of our electrodes as indicative of a brain “area” or “module” devoted to recognizing objects of the preferred category, as claimed in functional imaging studies ([Grill-Spector and Malach, 2004](#);

[Haxby et al., 2001](#); [Kanwisher et al., 1997](#)). However, we should be careful about such an interpretation of our recordings. First, we examined the responses to a limited number of object categories. Even though most electrodes responded best to one category as opposed to multiple categories ([Figure S2B](#)), we cannot rule out the possibility that some of these electrodes may also respond to other object categories that were not presented to the subjects. Second, it is difficult to precisely quantify the spatial extent of the selective responses with our recordings. Third, we observed different electrodes with distinct selectivity in a given brain region ([Figure S5](#)).

The location of the electrodes in our study was dictated by clinical criteria. Therefore, the electrode coverage was far from exhaustive. There may well be other areas in the human brain, not interrogated in the current efforts, which also show selectivity and invariance. Yet, we note that the location of the selective responses that we observed is generally compatible with human



**Figure 6. Fast and Robust Decoding of Object Category**

(A) Latencies for the electrodes that showed IFP responses invariant to scale or rotation changes (see [Experimental Procedures](#) and [Figure S6A](#) for the definition of latency). Error bars indicate 1 SD.

(B) Dynamics of classification performance. The format and conventions are the same as in [Figure 3B](#). The solid curve shows the classification performance values for scale invariance, and the dashed curve shows the classification performance values for rotation invariance. The classification performance values were averaged across the five object categories (see [Figures S7D–S7F](#) for the corresponding results for each individual category and different bin sizes).

and macaque monkey functional imaging studies as well as with single-unit and field potential studies. For example, both electrophysiological studies (McCarthy et al., 1999; Privman et al., 2007) and functional imaging studies in humans (Grill-Spector and Malach, 2004; Haxby et al., 2001; Kanwisher et al., 1997; Tsao et al., 2006) have described the selective activation of areas in the fusiform gyrus (commonly called the fusiform face area) in response to face stimuli. The selectivity in the medial temporal lobe areas is also compatible with single-unit and local field potential studies in humans (Kraskov et al., 2007; Kreiman, 2007). Although comparisons across species are not entirely trivial, the face selectivity in the fusiform gyrus areas is consistent with recent single-unit and imaging measurements in macaque monkeys (Tsao et al., 2006), and the selective responses in human inferior temporal cortex that we report here may correspond to the areas that show selectivity and invariance in the macaque monkey inferior temporal cortex (Desimone et al., 1984; Hung et al., 2005; Logothetis and Sheinberg, 1996; Perrett et al., 1992; Richmond et al., 1983; Tanaka, 1996). The involvement of the fusiform gyrus and inferior temporal cortex in visual recognition is further supported by lesion studies in subjects who show severe recognition deficits after temporal lobe damage (Damasio et al., 1990; De Renzi, 2000; Grill-Spector and Malach, 2004; Holmes and Gross, 1984; Logothetis and Sheinberg, 1996; Tippet et al., 1996).

Although visual selectivity is interesting in itself, the key challenge in recognition is to combine selectivity with invariance to transformations. It is easy (and useless) to implement algorithms that can achieve exquisite selectivity without being able to extrapolate across different transformations of the same object. Our recordings provide evidence of category invariance (i.e., similar selective responses to multiple exemplars within a category) (Freedman et al., 2001; Hung et al., 2005; Kanwisher et al., 1997; Kreiman et al., 2000; Sigala and Logothetis, 2002), scale invariance (i.e., similar selectivity across different scales) (Ito et al., 1995; Logothetis and Sheinberg, 1996; Tanaka, 1996), and viewpoint invariance (i.e., similar selectivity across depth rotations) (Ito et al., 1995; Logothetis et al., 1995; Logothetis and Sheinberg, 1996; Tanaka, 1996).

Our observations suggest that selective and invariant responses arise with approximately the same latency. These observations suggest that there may not be a need for extensive additional computation to recognize a transformed object. This observation is compatible with both single-unit studies in macaque monkeys (Desimone et al., 1984; Hung et al., 2005; Logothetis et al., 1995) and with computational models of object recognition (Poggio and Edelman, 1990; Serre et al., 2007a; Wallis and Rolls, 1997). We note that other investigators have suggested that discriminating fine information (e.g., object identity) requires longer latencies (Sugase et al., 1999).

The selectivity observed in functional imaging studies and the behavioral deficits in lesion studies inform us about the “where” but not about the “when” of neural activity. Scalp recordings and psychophysical observations have provided temporal bounds for visual recognition (Potter and Levy, 1969; Thorpe et al., 1996) (“when” information), but it is hard to infer the locations of processing steps along the visual cortex from those studies. The results that we describe here link the “where” and “when” and argue that the stream of visual information reaches the highest

stages of processing in the inferior temporal cortex within less than 200 ms after stimulus onset. These latencies are significantly shorter than the ones reported in the human medial temporal lobe, which range from 200 to 500 ms (Kreiman et al., 2000; Mormann et al., 2008). Why there is such a relatively long gap between the responses in visual cortex (as reported here in the human visual cortex and in many studies in the macaque monkey inferior temporal cortex) (Hung et al., 2005; Perrett et al., 1992; Richmond et al., 1983) and the responses in the human and macaque medial temporal lobe (Kreiman et al., 2000; Mormann et al., 2008; Suzuki et al., 1997) is an important question for future investigation.

From an anatomical viewpoint, it is clear that there are extensive recurrent connections throughout the cortical circuitry (e.g., Callaway, 2004; Douglas and Martin, 2004; Felleman and Van Essen, 1991). The key question here and in theories of visual recognition is the relative functional weight of bottom-up, horizontal, and top-down influences during visual processing. In discussions of recurrent connections and backprojections, it is important to distinguish between “short” and “long” loops. We argue that the brief latencies and decoding in single trials shown here are compatible with feedforward accounts of immediate visual recognition. This is based on the assumption that long, recurrent loops (e.g., backprojections from inferior temporal cortex to primary visual cortex and back to inferior temporal cortex) would trigger responses with longer latencies. The results shown here are still compatible with short recurrent loops (e.g., those involving horizontal connections within an area, or loops between adjacent areas such as V1-V2). It is clear that in many cases the selective visual responses extended well past 300 ms (see Figures S1E and S1F for two particularly clear examples). The analyses presented here focused on the initial 300 ms after stimulus onset, because these initial signals can provide more insight about the relative contribution of bottom-up and top-down signals. The later phase of the physiological signals is likely to involve many top-down components that may include memory processes, task/planning processes, and emotional processes (Kreiman et al., 2000; Mormann et al., 2008; Tomita et al., 1999).

In order to understand the neuronal circuits and mechanisms that give rise to visual recognition, it is not sufficient to map the areas involved. Our observations bridge the human vision field with the high-resolution information derived from electrophysiological studies in macaque monkeys. These results provide strong bounds about the dynamics of vision in single trials and constrain the development of theoretical models of visual object recognition.

## EXPERIMENTAL PROCEDURES

### Subjects

Subjects were 11 patients (6 male, 9 right-handed, 12–34 years old) with pharmacological intractable epilepsy. They were admitted into either Children’s Hospital Boston (CHB) or Brigham and Women’s Hospital (BWH) for the purpose of localizing the seizure foci for potential surgical resection. All the tests described here were approved by the institutional review boards at both hospitals and were carried out under the subjects’ consent.

### Recordings

In order to localize the seizure event foci, subjects were implanted with intracranial electrodes (Ad-Tech, Racine, WI, USA; each recording site was 2 mm in diameter with 1 cm separation) (Engel et al., 2005; Kahana et al., 1999). Recording sites were arranged in grids or strips. Each grid/strip

contained anywhere from 4 to 64 recording sites. The total number of recording sites per subject ranged from 48 to 126 ( $80.4 \pm 18.4$ , mean  $\pm$  SD). The signal from each recording site was amplified ( $\times 2500$ ) and filtered between 0.1 and 100 Hz with a sampling rate of 256 Hz at CHB (XLTEK, Oakville, ON, Canada) and 500 Hz at BWH (Bio-Logic, Knoxville, TN, USA). A notch filter was applied at 60 Hz. Data were stored for offline processing. Throughout the text, we refer to the recorded signal as the “intracranial field potential” or IFP. This nomenclature aims to distinguish these electrophysiological recordings from scalp electroencephalographic recordings (EEG) and also from local field potentials (LFPs) measured through microwires. Subjects stayed in the hospital 6–9 days and physiological data were continuously monitored during this period. All the data reported in this manuscript come from periods without any seizure events. The number and location of the recording sites were determined exclusively by clinical criteria and are reported in Table S1.

### Stimulus Presentation

Subjects were presented with contrast-normalized grayscale digital photographs of objects from five different categories: “animals,” “chairs,” “human faces,” “fruits,” and “vehicles” (Figure S1B). Contrast was normalized by fixing the standard deviation of the grayscale pixel levels. In four subjects, we presented two additional object categories (artificial Lego objects and shoes); here we describe only the five categories that were common to all subjects. There were five different exemplars per category. To assess the robustness to object changes, each exemplar was shown in one of five possible transformations (Figure S1C): (1) 3° size, “default” viewpoint (the default viewpoint is defined by the images in Figure S1B); (2) 1.5° size, default viewpoint; (3) 6° size, default viewpoint; (4) 3° size,  $\sim 45^\circ$  depth rotation with respect to the default viewpoint; or (5) 3° size,  $\sim 90^\circ$  depth rotation with respect to the “default” viewpoint. Each object was presented for 200 ms and was followed by a 600 ms gray screen (Figure S1A). Subjects performed a one-back task, indicating whether an exemplar was repeated or not regardless of scale or viewpoint changes. Images were presented in pseudorandom order.

### Data Analyses

We succinctly describe the algorithms and analyses here; further details about the analytical methods are provided in Supplemental Data.

### Electrode Localization

To localize the electrodes, we integrated the anatomical information of the brain provided by preoperative magnetic resonance imaging (MRI) and the spatial information of the electrode positions provided by postoperative computer tomography (CT). For each subject, the 3D brain surface was reconstructed and then an automatic parcellation was performed using Freesurfer (Dale et al., 1999). The electrode positions were mapped onto 80 brain areas; these areas are listed together with Talairach coordinates in Table S2. Based on these coordinates, the electrodes were superimposed on the reconstructed brain surface for visualization purposes in the figures.

### Classifier Analysis

We used a statistical learning approach (Bishop, 1995; Hung et al., 2005) to decode visual information from the IFP responses in single trials. More details about the classifier analysis are provided in Supplemental Data. Unless stated otherwise, throughout the text we refer to the “IFP response” as the range of the signal ( $\max(\text{IFP}) - \min(\text{IFP})$ ) in the interval  $T$ , and the default  $T$  is [50;300] ms after stimulus onset (see Figure S3 for other definitions). The classifier approach allows us to consider either each electrode independently or the encoding of information by an ensemble of multiple electrodes. In the case of single electrodes and single bins per electrode, the results obtained using a classifier are very similar to those obtained using an ROC analysis (Green and Swets, 1966) or a one-way ANOVA across categories on the IFP responses (Supplemental Data). Unless stated otherwise, the results shown throughout the manuscript correspond to binary classification between a given object category and the other object categories (chance = 50%). In the case of multiclass problems, we used a one-versus-all approach (chance = 20%; Figure S4B). In Figures 1 and 4, we also present an ROC curve (Green and Swets, 1966) (e.g., Figure 1E), showing the proportion of correct detections ( $P_{CD}$ , y axis) as a function of the probability of false alarms ( $P_{FA}$ , x axis). Importantly, in all cases, the data were divided into two nonoverlapping sets, a training set and a test set. The way of dividing the data into a training set and test set depended on the

specific question asked (see Results and Supplemental Data). When using the classifier to assess selectivity, the training set consisted of 70% of the data and the classification performance was evaluated on the remaining 30% of the data. When using the classifier to assess the degree of extrapolation across object changes, the training set consisted of the IFP responses to objects at one scale and depth rotation, whereas the test set consisted of the IFP responses to the remaining scales and depth rotations. The classifier was trained to learn the map between the physiological signals and the object category. The performance of the classifier was evaluated by comparing the classifier predictions on the test data with the actual labels for the test data. Throughout the text, we report the proportion of test repetitions correctly labeled as “classification performance.” In the case of binary classification, the chance level is 50% and is indicated in the plots by a dashed line. The maximum classification performance is 100%. Unless stated otherwise, we used a support vector machine classifier with a linear kernel. To evaluate the statistical significance of a given classification performance value, we compared the results against the distribution of classification performance values obtained under the null hypothesis defined by randomly shuffling the category labels (Figure S2A).

### Other Statistical Analyses

We also performed a one-way analysis of variance (ANOVA) on the IFP responses. The ANOVA results were consistent with the single-electrode statistical classifier results (Supplemental Data). Throughout the text, we report the classifier results because they provide a more rigorous quantitative approach and a stricter criterion for selectivity, they allow considering ensembles of electrodes in a straightforward way, and they provide a direct measure of performance in single trials. In order to assess whether category selectivity could be explained by selectivity to particular exemplars within a category, we rank ordered the responses to all the individual exemplars (e.g., Figure 1C). As a null hypothesis, we assumed that the distribution of rank orders across categories was uniform and we assessed deviations from this null hypothesis by randomly assigning exemplars to categories in 100,000 iterations. The  $p$  values quoted in the text for this analysis indicate the odds that  $m$  exemplars from any category would rank among the top  $n$  best responses.

### SUPPLEMENTAL DATA

Supplemental data include eight figures, two tables, and experimental procedures and can be found with this article online at [http://www.neuron.org/supplemental/S0896-6273\(09\)00171-8](http://www.neuron.org/supplemental/S0896-6273(09)00171-8).

### ACKNOWLEDGMENTS

We would like to thank the patients for their cooperation in these studies. We also thank Margaret Livingstone, Ethan Meyers, and Thomas Serre for comments on the manuscript, and Calin Buia, Sheryl Manganaro, Paul Dionne, and Malena Español for technical assistance. We also acknowledge the financial support from the Epilepsy Foundation, the Whitehall Foundation, the Klingenstein Fund, and Children’s Hospital Boston Ophthalmology Foundation.

Accepted: February 24, 2009

Published: April 29, 2009

### REFERENCES

- Bishop, C.M. (1995). *Neural Networks for Pattern Recognition* (Oxford: Clarendon Press).
- Bullier, J. (2001). Integrated model of visual processing. *Brain Res. Brain Res. Rev.* 36, 96–107.
- Callaway, E.M. (2004). Feedforward, feedback and inhibitory connections in primate visual cortex. *Neural Netw.* 17, 625–632.
- Dale, A.M., Fischl, B., and Sereno, M.I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* 9, 179–194.
- Damasio, A.R., Tranel, D., and Damasio, H. (1990). Face agnosia and the neural substrates of memory. *Annu. Rev. Neurosci.* 13, 89–109.

- Dean, P. (1976). Effects of inferotemporal lesions on the behavior of monkeys. *Psychol. Bull.* 83, 41–71.
- Deco, G., and Rolls, E.T. (2004). A neurodynamical cortical model of visual attention and invariant object recognition. *Vision Res.* 44, 621–642.
- De Renzi, E. (2000). Disorders of visual recognition. *Semin. Neurol.* 20, 479–485.
- Desimone, R., Albright, T., Gross, C., and Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *J. Neurosci.* 4, 2051–2062.
- Douglas, R.J., and Martin, K.A. (2004). Neuronal circuits of the neocortex. *Annu. Rev. Neurosci.* 27, 419–451.
- Engel, A.K., Moll, C.K., Fried, I., and Ojemann, G.A. (2005). Invasive recordings from the human brain: clinical insights and beyond. *Nat. Rev. Neurosci.* 6, 35–47.
- Felleman, D.J., and Van Essen, D.C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* 1, 1–47.
- Freedman, D.J., Riesenhuber, M., Poggio, T., and Miller, E.K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science* 291, 312–316.
- Fukushima, K. (1980). Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol. Cybern.* 36, 193–202.
- Garrido, M.I., Kilner, J.M., Kiebel, S.J., and Friston, K.J. (2007). Evoked brain responses are generated by feedback loops. *Proc. Natl. Acad. Sci. USA* 104, 20961–20966.
- Green, D., and Swets, J. (1966). *Signal Detection Theory and Psychophysics* (New York: Wiley).
- Grill-Spector, K., and Malach, R. (2004). The human visual cortex. *Annu. Rev. Neurosci.* 27, 649–677.
- 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.
- Hochstein, S., and Ahissar, M. (2002). View from the top: hierarchies and reverse hierarchies in the visual system. *Neuron* 36, 791–804.
- Holmes, E.J., and Gross, C.G. (1984). Stimulus equivalence after inferior temporal lesions in monkeys. *Behav. Neurosci.* 98, 898–901.
- Hung, C.P., Kreiman, G., Poggio, T., and DiCarlo, J.J. (2005). Fast read-out of object identity from macaque inferior temporal cortex. *Science* 310, 863–866.
- Ito, M., Tamura, H., Fujita, I., and Tanaka, K. (1995). Size and position invariance of neuronal responses in monkey inferotemporal cortex. *J. Neurophysiol.* 73, 218–226.
- Johnson, J.S., and Olshausen, B.A. (2005). The earliest EEG signatures of object recognition in a cued-target task are postsensory. *J. Vis.* 5, 299–312.
- Kahana, M.J., Sekuler, R., Caplan, J.B., Kirschen, M., and Madsen, J.R. (1999). Human theta oscillations exhibit task dependence during virtual maze navigation. *Nature* 399, 781–784.
- Kanwisher, N., McDermott, J., and Chun, M.M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* 17, 4302–4311.
- Keysers, C., Xiao, D.K., Foldiak, P., and Perrett, D.I. (2001). The speed of sight. *J. Cogn. Neurosci.* 13, 90–101.
- Kraskov, A., Quiroga, R.Q., Reddy, L., Fried, I., and Koch, C. (2007). Local field potentials and spikes in the human medial temporal lobe are selective to image category. *J. Cogn. Neurosci.* 19, 479–492.
- Kreiman, G. (2007). Single unit approaches to human vision and memory. *Curr. Opin. Neurobiol.* 17, 471–475.
- Kreiman, G., Koch, C., and Fried, I. (2000). Category-specific visual responses of single neurons in the human medial temporal lobe. *Nat. Neurosci.* 3, 946–953.
- Logothetis, N.K., and Sheinberg, D.L. (1996). Visual object recognition. *Annu. Rev. Neurosci.* 19, 577–621.
- Logothetis, N.K., Pauls, J., and Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Curr. Biol.* 5, 552–563.
- McCarthy, G., Puce, A., Belger, A., and Allison, T. (1999). Electrophysiological studies of human face perception. II: Response properties of face-specific potentials generated in occipitotemporal cortex. *Cereb. Cortex* 9, 431–444.
- Mormann, F., Kornblith, S., Quiroga, R.Q., Kraskov, A., Cerf, M., Fried, I., and Koch, C. (2008). Latency and selectivity of single neurons indicate hierarchical processing in the human medial temporal lobe. *J. Neurosci.* 28, 8865–8872.
- Ojemann, G.A. (1997). Treatment of temporal lobe epilepsy. *Annu. Rev. Med.* 48, 317–328.
- Perrett, D.I., Hietanen, J.K., Oram, M.W., and Benson, P.J. (1992). Organization and functions of cells responsive to faces in the temporal cortex. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 335, 23–30.
- Poggio, T., and Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature* 343, 263–266.
- Potter, M.C., and Levy, E.I. (1969). Recognition memory for a rapid sequence of pictures. *J. Exp. Psychol.* 81, 10–15.
- Privman, E., Nir, Y., Kramer, U., Kipervasser, S., Andelman, F., Neufeld, M., Mukamel, R., Yeshurun, Y., Fried, I., and Malach, R. (2007). Enhanced category tuning revealed by intracranial electroencephalograms in high-order human visual areas. *J. Neurosci.* 27, 6234–6242.
- Quiroga, R.Q., Reddy, L., Kreiman, G., Koch, C., and Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature* 435, 1102–1107.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychol. Bull.* 124, 372–422.
- Richmond, B.J., Wurtz, R.H., and Sato, T. (1983). Visual responses in inferior temporal neurons in awake rhesus monkey. *J. Neurophysiol.* 50, 1415–1432.
- Riesenhuber, M., and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nat. Neurosci.* 2, 1019–1025.
- Rolls, E.T. (1991). Neural organization of higher visual functions. *Curr. Opin. Neurobiol.* 1, 274–278.
- Schmolesky, M.T., Wang, Y., Hanes, D.P., Thompson, K.G., Leutgeb, S., Schall, J.D., and Leventhal, A.G. (1998). Signal timing across the macaque visual system. *J. Neurophysiol.* 79, 3272–3278.
- Serre, T., Kreiman, G., Kouh, M., Cadieu, C., Knoblich, U., and Poggio, T. (2007a). A quantitative theory of immediate visual recognition. *Prog. Brain Res.* 165, 33–56.
- Serre, T., Oliva, A., and Poggio, T. (2007b). Feedforward theories of visual cortex account for human performance in rapid categorization. *Proc. Natl. Acad. Sci. USA* 104, 6424–6429.
- Sigala, N., and Logothetis, N. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature* 415, 318–320.
- Sugase, Y., Yamane, S., Ueno, S., and Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex. *Nature* 400, 869–873.
- Suzuki, W.A., Miller, E.K., and Desimone, R. (1997). Object and place memory in the macaque entorhinal cortex. *J. Neurophysiol.* 78, 1062–1081.
- Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annu. Rev. Neurosci.* 19, 109–139.
- Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. *Nature* 381, 520–522.
- Tippett, L.J., Glosser, G., and Farah, M.J. (1996). A category-specific naming impairment after temporal lobectomy. *Neuropsychologia* 34, 139–146.
- Tomita, H., Ohbayashi, M., Nakahara, K., Hasegawa, I., and Miyashita, Y. (1999). Top-down signal from prefrontal cortex in executive control of memory retrieval. *Nature* 401, 699–703.
- Tsao, D.Y., Freiwald, W.A., Tootell, R.B., and Livingstone, M.S. (2006). A cortical region consisting entirely of face-selective cells. *Science* 311, 670–674.
- VanRullen, R., and Thorpe, S. (2002). Surfing a spike wave down the ventral stream. *Vision Res.* 42, 2593–2615.
- Wallis, G., and Rolls, E.T. (1997). Invariant face and object recognition in the visual system. *Prog. Neurobiol.* 51, 167–194.