Timing, timing, timing: Fast decoding of object information from intracranial field potentials in human visual cortex

Hesheng Liu[1,3], **Yigal Agam**[1], **Joseph R. Madsen**[2], and **Gabriel Kreiman**[1,4,5]

[1]Department of Neuroscience and Ophthalmology, Children's Hospital Boston, Harvard Medical School

[2]Department of Neurosurgery, Children's Hospital Boston, Harvard Medical School

[3]Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital

[4]Center for Brain Science, Harvard University

[5]Swartz Center for Theoretical Neuroscience, Harvard University

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 post-stimulus. 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 feed-forward 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; Quian Quiroga et al., 2005) and neurological studies in humans (Damasio et al., 1990; De Renzi, 2000; Tippett et al., 1996).

Correspondence: gabriel.kreiman@tch.harvard.edu .

Theories of visual recognition can be largely divided into two main groups: (i) theories advocating top-down influences and long recurrent feedback connections (e.g. (Bullier, 2001; Garrido et al., 2007; Hochstein and Ahissar, 2002)) and (ii) theories relying on a bottom-up approach and largely dependent on feed-forward processing steps (e.g. (Fukushima, 1980; Riesenhuber and Poggio, 1999; Rolls, 1991; vanRullen and Thorpe, 2002)). While 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 feed-forward 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; Keysers 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 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).

While 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 (IFPs) 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 to 34 years old) were presented with digital photographs of objects belonging to 5 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 5 different exemplars per category (Figure S1**B**). Images were grayscale, contrast-normalized and were presented for 200 ms in pseudo-random order, with an interval between images of 600 ms (Figure S1**A**), 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 1**A** , 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(IFP)-min(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 Figure S1 and **F** 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 (see Figure S3 and Supplementary Material). In the example in Figure 1**A** , a one-way ANOVA across categories on the IFP responses yielded $p < 10^{-6}$ (Experimental Procedures). The selective response was consistent across repetitions (*z*-score=3.2, Supplementary Material) and also across the different exemplars for each category (Figure 1**B**). When we ranked the objects according to the IFP response, the 5 best objects were the 5 exemplars from the "fruits" category (Figure 1**C** , 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 Figure 1**F–J** . 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 1**F** , blue trace). More examples are shown in Figure S1 **D–G** for more examples.

While 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 1**D** 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 1**E** shows the corresponding 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 (Supplementary Material). 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 Figure 1**A** and Figure 1**F** , we could decode the preferred object categories ("fruits" and "human faces" respectively) from the IFP responses in single trials with performance levels of 69±6% and 65±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 bootstrap procedure whereby we randomly shuffled the object category labels (Supplementary Material). Based on this bootstrap procedure, we used a threshold of 3 standard deviations from the mean based on the null hypothesis (Figure S2**A**). 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 S2**B**). Across all the electrodes, we observed selective responses to the 5 different object categories that were tested (Figure S2**C**)

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 S2**C**). We note that the observation of an electrode that shows selectivity for a particular object 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 S2**D** , 61%±4%, mean±s.d.). Other statistical classifiers and other definitions of IFP response yielded similar conclusions (Figure S3 and Supplementary Material).

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 (Supplementary Material). 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 (Supplementary Material). 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 fifteen different image properties including contrast, number of bright or dark boxes, etc. and computed the correlation coefficient between the IFP responses and the 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 (Supplementary Material).

Of the 912 electrodes, 111 (12%) showed selectivity. These electrodes were non-uniformly distributed across the different lobes (occipital: 35%, temporal: 14%, frontal: 7%, parietal 4%). To determine the location of the electrodes, we co-registered the pre-operative MR images with the post-operative CT 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), mid-occipital 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 (Supplementary Material). Figure S4**A** 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 S4**B** . 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 non-linear 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 IFP to estimate the latency of the selective responses. We followed the procedure in (Thorpe et al., 1996): (i) we defined an electrode to be "selective" if there were 25 consecutive bins (bin size = 2 ms at BWH and 3.9 ms at CHB) where a one-way ANOVA on the IFP responses across object categories yielded $p < 0.01$; (ii) 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 S6**A**). The latencies depend on the threshold parameters; this parameter dependence is described in Figure S6**B–C** . The parameters 25, 10 and 0.01 represent conservative cut-off 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 3**A** ; the mean latency across all the electrodes was 115±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 S6**D**). 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 3**B**). 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 Figure S7**A–C** for the results for each object category and using bin sizes of 25, 50 ms and 100 ms). In spite of the task difficulty, we could still discriminate category information as early as ~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; Keysers 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 degrees of visual angle) and also two additional rotations (~45 and ~90 degrees) and two additional scales (1.5 and 6 degrees of visual angle; Figure S1**C**). 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 S1**B**). 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 4**A** . 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 non-preferred categories and the ROC analyses (Figure 4**B**) 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 train set / test set combinations yielded similar results (Supplementary Material). The decoder showed invariance, with an average classification performance across the four conditions in Figure 4 of 73±8%. Overall, invariance to scale and depth rotation was observed in at least one electrode 10 of the 11 subjects. Forty-nine electrodes (44% of the selective electrodes) showed invariance to scale; the classification performance ranged from 60% to 82% (Figure S2**F** , 64%±4%, mean±s.d.). Sixty-four electrodes (58% of the selective electrodes) showed invariance to depth rotation; the classification performance ranged from 60% to 81% (Figure S2**E** ; 65%±4%, mean±s.d.). 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 5**A** ; 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 are shown in Table S1. The classification performance values for a neural ensemble of 11 electrodes ranged from 58% to above 90% (Figure S4 **C–D**). 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 5**B** ; 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 S6**A**). Latencies were computed separately for each one of the 5 combinations of scale and depth rotation. There was no significant difference among the latencies for the different object transformations (Figure 6**A**). 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 6**B** ; see also Figure S7 **D–F**).

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 of major advances in computational power, human recognition performance currently outperforms computer vision algorithms (e.g. (Serre et al., 2007a; Serre et al., 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; Keysers 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 S2**B**), 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 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 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 that 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; Tippett 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 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 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 and back-projections, it is important to distinguish between "short" and "long" loops. We argue that the brief latencies and decoding in single trials shown here is compatible with feed-forward accounts of immediate visual recognition. This is based on the assumption that long, recurrent loops (e.g. back-projections 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 Figure S1 **E** and **F** 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 to 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 IRBs 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 (each recording site was 2 mm diameter with 1 cm separation, Ad-Tech, Racine, WI) (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±18.4, mean±s.d.). The signal from each recording site was amplified (x2500) and filtered between 0.1 and 100 Hz with a sampling rate of 256 Hz at CHB (XLTEK, Oakville, ON) and 500 Hz at BWH (Bio-Logic, Knoxville, TN). A notch filter was applied at 60 Hz. Data was stored for off-line processing. Throughout the text, we refer to the recorded signal as "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 to 9 days and physiological data was continuously monitored during this period. All the data reported in this manuscript comes from periods without any seizure events. The number and location of the recording sites were determined exclusively by clinical criteria and is reported in Table S1.

Stimulus presentation

Subjects were presented with contrast-normalized grayscale digital photographs of objects from 5 different categories: "animals", "chairs", "human faces", "fruits" and "vehicles" (Figure S1**B**). Contrast was normalized by fixing the standard deviation of the grayscale pixel levels. In 4 subjects, we presented 2 additional object categories (artificial Lego objects and shoes); here we describe only the 5 categories that were common to all subjects. There were 5 different exemplars per category. To assess the robustness to object changes, each exemplar was shown in one of 5 possible transformations (Figure S1**C**): (i) 3 degrees size, "default" viewpoint (the "default" viewpoint is defined by the images in Figure S1**B**), (ii) 1.5 degrees size, "default" viewpoint, (iii) 6 degrees size, "default" viewpoint, (iv) 3 degrees size, ~45 degree depth rotation with respect to the "default" viewpoint, (v) 3 degrees size, ~90 degree 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 S1**A**). 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 the Supplementary Material.

Electrode localization—To localize the electrodes, we integrated the anatomical information of the brain provided by pre-operative Magnetic Resonance Imaging (MRI) and the spatial information of the electrode positions provided by post-operative Computer Tomography (CT). For each subject, the 3-D brain surface was reconstructed and then an automatic parcellation was performed using Freesurfer. 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 on single trials. More details

about the classifier analysis are provided in the Supplementary Material. Unless stated otherwise, throughout the text we refer to the "IFP response", as the range of the signal (max(IFP)-min(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 (Supplementary Material). 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 S4**B**). In Figures 1 and 4, we also present an ROC curve (Green and Swets, 1966) (e.g. Figure 1**E**), 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 non-overlapping 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 Supplementary Material). 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 while 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 S2**A**).

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 (Supplementary Material). 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 1**C**). 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

List of references

Bishop, CM. Neural Networks for Pattern Recognition. Oxford: Clarendon Press; 1995.

Bullier J. Integrated model of visual processing. Brain Res Brain Res Rev 2001;36:96–107. [PubMed: 11690606]

Callaway EM. Feedforward, feedback and inhibitory connections in primate visual cortex. Neural Netw 2004;17:625–632. [PubMed: 15288888]

Damasio A, Tranel D, Damasio H. Face agnosia and the neural substrtes of memory. Annual Review of Neuroscience 1990;13:89–109.

De Renzi E. Disorders of visual recognition. Semin Neurol 2000;20:479–485. [PubMed: 11149704]

Dean P. Effects of inferotemporal lesions on the behavior of monkeys. Psychological Bulletin 1976;83:41–71. [PubMed: 828276]

Deco G, Rolls ET. A neurodynamical cortical model of visual attention and invariant object recognition. Vision Res 2004;44:621–642. [PubMed: 14693189]

Desimone R, Albright T, Gross C, Bruce C. Stimulus-selective properties of inferior temporal neurons in the macaque. Journal of Neuroscience 1984;4:2051–2062. [PubMed: 6470767]

Douglas RJ, Martin KA. Neuronal circuits of the neocortex. Annu Rev Neurosci 2004;27:419–451. [PubMed: 15217339]

Engel AK, Moll CK, Fried I, Ojemann GA. Invasive recordings from the human brain: clinical insights and beyond. Nat Rev Neurosci 2005;6:35–47. [PubMed: 15611725]

Felleman DJ, Van Essen DC. Distributed hierarchical processing in the primate cerebral cortex. Cerebral Cortex 1991;1:1–47. [PubMed: 1822724]

Freedman D, Riesenhuber M, Poggio T, Miller E. Categorical representation of visual stimuli in the primate prefrontal cortex. Science 2001;291:312–316. [PubMed: 11209083]

Fukushima K. Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biological Cybernetics 1980;36:193–202. [PubMed: 7370364]

Garrido MI, Kilner JM, Kiebel SJ, Friston KJ. Evoked brain responses are generated by feedback loops. Proc Natl Acad Sci U S A 2007;104:20961–20966. [PubMed: 18087046]

Green, D.; Swets, J. Signal detection theory and psychophysics. New York: Wiley; 1966.

Grill-Spector K, Malach R. The human visual cortex. Annual Review of Neuroscience 2004;27:649–677.

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 2001;293:2425–2430. [PubMed: 11577229]

Hochstein S, Ahissar M. View from the top: hierarchies and reverse hierarchies in the visual system. Neuron 2002;36:791–804. [PubMed: 12467584]

Holmes E, Gross C. Stimulus equivalence after inferior temporal lesions in monkeys. Behavioral Neuroscience 1984;98:898–901. [PubMed: 6435646]

Hung C, Kreiman G, Poggio T, DiCarlo J. Fast Read-out of Object Identity from Macaque Inferior Temporal Cortex. Science 2005;310:863–866. [PubMed: 16272124]

Ito M, Tamura H, Fujita I, Tanaka K. Size and position invariance of neuronal responses in monkey inferotemporal cortex. J Neurophysiol 1995;73:218–226. [PubMed: 7714567]

Johnson J, Olshausen B. The earliest EEG signatures of object recognition in a cued-target task are postsensory. Journal of Vision 2005;5:299–312. [PubMed: 15929653]

Kahana M, Sekuler R, Caplan J, Kirschen M, Madsen JR. Human theta oscillations exhibit task dependence during virtual maze navigation. Nature 1999;399:781–784. [PubMed: 10391243]

Kanwisher N, McDermott J, Chun MM. The fusiform face area: a module in human extrastriate cortex specialized for face perception. Journal of Neuroscience 1997;17:4302–4311. [PubMed: 9151747]

Keysers C, Xiao DK, Foldiak P, Perret DI. The speed of sight. Journal of Cognitive Neuroscience 2001;13:90–101. [PubMed: 11224911]

Kraskov A, Quian Quiroga R, Reddy L, Fried I, Koch C. Local Field Potentials and Spikes in the Human Medial Temporal Lobe are Selective to Image Category. Journal of Cognitive Neuroscience 2007;19:479–492. [PubMed: 17335396]

Kreiman G. Single neuron approaches to human vision and memories. Current Opinion in Neurobiology 2007;17:471–475. [PubMed: 17703936]

Kreiman G, Koch C, Fried I. Category-specific visual responses of single neurons in the human medial temporal lobe. Nature Neuroscience 2000;3:946–953.

Logothetis NK, Pauls J, Poggio T. Shape representation in the inferior temporal cortex of monkeys. Current Biology 1995;5:552–563. [PubMed: 7583105]

Logothetis NK, Sheinberg DL. Visual object recognition. Annual Review of Neuroscience 1996;19:577–621.

McCarthy G, Puce A, Belger A, Allison T. Electrophysiological studies of human face perception. II: Response properties of face-specific potentials generated in occipitotemporal cortex. Cerebral Cortex 1999;9:431–444. [PubMed: 10450889]

Mormann F, Kornblith S, Quiroga RQ, Kraskov A, Cerf M, Fried I, Koch C. Latency and selectivity of single neurons indicate hierarchical processing in the human medial temporal lobe. J Neurosci 2008;28:8865–8872. [PubMed: 18768680]

Ojemann GA. Treatment of temporal lobe epilepsy. Annual Review of Medicine 1997;48:317–328.

Perrett D, Hietanen J, Oeam M, Benson P. Organization and functions of cells responsive to faces in the temporal cortex. Philosophical Transactions of the Royal Society 1992;355:23–30.

Poggio T, Edelman S. A network that learns to recognize three-dimensional objects. Nature 1990;343:263–266. [PubMed: 2300170]

Potter M, Levy E. Recognition memory for a rapid sequence of pictures. Journal of Experimental Psychology 1969;81:10–15. [PubMed: 5812164]

Privman E, Nir Y, Kramer U, Kipervasser S, Andelman F, Neufeld M, Mukamel R, Yeshurun Y, Fried I, Malach R. Enhanced category tuning revealed by iEEG in high order human visual areas. Journal of Neuroscience 2007;6:6234–6242. [PubMed: 17553996]

Quian Quiroga R, Reddy L, Kreiman G, Koch C, Fried I. Invariant visual representation by single neurons in the human brain. Nature 2005;435:1102–1107. [PubMed: 15973409]

Rayner K. Eye movements in reading and information processing: 20 years of research. Psychol Bull 1998;124:372–422. [PubMed: 9849112]

Richmond B, Wurtz R, Sato T. Visual responses in inferior temporal neurons in awake Rhesus monkey. Journal of Neurophysiology 1983;50:1415–1432. [PubMed: 6663335]

Riesenhuber M, Poggio T. Hierarchical models of object recognition in cortex. Nature Neuroscience 1999;2:1019–1025.

Rolls E. Neural organization of higher visual functions. Current Opinion in Neurobiology 1991;1:274–278. [PubMed: 1821190]

Schmolesky M, Wang Y, Hanes D, Thompson K, Leutgeb S, Schall J, Leventhal A. Signal timing across the macaque visual system. Journal of Neurophysiology 1998;79:3272–3278. [PubMed: 9636126]

Serre T, Kreiman G, Kouh M, Cadieu C, Knoblich U, Poggio T. A quantitative theory of immediate visual recognition. Progress In Brain Research 2007a;165C:33–56.

Serre T, Oliva A, Poggio T. Feedforward theories of visual cortex account for human performance in rapid categorization. PNAS 2007b;104:6424–6429. [PubMed: 17404214]

Sigala N, Logothetis N. Visual categorization shapes feature selectivity in the primate temporal cortex. Nature 2002;415:318–320. [PubMed: 11797008]

Sugase Y, Yamane S, Ueno S, Kawano K. Global and fine information coded by single neurons in the temporal visual cortex. Nature 1999;400:869–873. [PubMed: 10476965]

Suzuki W, Miller E, Desimone R. Object and place memory in the macaque entorhinal cortex. Journal of Neurophysiology 1997;78:1062–1081. [PubMed: 9307135]

Tanaka K. Inferotemporal cortex and object vision. Annual Review of Neuroscience 1996;19:109–139.

Thorpe S, Fize D, Marlot C. Speed of processing in the human visual system. Nature 1996;381:520–522. [PubMed: 8632824]

Tippett L, Glosser G, Farah M. A category-specific naming impairment after temporal lobectomy. Neuropsychologia 1996;34:139–146. [PubMed: 8852876]

Tomita H, Ohbayashi M, Nakahara K, Hasegawa I, Miyashita Y. Top-down signal from prefrontal cortex in executive control of memory retrieval. Nature 1999;401:699–703. [PubMed: 10537108]

Tsao DY, Freiwald WA, Tootell RB, Livingstone MS. A cortical region consisting entirely of face-selective cells. Science 2006;311:670–674. [PubMed: 16456083]

vanRullen R, Thorpe S. Surfing a Spike Wave down the Ventral Stream. Vision Research 2002;42:2593–2615. [PubMed: 12446033]

Wallis G, Rolls ET. Invariant face and object recognition in the visual system. Progress in Neurobiology 1997;51:167–194. [PubMed: 9247963]
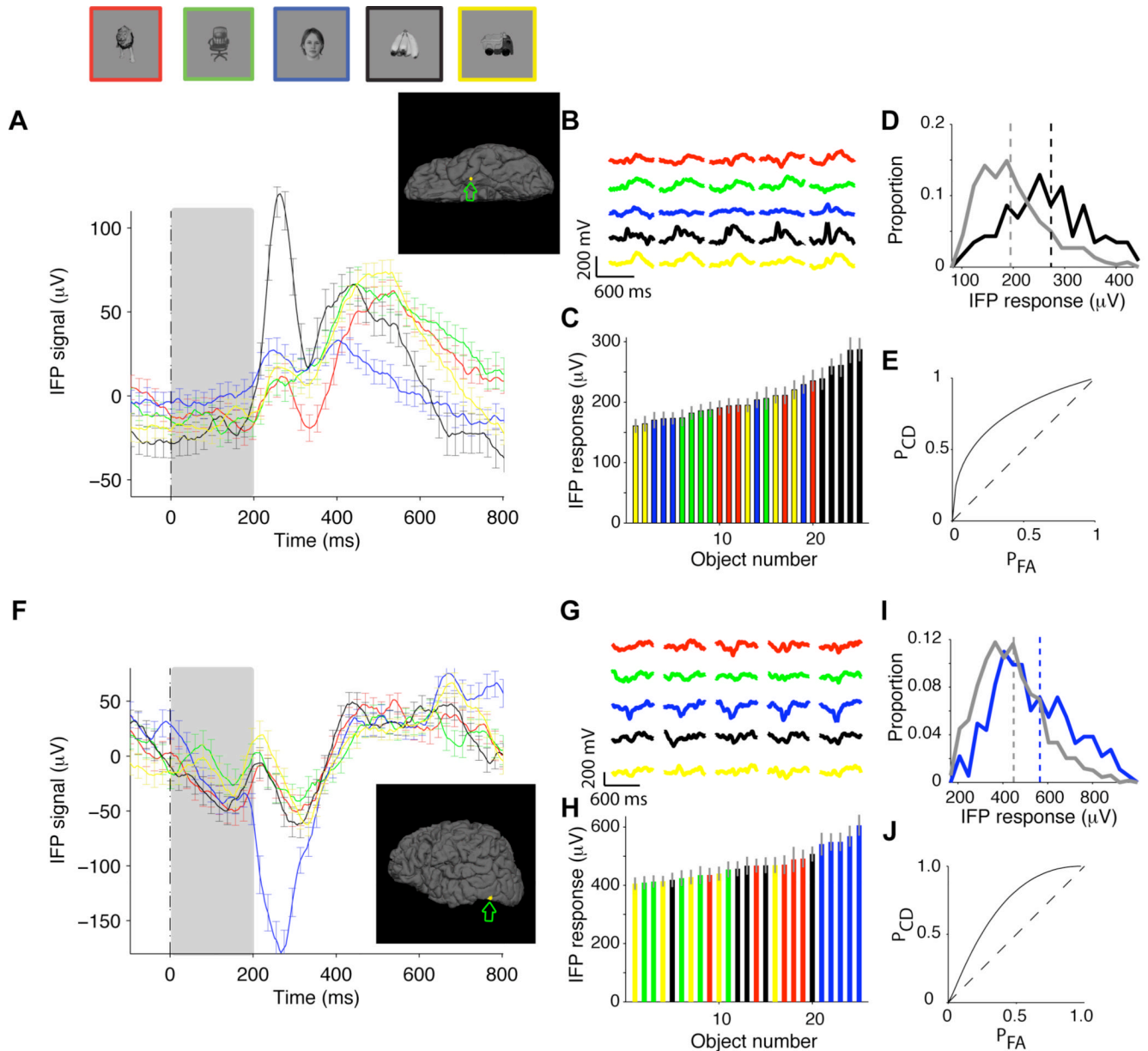
Figure 1. Selectivity in intracranial field potential (IFP) recordings from the human cortex
A. Example showing the responses of an electrode in the left temporal pole (Talairach coordinates: −27.0, −11.8, −35.3; see inset depicting the electrode position). The plot shows the average IFP response to each of 5 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 corresponds to the average over 5 exemplars; see Figure S1**B** for the 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 every 10 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 the object category (see **A**) and are aligned to stimulus onset. **C**. The *IFP response* was defined as the range of the signal (max(IFP)-min(IFP)) in the 50 to 300 ms interval post-stimulus onset (see Figure S3 and

Supplementary Material for other IFP response definitions). A one-way ANOVA on the IFP response yielded $p<10^{-6}$. Average IFP response (±SEM) to each of the 25 exemplar objects. The responses are color-coded based on the object category and the responses are sorted by their magnitude. Note that the 5 images containing "fruits" represent the top 5 responses (bootstrap $p<10^{-5}$, Experimental Procedures). **D**. Distribution of IFP responses for the 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±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±5%.
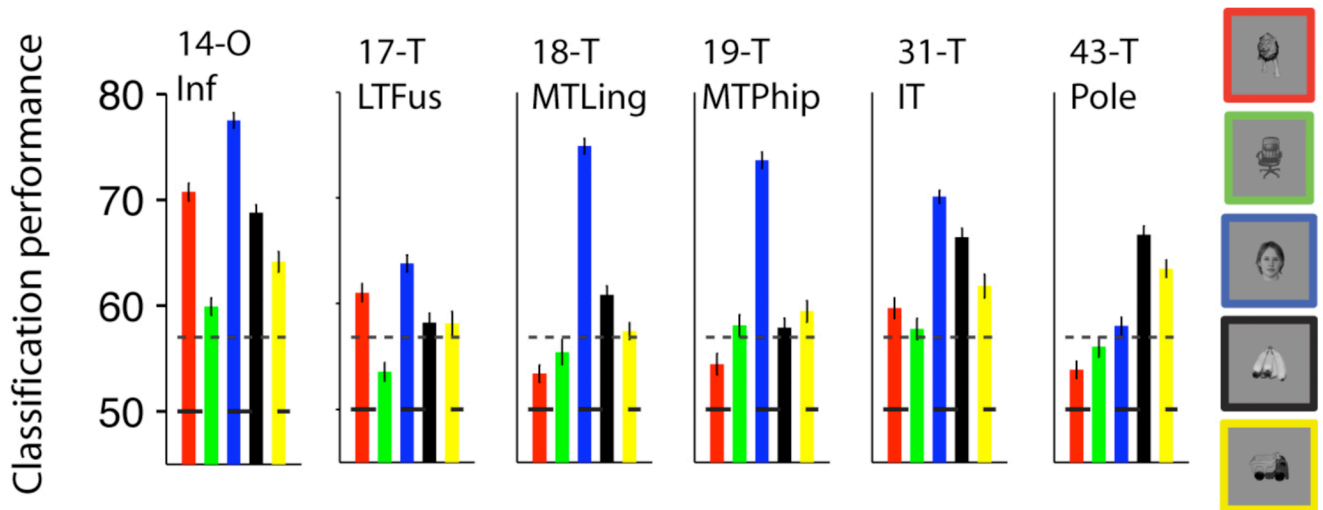
Figure 2. Locations showing strongest selectivity
Classification performance using neural ensembles from different brain locations. We randomly selected 10 electrodes from the pool of all the electrodes in each location. The subplots show the classification performance using the responses from these neural ensembles for the 6 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 S2**A**).
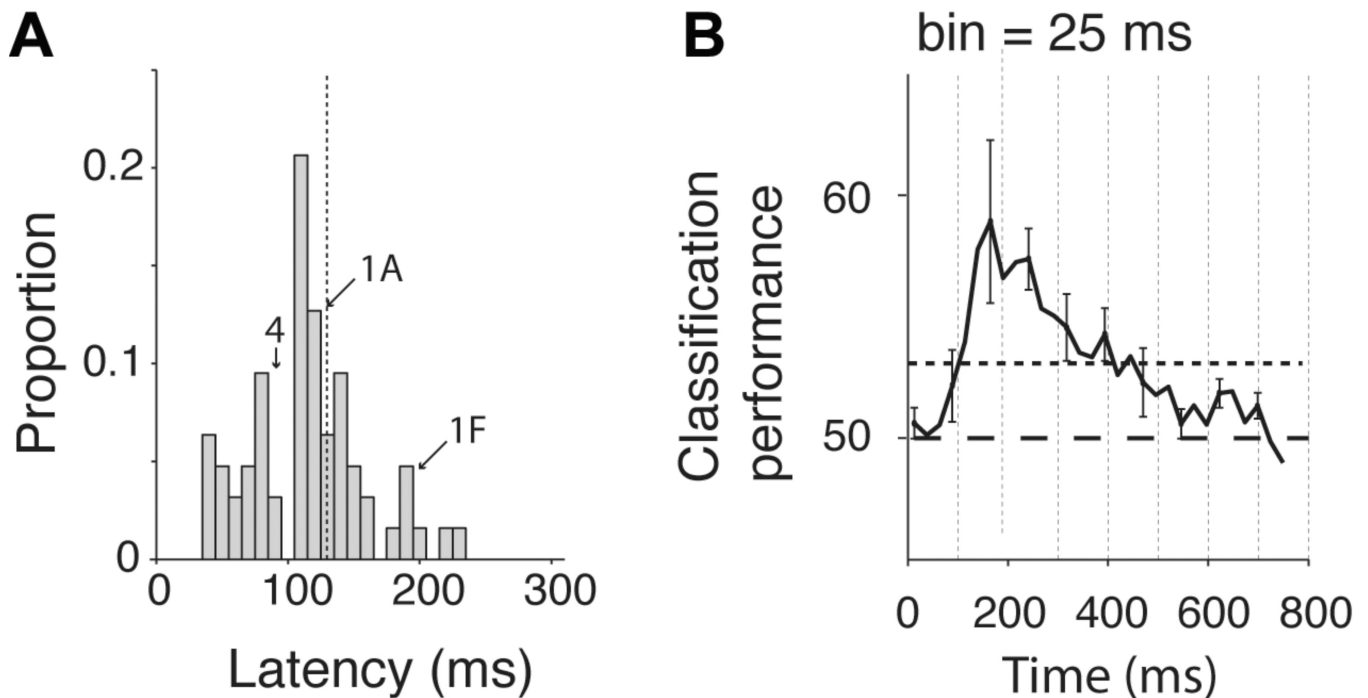
Figure 3. Fast decoding of object category
A. Distribution of IFP latencies for all the electrodes that showed selective responses (*n*=94, see text). The latency was defined as the first time point where 10 consecutive points yielded *p* < 0.01 in a one-way ANOVA across object categories (see Experimental Procedures and Figure S6**A**). The dashed line shows the mean of the distribution and the arrows show the position of the example electrodes in Figure 1**A**, 1**F** and 4. **B** Average classification performance as a function of time from stimulus onset. We built a neural ensemble vector that contained the IFP power in individual bins of 25 ms duration (12 sampling points at BWH, 6 sampling points at CHB) using 11 electrodes with the highest r_v value (ratio of variance *across* categories to variance *within* categories, computed using only the training data and using the 50 to 300 ms interval, see Experimental Procedures and (Hung et al., 2005)). This figure shows the classification performance as a function of time from stimulus onset averaged over all 5 categories (see Figure S7 **A–C** 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.
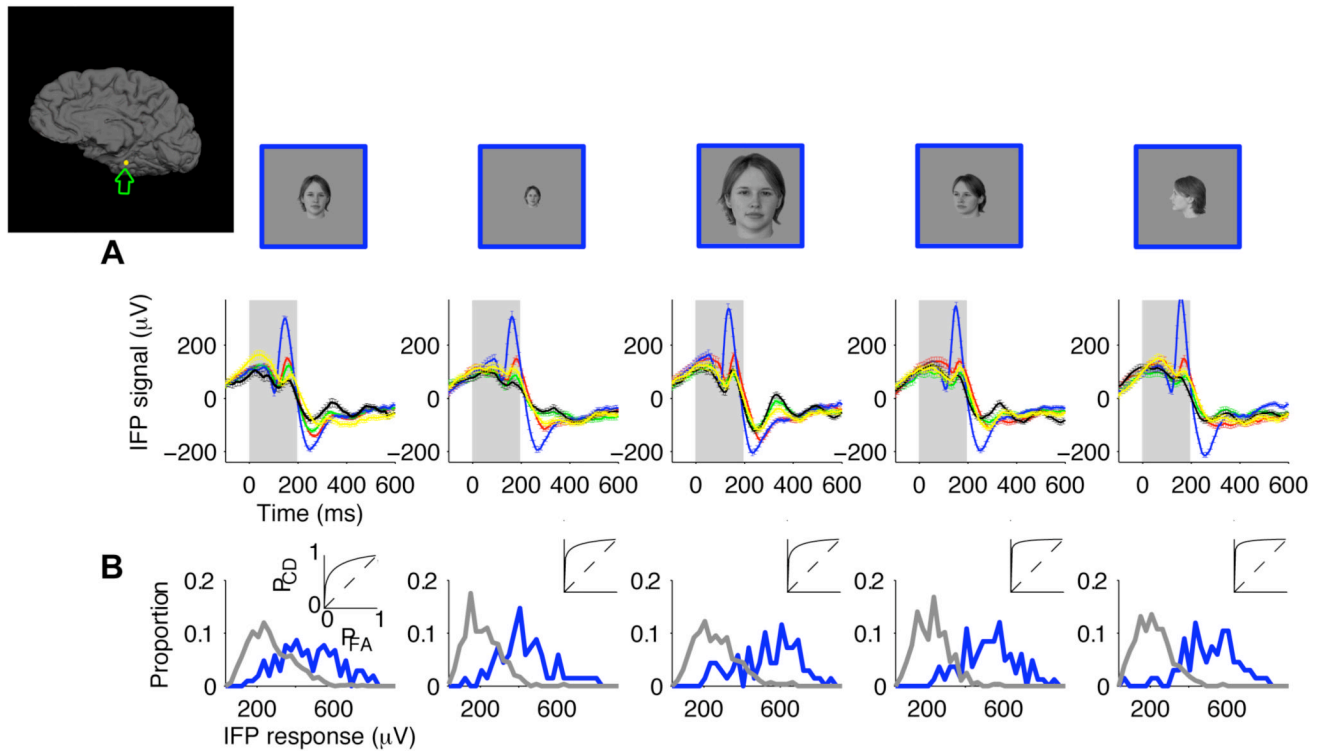
Figure 4. Robustness to scale and depth rotation
A. Responses of an electrode located in the parahippocampal part of the right medio-temporal-gyrus (Talairach coordinates: 32.6, −34.8, −13.6; see inset depicting the electrode position). Each column depicts the IFP responses to images where the objects 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 S1**B** for the "default" viewpoint for the other categories) at a scale of 3 degrees (column 1, total n=329), with the same default viewpoint at a scale of 1.5 degrees (column 2, total n=359), with the same default viewpoint at a scale of 6 degrees (column 3, total n=363), at a scale of 3 degrees and ~45 degree depth rotation (column 4, total n=370) and at a scale of 3 degrees and ~90 degree 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 corresponds to the average over all 5 exemplars for each category. The format for the response subplots is the same as in Figure 1**A**. Error bars are 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 to 300 ms window) for "human faces" (blue) versus other object categories (gray). The insets show the corresponding ROC curves.
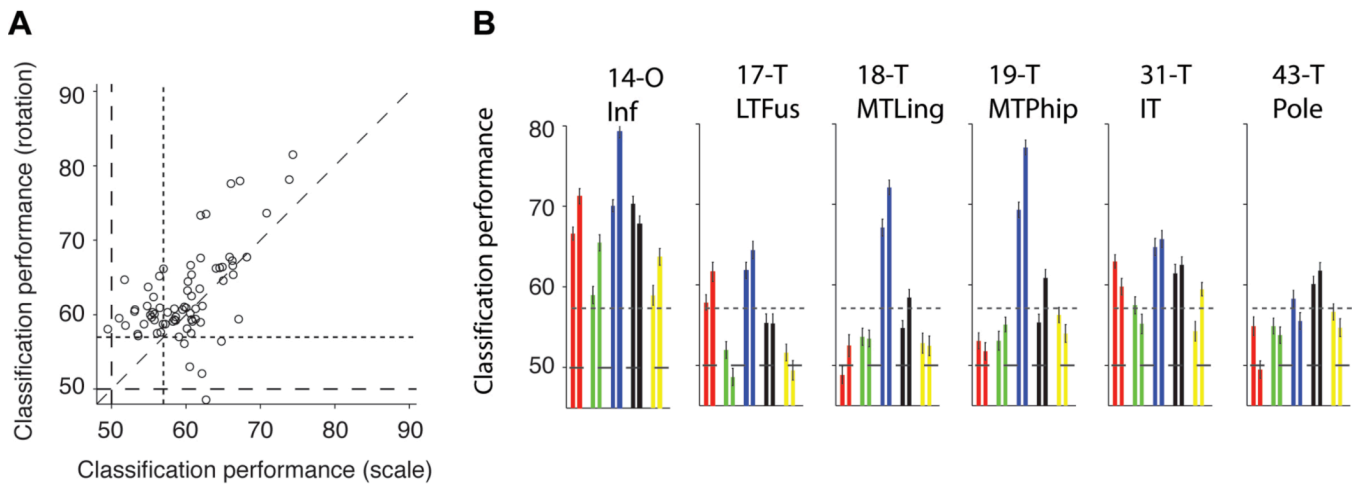
Figure 5. Locations showing strongest robustness to object changes

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). The corresponding results for all locations are shown in Figure S5 **D–F** .
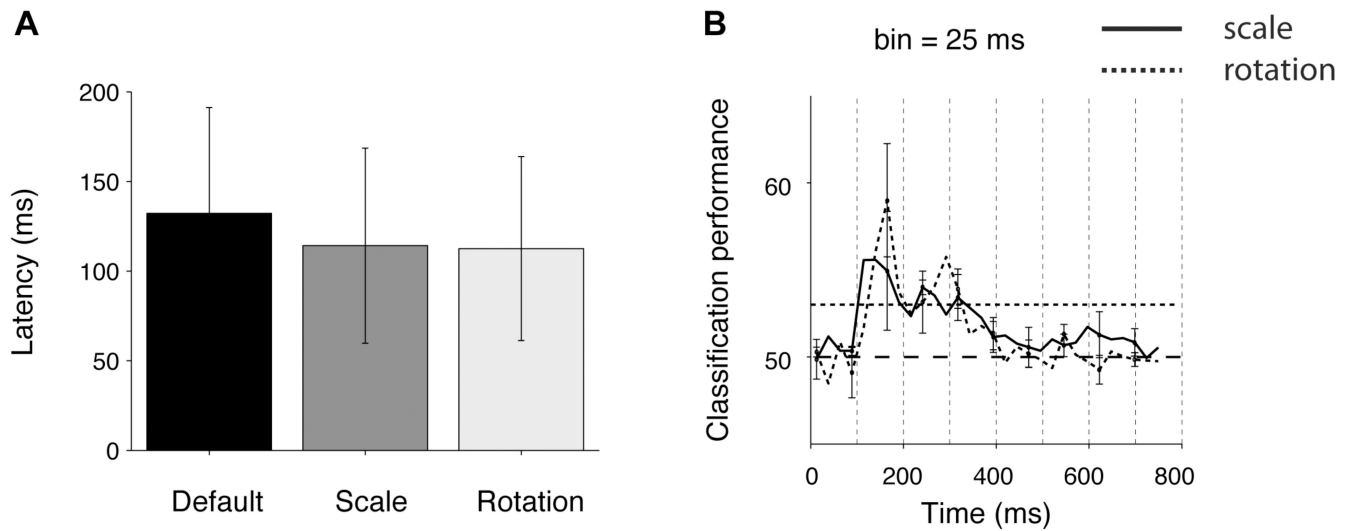
Figure 6. Fast and robust decoding of object category
A. Latencies for the electrodes that showed IFP responses invariant to scale or rotation changes (see Experimental Procedures and Figure S6**A** for the definition of latency). Error bars are one SD. **B**. Dynamics of classification performance. The format and conventions are the same as in Figure 3**B**. 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 5 object categories (see Figure S7 **D–F** for the corresponding results for each individual category and different bin sizes).