

Spatiotemporal Dynamics Underlying Object Completion in Human Ventral Visual Cortex

Hanlin Tang,^{1,2} Calin Buia,² Radhika Madhavan,² Nathan E. Crone,³ Joseph R. Madsen,⁴ William S. Anderson,⁵ and Gabriel Kreiman^{1,2,6,*}

¹Program in Biophysics, Harvard University, Boston, MA 02115, USA

²Department of Ophthalmology, Boston Children's Hospital, Harvard Medical School, Boston, MA 02115, USA

³Department of Neurology, Johns Hopkins School of Medicine, Baltimore, MD 21287, USA

⁴Department of Neurosurgery, Boston Children's Hospital, Harvard Medical School, Boston, MA 02115, USA

⁵Department of Neurosurgery, Johns Hopkins School of Medicine, Baltimore, MD 21287, USA

⁶Center for Brain Science, Harvard University, Boston, MA 02115, USA

*Correspondence: gabriel.kreiman@childrens.harvard.edu

<http://dx.doi.org/10.1016/j.neuron.2014.06.017>

SUMMARY

Natural vision often involves recognizing objects from partial information. Recognition of objects from parts presents a significant challenge for theories of vision because it requires spatial integration and extrapolation from prior knowledge. Here we recorded intracranial field potentials of 113 visually selective electrodes from epilepsy patients in response to whole and partial objects. Responses along the ventral visual stream, particularly the inferior occipital and fusiform gyri, remained selective despite showing only 9%–25% of the object areas. However, these visually selective signals emerged ~100 ms later for partial versus whole objects. These processing delays were particularly pronounced in higher visual areas within the ventral stream. This latency difference persisted when controlling for changes in contrast, signal amplitude, and the strength of selectivity. These results argue against a purely feedforward explanation of recognition from partial information, and provide spatiotemporal constraints on theories of object recognition that involve recurrent processing.

INTRODUCTION

During natural viewing conditions, we often have access to only partial information about objects due to limited viewing angles, poor luminosity, or occlusion. How the visual system can recognize objects from limited information while still maintaining fine discriminability between like objects remains poorly understood and represents a significant challenge for computer vision algorithms and theories of vision.

Visual shape recognition is orchestrated by a cascade of processing steps along the ventral visual stream (Connor et al., 2007; Logothetis and Sheinberg, 1996; Rolls, 1991; Tanaka, 1996). Neurons in the highest echelons of the macaque monkey

ventral stream, the inferior temporal cortex (ITC), demonstrate strong selectivity to complex objects (Ito et al., 1995; Miyashita and Chang, 1988; Richmond et al., 1983; Rolls, 1991). In the human brain, several areas within the occipital-temporal lobe showing selective responses to complex shapes have been identified using neuroimaging (Grill-Spector and Malach, 2004; Haxby et al., 1991; Kanwisher et al., 1997; Taylor et al., 2007) and invasive physiological recordings (Allison et al., 1999; Liu et al., 2009; Privman et al., 2007). Converging evidence from behavioral studies (Kirchner and Thorpe, 2006; Thorpe et al., 1996), human scalp electroencephalography (Thorpe et al., 1996), and monkey (Hung et al., 2005; Keyser et al., 2001; Optican and Richmond, 1987) and human (Allison et al., 1999; Liu et al., 2009) neurophysiological recordings have established that selective responses to and rapid recognition of isolated whole objects can occur within 100 ms of stimulus onset. As a first-order approximation, the speed of visual processing suggests that initial recognition may occur in a largely feedforward fashion, whereby neural activity progresses along the hierarchical architecture of the ventral visual stream with minimal contributions from feedback connections between areas or within-area recurrent computations (Deco and Rolls, 2004; Fukushima, 1980; Riesenhuber and Poggio, 1999).

Recordings in ITC of monkeys (Desimone et al., 1984; Hung et al., 2005; Ito et al., 1995; Logothetis and Sheinberg, 1996) and humans (Liu et al., 2009) have revealed a significant degree of tolerance to object transformations. Visual recognition of isolated objects under certain transformations such as scale or position changes does not incur additional processing time at the behavioral or physiological level (Biederman and Cooper, 1991; Desimone et al., 1984; Liu et al., 2009; Logothetis et al., 1995) and can be described using purely bottom-up computational models. While bottom-up models may provide a reasonable approximation for rapid recognition of whole isolated objects, top-down as well as horizontal projections abound throughout visual cortex (Callaway, 2004; Felleman and Van Essen, 1991). The contribution of these projections to the strong robustness of object recognition to various transformations remains unclear. In particular, recognition of objects from partial information is a difficult problem for purely feedforward architectures and may involve significant contributions from recurrent connections as

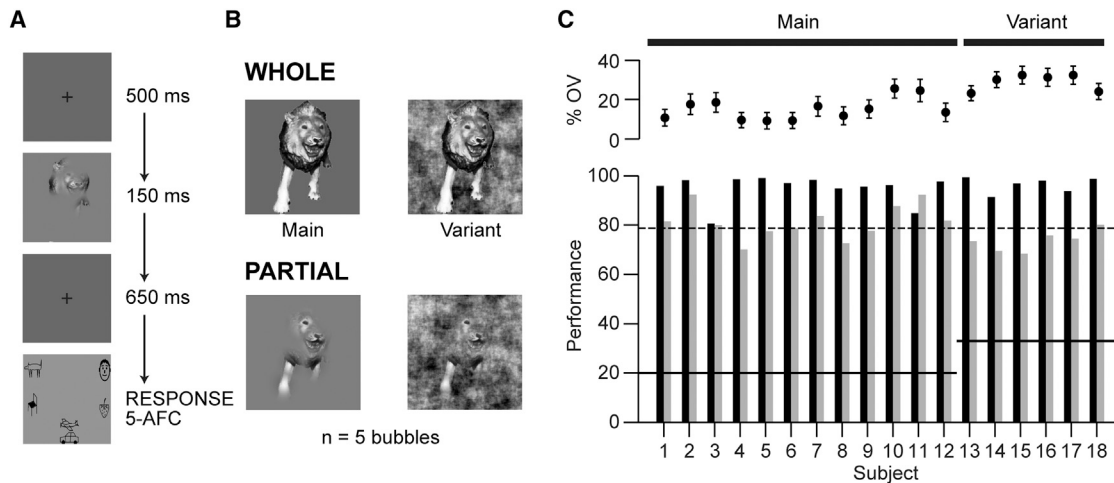


Figure 1. Experimental Design and Behavioral Performance

(A) After 500 ms fixation, an image containing a whole object or a partial object was presented for 150 ms. Subjects categorized objects into one of five categories (5-alternative forced choice) following a choice screen. Presentation order was pseudorandomized.

(B) Example images used in the task. Objects were either unaltered (whole) or presented through Gaussian bubbles (partial). For 12 subjects the background was a gray screen (Main experiment) and for 6 subjects the background was phase-scrambled noise (Variant experiment). In this example, the object is seen through five bubbles (18% of object area shown). The number of bubbles was titrated for each subject to achieve 80% performance. Stimuli consisted of 25 different objects belonging to five categories (for more examples, see Figure S1).

(C) Above, percentage of the object visible (mean \pm SD) for each subject in the Main experiment (left) and the contrast-normalized Variant (right). Below, percentage of correct trials (performance) for whole (black) and partial (gray) objects.

shown in attractor networks (Hopfield, 1982; O'Reilly et al., 2013) or Bayesian inference models (Lee and Mumford, 2003).

Previous studies have examined the brain areas involved in pattern completion with human neuroimaging (Lerner et al., 2004; Schiltz and Rossion, 2006; Taylor et al., 2007), the selectivity of physiological signals elicited by partial objects (Issa and Dicarlo, 2012; Kovacs et al., 1995; Nielsen et al., 2006; Ruffisshauer et al., 2011), and behavioral delays when recognizing occluded or partial objects (Biederman, 1987; Brown and Koch, 2000; Johnson and Olshausen, 2005). Several studies have focused on “amodal completion,” i.e., the linking of disconnected parts to a single “gestalt,” using geometric shapes or line drawings and strong occluders that provided depth cues (Brown and Koch, 2000; Chen et al., 2010; Johnson and Olshausen, 2005; Murray et al., 2001; Nakayama et al., 1995; Sehatpour et al., 2008). In addition to determining that different parts belong to a whole, the brain has to jointly process the parts to recognize the object (Gosselin and Schyns, 2001; Nielsen et al., 2006; Ruffisshauer et al., 2011), which we study here.

We investigated the spatiotemporal dynamics underlying object completion by recording field potentials from intracranial electrodes implanted in epilepsy patients while subjects recognized objects from partial information. Even with very few features present (9%–25% of object area shown), neural responses in the ventral visual stream retained object selectivity. These visually selective responses to partial objects emerged about 100 ms later than responses to whole objects. The processing delays associated with interpreting objects from partial information increased along the visual hierarchy. These delays stand in contrast to position and scale transformations. Together, these results argue against a feedforward explanation for recognition

of partial objects and provide evidence for the involvement of highest visual areas in recurrent computations orchestrating pattern completion.

RESULTS

We recorded intracranial field potentials (IFPs) from 1,699 electrodes in 18 subjects (11 male, 17 right handed, 8–40 years old) implanted with subdural electrodes to localize epileptic seizure foci. Subjects viewed images containing grayscale objects presented for 150 ms. After a 650 ms delay period, subjects reported the object category (animals, chairs, human faces, fruits, or vehicles) by pressing corresponding buttons on a gamepad (Figure 1A). In 30% of the trials, the objects were unaltered (referred to as the “Whole” condition). In 70% of the trials, partial object features were presented through randomly distributed Gaussian “bubbles” (Figure 1B, Experimental Procedures, referred to as the “Partial” condition) (Gosselin and Schyns, 2001). The number of bubbles was calibrated at the start of the experiment such that performance was \sim 80% correct. The number of bubbles (but not their location) was then kept constant throughout the rest of the experiment. For 12 subjects, the objects were presented on a gray background (the “Main” experiment). While contrast was normalized across whole objects, whole objects and partial objects had different contrast levels because of the gray background. In six additional subjects, a modified experiment (the “Variant” experiment) was performed where contrast was normalized between whole and partial objects by presenting objects on a background of phase-scrambled noise (Figure 1B).

The performance of all subjects was around the target correct rate (Figure 1C, $79\% \pm 7\%$, mean \pm SD). Performance was

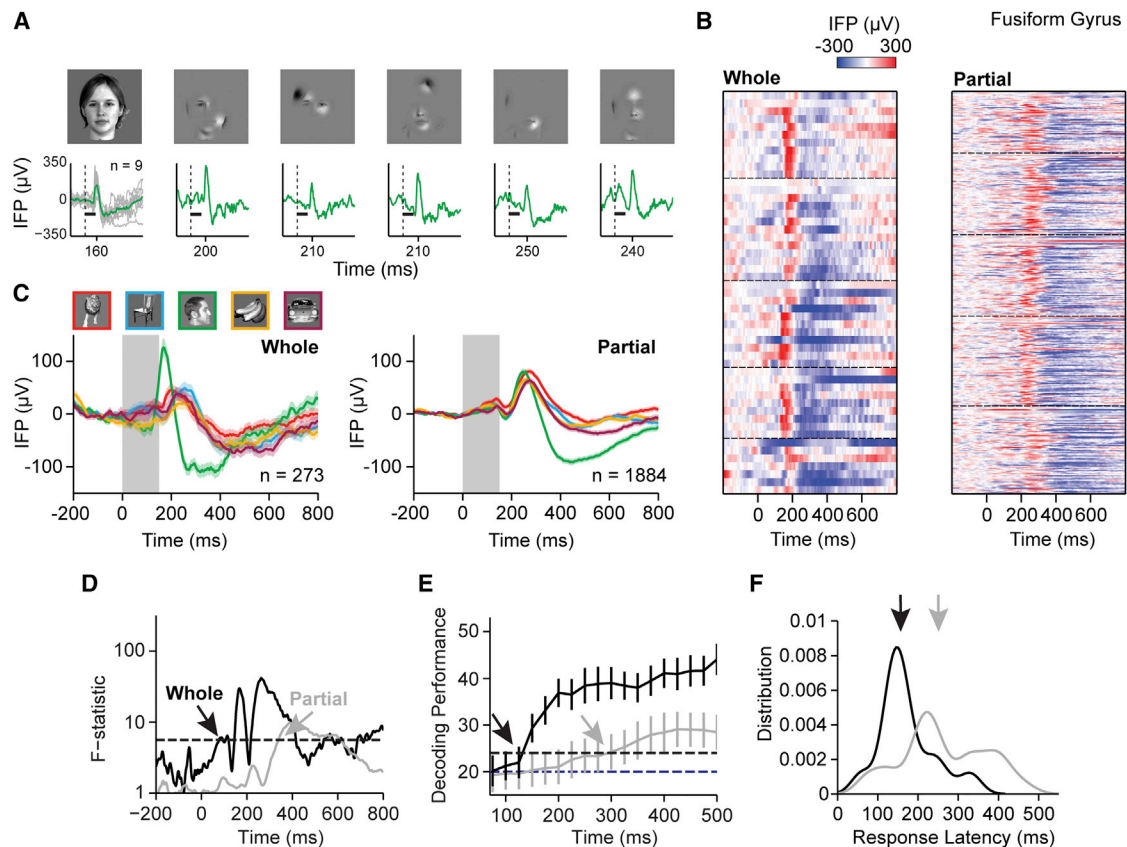


Figure 2. Example Physiological Responses from Main Experiment

Example responses from an electrode in the left fusiform gyrus.

(A) Intracranial field potential (IFP) responses to an individual exemplar object. For the Whole condition, the average response (green) and single trial traces (gray) are shown. For the Partial condition, example single trial responses (green, $n = 1$) to different partial images of the same exemplar (top row) are shown. The response peak time is marked on the x axis. The dashed line indicates the stimulus onset time, and the black bar indicates stimulus presentation duration (150 ms).

(B) Raster of the neural responses for whole (left, 52 trials) and partial (right, 395 trials) objects for the category that elicited the strongest responses (human faces). Rows represent individual trials. Dashed lines separate responses to the five face exemplars. The color indicates the IFP at each time point (bin size = 2 ms, see scale on top).

(C) Average IFP response to whole (left) and partial (right) objects belonging to five different categories (animals, chairs, human faces, fruits, and vehicles, see color map on top). Shaded areas around each line indicate SEM. The gray rectangle denotes the image presentation time (150 ms). The total number of trials is indicated on the bottom right of each subplot.

(D) Selectivity was measured by computing the F-statistic at each time point for whole (black) and partial (gray) objects. Arrows indicate the first time point when the F-statistic was greater than the statistical threshold (black dashed line) for 25 consecutive milliseconds.

(E) Decoding performance (mean \pm SD) using a linear multiclass discriminant algorithm in classifying trials into one of five categories. Arrows indicate the first time when decoding performance reached the threshold for statistical significance (black dashed line). Chance is 20% (blue dashed line).

(F) Distribution of the visual response latency across trials for whole (black) and partial (gray) objects, based on when the IFP in individual trials was significantly above baseline activity. The distribution is based on kernel density estimate (bin size = 6 ms). The arrows denote the distribution averages.

significantly above chance (Main experiment: chance = 20%, 5-alternative forced choice; Variant experiment: chance = 33%, 3-alternative forced choice) even when only 9%–25% of the object was visible. As expected, performance for the Whole condition was near ceiling ($95\% \pm 5\%$, mean \pm SD). The analyses in this manuscript were performed on correct trials only.

Object Selectivity Was Retained despite Presenting Partial Information

Consistent with previous studies, multiple electrodes showed strong visually selective responses to whole objects (Allison

et al., 1999; Davidesco et al., 2013; Liu et al., 2009). An example electrode from the Main experiment, located in the fusiform gyrus, had robust responses to several exemplars in the Whole condition, such as the one illustrated in the first panel of Figure 2A. These responses could also be observed in individual trials of face exemplars (gray traces in Figures 2A and 2B, left). This electrode was preferentially activated in response to faces compared to the other objects (Figure 2C, left). Responses to stimuli other than human faces were also observed, such as the responses to several animal (red) and fruit (orange) exemplars (Figure S1B available online).

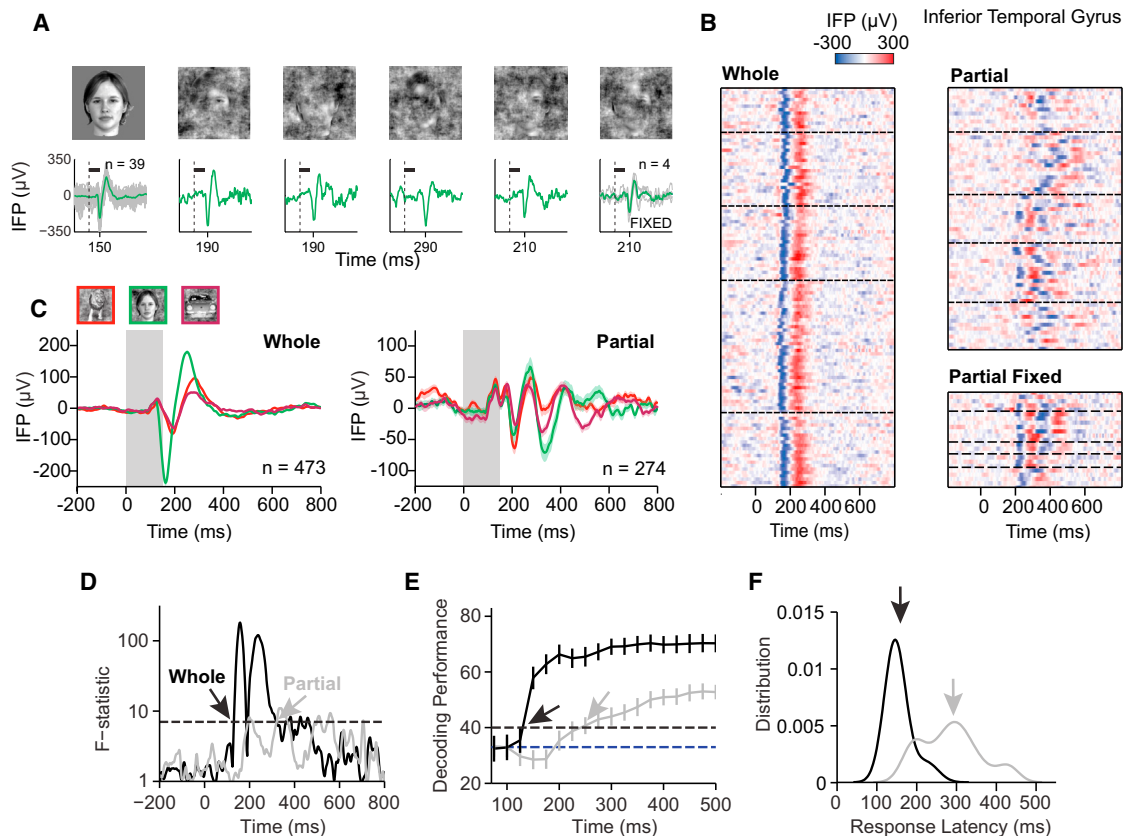


Figure 3. Second Example of Physiological Responses from Variant Experiment

Example responses from an electrode in the left inferior temporal gyrus. The format and conventions are as in Figure 2, except that only three categories were tested, and the Partial Fixed condition was added in (A) and (B) (Experimental Procedures). Note that the statistical thresholds for the F-statistic and decoding performance differ from those in Figure 2 because of the different number of categories. More examples are shown in Figures S2 and S3.

The responses in this example electrode were preserved in the Partial condition, where only $11\% \pm 4\%$ (mean \pm SD) of the object was visible. Robust responses to partial objects were observed in single trials (Figures 2A and 2B, right). These responses were similar even when largely disjoint sets of features were presented (e.g., compare Figure 2A, third and fourth images). Because the bubble locations varied from trial to trial, there was significant variability in the latency of the visual response (Figure 2B, right); this variability affected the average responses to each category of partial objects (Figure 2C, right). Despite this variability, the electrode remained selective and kept the stimulus preferences at the category and exemplar level (Figures 2C and S1B).

The responses of an example electrode from the Variant experiment support similar conclusions (Figure 3). Even though only $21\% \pm 4\%$ (mean \pm SD) of the object was visible, there were robust responses in single trials (Figures 3A and 3B), and strong selectivity both for whole objects and partial objects at the category and exemplar level (Figures 3C and S1C). While the selectivity was consistent across single trials, there was significantly more trial-to-trial variation in the timing of the responses to partial objects compared to whole objects (Figure 3B, top right).

To measure the strength of selectivity, we employed two approaches. The first approach (“ANOVA”) was a nonparametric one-way analysis of variance test to evaluate whether and when the average category responses differed significantly. An electrode was denoted “selective” if, during 25 consecutive milliseconds, the ratio of variances across versus within categories (F-statistic) was greater than a significance threshold determined by a bootstrapping procedure to ensure a false discovery rate of $q < 0.001$ ($F = 5.7$) (Figures 2D and 3D). Similar results were obtained when considering d' as a measure of selectivity (Experimental Procedures). The ANOVA test evaluates whether the responses are statistically different when averaged across trials, but the brain needs to discriminate among objects in single trials. To evaluate the degree of selectivity in single trials, we employed a statistical learning approach to measure when information in the neural response became available to correctly classify the object into one of the five categories (denoted “Decoding”; Figure 2E, chance = 20%; Figure 3E, chance = 33%). An electrode was considered selective if the decoding performance exceeded a threshold determined to ensure $q < 0.001$ (Experimental Procedures).

Of the 1,699 electrodes, 210 electrodes (12%) and 163 electrodes (10%) were selective during the Whole condition in the

Table 1. Number of Selective Electrodes

Experiment	Frequency Band	Whole	Shuffled	Both	Shuffled	Figures
Main	broadband	83	(1.66 ± 0.07)	22	(0.06 ± 0.01)	Figures 4, 5A–5E, and 6
Variant	broadband	30	(1.12 ± 0.12)	8	(0.04 ± 0.03)	Figures 4E, 4F, 5A–5C, and 6
Main	gamma	53	(1.56 ± 0.05)	14	(0.04 ± 0.01)	Figures 5F and 6D

For the experiment and frequency bands reported in the main text, this table shows the number of electrodes selective to whole images (“Whole”) or to both whole and partial images (“Both”). Also reported is the average number of selective electrodes found when the object category labels were shuffled (mean ± SEM, n = 1,000 iterations).

ANOVA and Decoding tests, respectively. We focused subsequent analyses only on the 113 electrodes selective in both tests (83 from the Main experiment and 30 from the Variant experiment; Table 1). As a control, shuffling the object labels yielded only 2.78 ± 0.14 selective electrodes (mean ± SEM, 1,000 iterations; 0.16% of the total). Similar to previous reports, the preferred category of different electrodes spanned all five object categories, and the electrode locations were primarily distributed along the ventral visual stream (Figures 4E and 4F) (Liu et al., 2009). As demonstrated for the examples in Figures 2 and 3, 30 electrodes (24%) remained visually selective in the Partial condition (Main experiment: 22; Variant experiment: 8), whereas the shuffling control yielded an average of 0.06 and 0.04 selective electrodes in the Main and Variant experiments, respectively (Table 1).

The examples in Figures 2C and 3C seem to suggest that the response amplitudes were larger in the Whole condition. However, this effect was due to averaging over trials and the increased trial-to-trial variability in the response latency for the Partial condition. No amplitude changes are apparent in the single trial data (Figures 2B and 3B). The range of the IFP responses to the preferred category from 50 to 500 ms was not significantly different for whole versus partial objects (Figure 4A, $p = 0.68$, Wilcoxon rank-sum test). However, the strength of category selectivity was suppressed in the Partial condition. The median F-statistic was 23 for the Whole condition and 14 for the Partial condition (Figure 4B, $p < 10^{-4}$, Wilcoxon signed-rank test, an F-statistic value of 1 indicates no selectivity). The median decoding performance was 33% for the Whole condition and 26% for the Partial condition (Figure 4C, $p < 10^{-4}$, Wilcoxon signed-rank test). Because the Variant experiment contained only three categories, measures of selectivity such as the F-statistic or decoding performance are scaled differently from the Main experiment, so Figures 4A–4D only show data from the Main experiment. Analysis of the electrodes in the Variant experiment revealed similar conclusions.

The observation that even nonoverlapping sets of features can elicit robust responses (e.g., third and fourth panel in Figure 2A) suggests that the electrodes tolerated significant trial-to-trial variability in the visible object fragments. To quantify this observation across the population, we defined the percentage of overlap between two partial images of the same object by computing the number of pixels shared by the image pair divided by the object area (Figure 4D, insert). We considered partial images where the response to the preferred category was highly discriminable from the response to the nonpreferred categories (Experimental Procedures). Even for these trials with

robust responses, 45% of the 10,438 image pairs had less than 5% overlap, and 11% of the pairs had less than 1% overlap (Figure 4D). Furthermore, in every electrode, there existed pairs of robust responses where the partial images had <1% overlap.

To compare different brain regions, we measured the percentage of electrodes in each gyrus that were selective in either the Whole condition or in both conditions (Figures 4E and 4F). Consistent with previous reports, electrodes along the ventral visual stream were selective in the Whole condition (Figure 4F, black bars) (Allison et al., 1999; Davidesco et al., 2013; Liu et al., 2009). The locations with the highest percentages of electrodes selective to partial objects were primarily in higher visual areas, such as the fusiform gyrus and inferior occipital gyrus (Figure 4F, gray bars, $p = 2 \times 10^{-6}$ and 5×10^{-4} , respectively, Fisher’s exact test). In sum, electrodes in the highest visual areas in the human ventral stream retained visual selectivity to partial objects, their responses could be driven by disjoint sets of object parts, and the response amplitude but not the degree of selectivity was similar to that of whole objects.

Delayed Responses to Partial Objects

In addition to the changes in selectivity described above, the responses to partial objects were delayed compared to the corresponding responses to whole objects (e.g., compare Whole versus Partial in the single trial responses in Figures 2A, 2B, 3A, and 3B). To compare the latencies of responses to whole and partial objects, we measured both selectivity latency and visual response latency. Selectivity latency indicates when sufficient information becomes available to distinguish among different objects or object categories, whereas the response latency denotes when the visual response differs from baseline (Experimental Procedures).

Quantitative estimates of latency are difficult because they depend on multiple variables, including number of trials, response amplitudes, and thresholds. Here we independently applied different measures of latency to the same data set. The selectivity latency in the responses to whole objects for the electrode shown in Figure 2 was 100 ± 8 ms (mean ± 99% CI) based on the first time point when the F-statistic crossed the statistical significance threshold (Figure 2D, black arrow). The selectivity latency for the partial objects was 320 ± 6 ms (mean ± 99% CI), a delay of 220 ms. A comparable delay of 180 ms between Partial and Whole conditions was obtained using the single trial decoding analyses (Figure 2E). Similar delays were apparent for the example electrode in Figure 3.

We considered all electrodes in the Main experiment that showed selective responses to both whole objects and partial

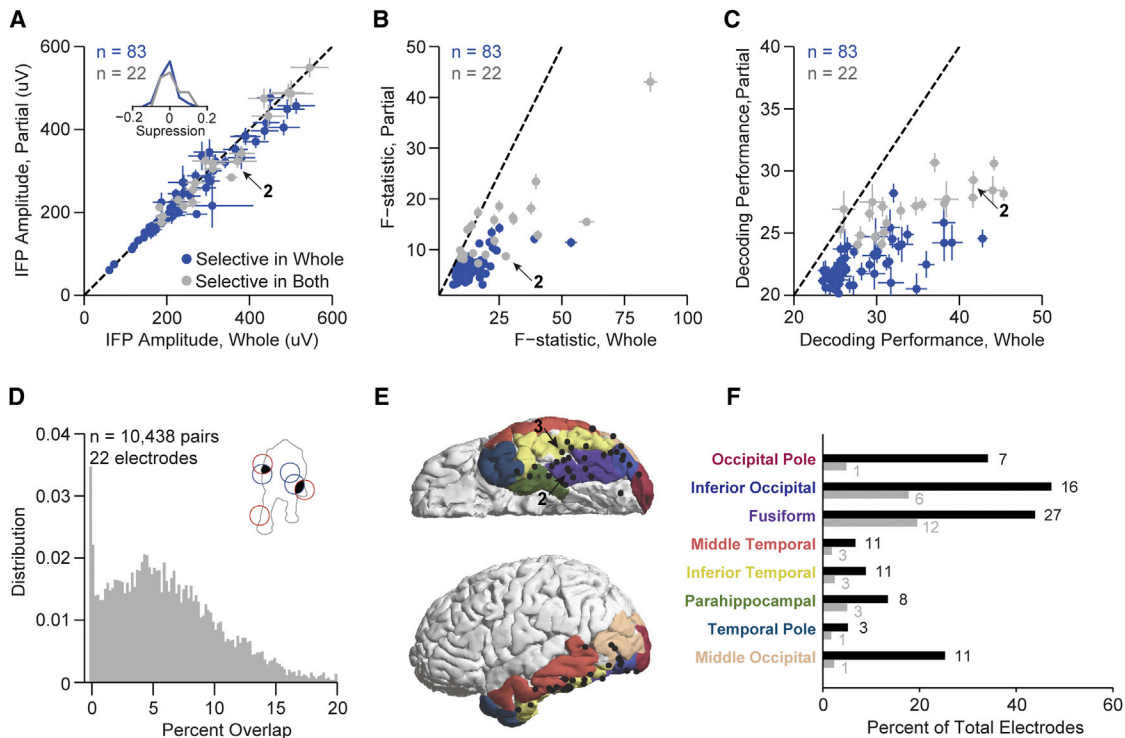


Figure 4. Neural Responses Remained Visually Selective despite Partial Information

(A) Average IFP amplitude $A = (1/N) \sum_{i=1}^N \max(IFP_i(t)) - \min(IFP_i(t))$ across trials (N) in response to partial versus whole objects for electrodes that were visually selective in the Whole condition (blue, $n = 61 + 22$), and electrodes that were visually selective in both Whole and Partial conditions (gray, $n = 22$) (Main experiment). Most of the data clustered around the diagonal (dashed line). Inset, distribution of suppression index: $(A_{\text{whole}} - A_{\text{partial}}) / A_{\text{whole}}$. Error bars represent 99% confidence interval on the estimate of the mean.

(B) Comparison between selectivity for partial versus whole objects measured by the F-statistic. Most of the data were below the diagonal (dashed line). The arrow points to the example from Figure 2. Here we only show data from the Main experiment (F values are hard to compare across experiments because of the different number of categories; hence, the example from Figure 3 is not in this plot).

(C) Comparison between selectivity for partial versus whole objects measured by the single trial decoding performance (Experimental Procedures). Most of the data are below the diagonal (dashed line). Chance performance is 20%.

(D) For all pairs of discriminable trials ($n = 10,438$ pairs from 22 selective electrodes), we computed the distribution of the percent overlap in shared pixels. The percent overlap between two pairs of trials (inset, red and blue bubbles) was defined as the number of shared pixels (black) divided by the total object area (area inside gray outline).

(E) Locations of electrodes that showed visual selectivity in both Whole and Partial conditions. Electrodes were mapped to the same reference brain. Example electrodes from Figures 2 and 3 are marked by arrows. Colors indicate different brain gyri.

(F) Percent of total electrodes in each region that were selective in either the Whole condition (black) or in both conditions (gray). Color in the location name corresponds to the brain map in (E). The number of selective electrodes is shown next to each bar. Only regions with at least one electrode selective in both conditions are shown.

objects ($n = 22$). For the responses to whole objects, the median latency across these electrodes was 155 ms, which is consistent with previous estimates (Allison et al., 1999; Liu et al., 2009). The responses to partial objects showed a significant delay in the selectivity latency as measured using ANOVA (median latency difference between Partial and Whole conditions = 117 ms, Figure 5A, black dots, $p < 10^{-5}$) or Decoding (median difference = 158 ms, Figure 5B, black dots, $p < 10^{-5}$). Similar effects were observed when considering two-class selectivity metrics such as d' (Figures S4A and S4B).

We examined several potential factors that might correlate with the observed latency differences. Stimulus contrast is known to cause significant changes in response magnitude and latency across the visual system (e.g., Reich et al., 2001;

Shapley and Victor, 1978). As noted above, there was no significant difference in the response magnitudes between Whole and Partial conditions (Figure 4A). Furthermore, in the Variant experiment, where all the images had the same contrast, we still observed latency differences between conditions (median difference = 73 ms [ANOVA], Figure 5A; and median difference = 93 ms [Decoding], Figure 5B, gray circles).

Because the spatial distribution of bubbles varied from trial to trial, each image in the Partial condition revealed different visual features. As a consequence, the response waveform changed from trial to trial in the Partial condition (e.g., compare the strikingly small trial-to-trial variability in the responses to whole objects with the considerable variability in the responses to partial objects, Figure 3B). Yet, the latency differences between Whole

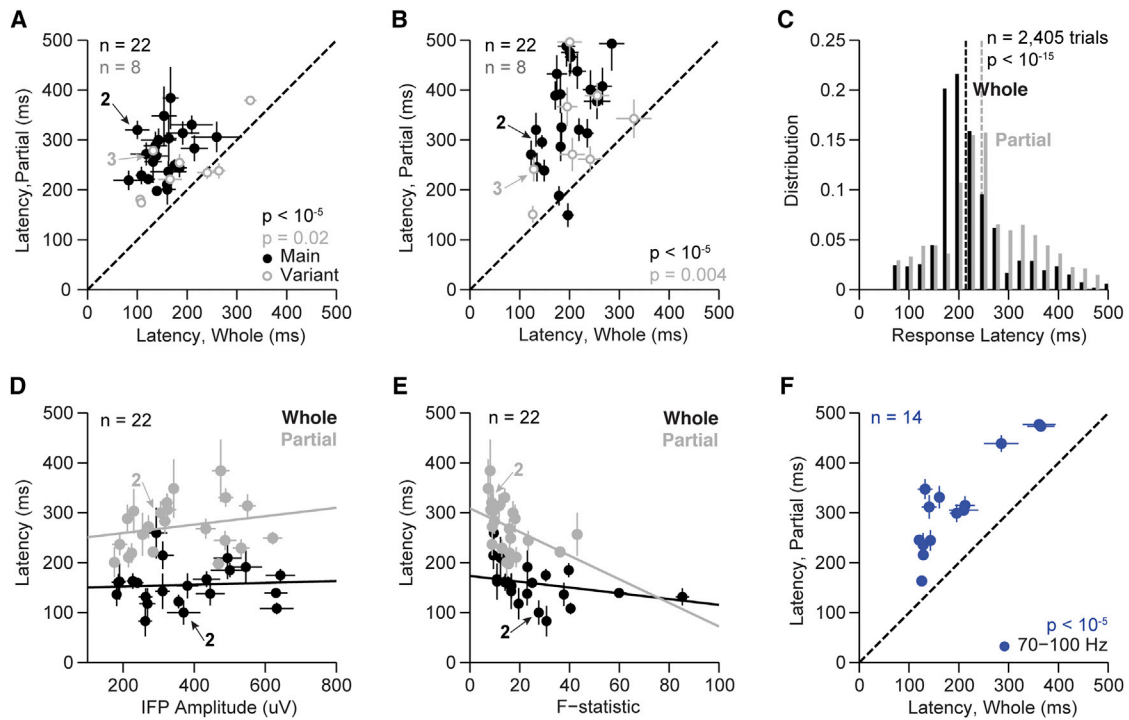


Figure 5. Increased Latency for Object Completion

We considered several definitions of latency (see text).

(A) Latency of selective responses, as measured through ANOVA (e.g., Figure 2D) for electrodes selective in both Whole and Partial conditions from the Main (black, $n = 22$) and Variant (gray, $n = 8$) experiments. The latency distributions were significantly different (Wilcoxon signed-rank test, Main experiment: $p < 10^{-5}$; Variant experiment: $p = 0.02$). Error bars represent 99% confidence interval on the estimate of the mean.

(B) Latency as measured by the machine learning decoding analysis (e.g., Figure 2E). These latency distributions were significantly different (Wilcoxon signed-rank test, Main experiment: $p < 10^{-5}$; Variant experiment: $p = 0.004$).

(C) Distribution of visual response latencies in single trials for whole (black) and partial (gray) objects (as illustrated in Figure 2F). These distributions were significantly different (Wilcoxon rank-sum test, $p < 10^{-15}$). The vertical dashed lines denote the means of each distribution.

(D) There was no significant correlation between selectivity latency (measured using ANOVA) and IFP amplitude (defined in Figure 4A) (Whole: $r = 0.13$, $p = 0.29$; Partial: $r = 0.15$, $p = 0.27$).

(E) The correlation between selectivity latency and the selectivity as evaluated by the F-statistic was significant in the Partial condition ($r = -0.43$, $p = 0.03$), but not in the Whole condition ($r = -0.36$, $p = 0.06$). However, the latency difference between conditions was still significant when accounting for changes in the strength of selectivity (analysis of covariance [ANCOVA], $p < 10^{-8}$), as can be observed by comparing latencies in subpopulations of matched selectivity.

(F) Latency of selective responses from electrodes using power in the 70–100 Hz (gamma, blue) frequency bands. Statistical significance measured with the Wilcoxon signed-rank test ($p < 10^{-5}$).

and Partial conditions were apparent even in single trials (e.g., Figures 2A and 3A). These response latencies depended on the sets of features revealed in each trial. In a subset of trials where we presented repetitions of partial objects with one fixed position of bubbles (the “Partial Fixed” condition), the IFP timing was more consistent across trials (Figure 3C, right bottom), but the latencies were still longer for partial objects than for whole objects.

To further investigate the role of stimulus heterogeneity, we measured the response latency in each trial by determining when the IFP amplitude exceeded a threshold set as three SDs above the baseline activity (Figures 2F and 3F). The average response latencies in the Whole and Partial conditions for the preferred category for the first example electrode were 172 and 264 ms, respectively (Figure 2F, Wilcoxon rank-sum test, $p < 10^{-6}$). The distribution of response latencies in the Whole condition was highly peaked (Figures 2F and 3F), whereas the distribu-

tion of latencies in the Partial condition showed a larger variation, driven by the distinct visual features revealed in each trial. This effect was not observed in all the electrodes; some electrodes showed consistent, albeit delayed, latencies across trials in the Partial condition (Figure S3). Across the population, delays were observed in the visual response latencies (Figure 5C, Wilcoxon rank-sum test, $p < 10^{-15}$), even when the latencies were measured with only the most selective responses (Figure S6).

We asked whether the observed delays could be related to differences in the IFP response strength or the degree of selectivity by conducting an analysis of covariance (ANCOVA). The latency difference between conditions was significant even when accounting for differences in IFP amplitude ($p < 10^{-9}$) or strength of selectivity ($p < 10^{-8}$). Additionally, subpopulations of electrodes with matched amplitude or matched selectivity still showed significant differences in the selectivity latency (Figures 5D and 5E).

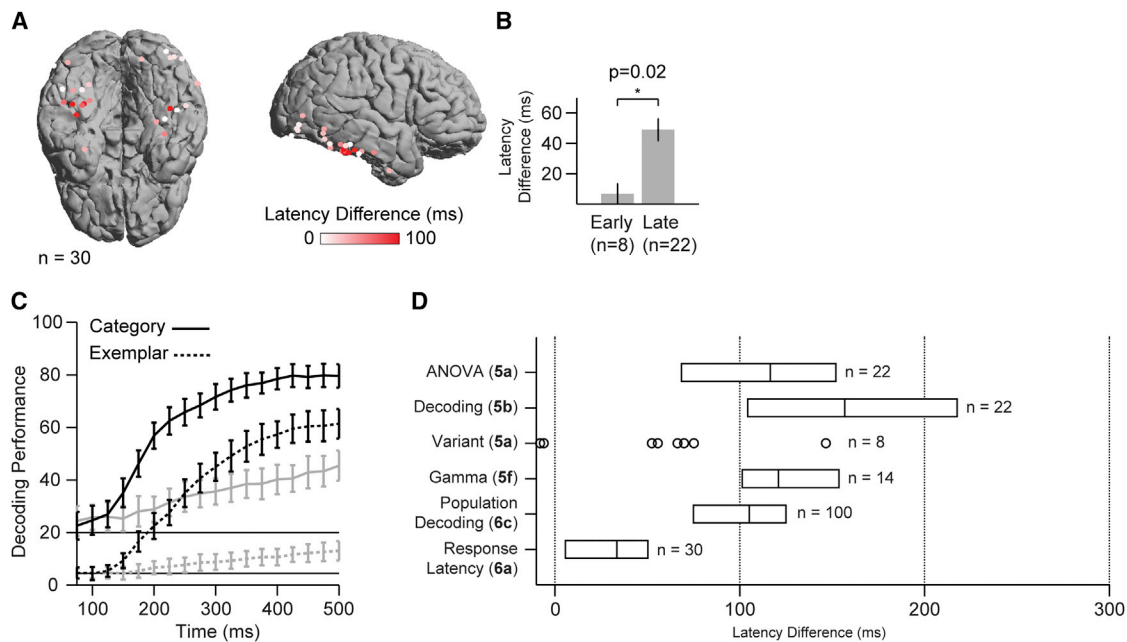


Figure 6. Summary of Latency Measurements

(A) Brain map of electrodes selective in both conditions, colored by the difference in the response latency (Partial–Whole; see color scale on the bottom).
 (B) Comparison of response latency differences (Partial–Whole) between electrodes in occipital lobe (early visual) and temporal lobe (late visual). Error bars represent SEM.
 (C) Decoding performance from pseudopopulation of 60 electrodes for categorization (thick lines) or exemplar identification (dotted lines) for Whole (black) or Partial (gray) conditions (Experimental Procedures). Horizontal lines indicate chance for categorization (20%) and identification (4%). Error bars represent SD. The 60 electrodes used in this analysis were selected using their rank order based on their individual decoding performance on training data.
 (D) Summary of latency difference between Partial and Whole conditions for multiple definitions of latency (parentheses mark the figure source). Positive values mean increased latency in the Partial condition. Box plots represent the median and quartile across the selective electrodes. For the Variant experiment, individual electrodes are plotted since the total number of electrodes n is small. For the population decoding results, the n denotes the number of repetitions using 60 electrodes.

Even though the average amplitudes were similar for whole and partial objects (Figure 4A), the variety of partial images could include a wider distribution with weak stimuli that failed to elicit a response. To further investigate whether such potential weaker responses could contribute to the latency differences, we performed two additional analyses. First, we subsampled the trials containing partial images to match the response amplitude distribution of the whole objects for each category. Second, we identified those trials where the decoder was correct at 500 ms and evaluated the decoding dynamics before 500 ms under these matched performance conditions. The selectivity latency differences between partial and whole objects remained when matching the amplitude distribution or the decoding performance ($p < 10^{-5}$, Figure S4C and S4D; $p < 10^{-7}$, Figures S4E–S4G).

Differences in eye movements between Whole and Partial conditions could potentially contribute to latency delays. We minimized the impact of eye movements by using a small stimulus size (5°), fast presentation (150 ms), and trial order randomization. Furthermore, we recorded eye movements along with the neural responses in two subjects. There were no clear differences in eye movements between whole and partial objects in these two subjects (Figure S5), and those subjects contributed 5 of the 22 selective electrodes in the Main experiment. To

further characterize the eye movements that subjects typically make under these experimental conditions, we also recorded eye movements from 20 healthy volunteers and found no difference in the statistics of saccades and fixation between Whole and Partial conditions (Figure S5; note that these are not the same subjects who participated in the physiological experiments).

Several studies have documented visual selectivity in different frequency bands of the IFP responses including broadband and gamma band signals (Davidesco et al., 2013; Liu et al., 2009; Vidal et al., 2010). We also observed visually selective responses in the 70–100 Hz gamma band (e.g., Figure S2). Delays during the Partial condition documented above for the broadband signals were also observed when measuring the selectivity latency in the 70–100 Hz frequency band (median latency difference = 157 ms, $n = 14$ electrodes, Figure 5F).

To compare delays across different brain regions and different subjects, we mapped each electrode onto the same reference brain. Delays in the response latency between Partial and Whole conditions had a distinct spatial distribution: most of the delays occurred in higher visual areas such as the fusiform gyrus and inferior temporal gyrus (Figure 6A). There was a significant correlation between the latency difference and the electrode position along the anterior-posterior axis of the temporal lobe

(Spearman's correlation = 0.43, permutation test, $p = 0.02$). In addition, the latency difference was smaller for electrodes in early visual areas (occipital cortex) versus late visual areas (temporal lobe), as shown in [Figure 6B](#) ($p = 0.02$, t test). For the two gyri where we had $n > 5$ electrodes selective in both conditions, delays were more prominent in the fusiform gyrus than the inferior occipital gyrus ($p = 0.01$, t test).

The analyses presented thus far only measured selectivity latency for individual electrodes, but the subject has access to activity across many regions. To estimate the selectivity latency from activity across different regions, we combined information from multiple electrodes and across subjects by constructing pseudopopulations ([Hung et al., 2005](#)). For each trial, electrode responses were randomly sampled without replacement from stimulus-matched trials (same exemplar and condition) and then concatenated to produce one response vector for each pseudopopulation trial ([Experimental Procedures](#)). This procedure involves several assumptions including independence and ignores potentially important correlations between electrodes within a trial ([Meyers and Kreiman, 2011](#)). Electrodes were rank-ordered based on their individual decoding performance, and varying population sizes were examined. Decoding performance using electrode ensembles was both fast and accurate ([Figure 6C](#)). Category information emerged within 150 ms for whole objects (black thick line) and 260 ms for partial objects (gray thick line), and reached 80% and 45% correct rate, respectively (chance = 20%). Even for the more difficult problem of identifying the stimulus exemplar (chance = 4%), decoding performance emerged within 135 ms for whole objects (black dotted line) and 273 ms for partial objects (gray dotted line). Exemplar decoding accuracy reached 61% for whole objects and 14% for partial objects. These results suggest that, within the sampling limits of our techniques, electrode ensembles also show delayed selectivity for partial objects.

In sum, we have independently applied several different estimates of latency that use statistical (ANOVA), machine learning (decoding), or threshold (response latency) techniques. These latency measures were estimated using information derived from both broadband signals and specific frequency bands, using individual electrodes as well as electrode ensembles, taking into account changes in contrast, signal strength, and degree of selectivity. Each definition of latency requires different assumptions and emphasizes different aspects of the response, leading to variations in the absolute values of the latency estimates. Yet, independently of the specific definition, the latencies for partial objects were consistently delayed with respect to the latencies to whole objects (the multiple analyses are summarized in [Figure 6D](#), see also [Figure S6](#)).

DISCUSSION

The visual system must maintain visual selectivity while remaining tolerant to a myriad of object transformations. This study shows that neural activity in the human occipitotemporal cortex remained visually selective (e.g., [Figure 2](#)) even when limited partial information about each object was presented (on average, only 18% of each object was visible). Despite the trial-to-trial variation in the features presented, the field potential response

waveform, amplitude, and object preferences were similar between the Whole and Partial conditions ([Figures 2–4](#)). However, the neural responses to partial objects required ~ 100 ms of additional processing time compared to whole objects ([Figures 5 and 6](#)). While the exact value of this delay may depend on stimulus parameters and task conditions, the requirement for additional computation was robust to different definitions of latencies including single trial analyses, different frequency bands, and different statistical comparisons ([Figure 6D](#)) and persisted when accounting for changes in image contrast, signal strength, and the strength of selectivity ([Figure 5](#)). This additional processing time was more pronounced in higher areas of the temporal lobe, including ITC and the fusiform gyrus, than in earlier visual areas ([Figure 6A](#)).

Previous human neuroimaging, scalp electroencephalography, and IFP recordings have characterized object completion by comparing responses to occluded objects with feature-matched scrambled counterparts ([Lerner et al., 2004](#); [Sehatpour et al., 2008](#)) or by comparing responses to object parts and wholes ([Schiltz and Rossion, 2006](#); [Taylor et al., 2007](#)). Taking a different approach, neurophysiological recordings in the macaque ITC have examined how robust shape selectivity or encoding of diagnostic features are to partial occlusion ([Issa and Dicarlo, 2012](#); [Kovacs et al., 1995](#); [Missal et al., 1997](#); [Nielsen et al., 2006](#)). Comparisons across species (monkeys versus humans) or across different techniques (IFP recordings versus fMRI) have to be interpreted with caution. However, the locations where we observed selective responses to partial objects, particularly ITC and fusiform gyrus ([Figures 4E and 4F](#)), are consistent with and provide a link between macaque neurophysiological recordings of selective responses and human neuroimaging of the signatures of object completion.

Presentation of whole objects elicits rapid responses that show initial selectivity within 100–200 ms after stimulus onset ([Hung et al., 2005](#); [Keyzers et al., 2001](#); [Liu et al., 2009](#); [Optican and Richmond, 1987](#); [Thorpe et al., 1996](#)). The speed of the initial selective responses is consistent with a largely bottom-up cascade of processes leading to recognition ([Deco and Rolls, 2004](#); [Fukushima, 1980](#); [Riesenhuber and Poggio, 1999](#); [Rolls, 1991](#)). For partial objects, however, visually selective responses were significantly delayed with respect to whole objects ([Figures 5 and 6](#)). These physiological delays are inconsistent with a purely bottom-up signal cascade, and stand in contrast to other transformations (scale, position, rotation) that do not induce additional neurophysiological delays ([Desimone et al., 1984](#); [Ito et al., 1995](#); [Liu et al., 2009](#); [Logothetis et al., 1995](#); [Logothetis and Sheinberg, 1996](#)).

Delays in response timing have been used as an indicator for recurrent computations and/or top-down modulation ([Buschman and Miller, 2007](#); [Keyzers et al., 2001](#); [Lamme and Roelfsema, 2000](#); [Schmolsky et al., 1998](#)). In line with these arguments, we speculate that the additional processing time implied by the delayed physiological responses can be ascribed to recurrent computations that rely on prior knowledge about the objects to be recognized ([Ahissar and Hochstein, 2004](#)). Horizontal and top-down projections throughout visual cortex could instantiate such recurrent computations ([Callaway, 2004](#); [Felleman and Van Essen, 1991](#)). Several areas where such top-down and

horizontal connections are prevalent showed selective responses to partial objects (Figures 4E and 4F).

It is unlikely that these delays were due to the selective signals to partial objects propagating at a slower speed through the visual hierarchy in a purely feedforward fashion. Selective electrodes in earlier visual areas did not have a significant delay in the response latency, which argues against latency differences being governed purely by low-level phenomena. Delays in the response latency were larger in higher visual areas, suggesting that top-down and/or horizontal signals within those areas of the temporal lobe are important for pattern completion (Figure 6A). Additionally, feedback is known to influence responses in visual areas within 100–200 ms after stimulus onset, as evidenced in studies of attentional modulation that involve top-down projections (Davidesco et al., 2013; Lamme and Roelfsema, 2000; Reynolds and Chelazzi, 2004). Those studies report onset latencies of feedback effects similar to the delays observed here in the same visual areas along the ventral stream. Cognitive effects on scalp electroencephalogram (EEG) responses that presumably involve feedback processing have also been reported at similar latencies (Schyns et al., 2007). The differences reported here between approximately early and high parts of the ventral visual stream are reminiscent of neuroimaging results that compare part and whole responses along the ventral visual stream (Schiltz and Rossion, 2006; Taylor et al., 2007).

The selective responses to partial objects were not exclusively driven by a single object patch (Figures 2A, 2B, 3A, and 3B). Rather, they were tolerant to a broad set of partial feature combinations. While our analysis does not explicitly rule out common features shared by different images with largely nonoverlapping pixels, the large fraction of trials with images with low overlap that elicited robust and selective responses makes this explanation unlikely (Figure 4D). The response latencies to partial objects were dependent on the features revealed: when we fixed the location of the bubbles, the response timing was consistent from trial to trial (Figure 3C).

The distinction between purely bottom-up processing and recurrent computations confirms predictions from computational models of visual recognition and attractor networks. Whereas recognition of whole objects has been successfully modeled by purely bottom-up architectures (Deco and Rolls, 2004; Fukushima, 1980; Riesenhuber and Poggio, 1999), those models struggle to identify objects with only partial information (Johnson and Olshausen, 2005; O'Reilly et al., 2013). Instead, computational models that are successful at pattern completion involve recurrent connections (Hopfield, 1982; Lee and Mumford, 2003; O'Reilly et al., 2013). Different computational models of visual recognition that incorporate recurrent computations include connections within the ventral stream (e.g., from ITC to V4) and/or from prefrontal areas to the ventral stream. Our results implicate higher visual areas (Figures 4E and 6A) as participants in the recurrent processing network involved in recognizing objects from partial information. Additionally, the object-dependent and unimodal distribution of response latencies to partial objects (e.g., Figure 2F) suggest models that involve graded evidence accumulation as opposed to a binary switch.

The current observations highlight the need for dynamical models of recognition to describe where, when, and how recur-

rent processing interacts with feedforward signals. Our findings provide spatial and temporal bounds to constrain these models. Such models should achieve recognition of objects from partial information within 200–300 ms, demonstrate delays in the visual response that are feature dependent, and include a graded involvement of recurrent processing in higher visual areas. We speculate that the proposed recurrent mechanisms may be employed not only in the context of object fragments but also in visual recognition for other types of transformations that impoverish the image or increase task difficulty. The behavioral and physiological observations presented here suggest that the involvement of recurrent computations during object completion, involving horizontal and top-down connections, result in a representation of visual information in the highest echelons of the ventral visual stream that is selective and robust to a broad range of possible transformations.

EXPERIMENTAL PROCEDURES

Physiology Subjects

Subjects were 18 patients (11 male, 17 right handed, 8–40 years old, Table S1) with pharmacologically intractable epilepsy treated at Children's Hospital Boston (CHB), Brigham and Women's Hospital (BWH), or Johns Hopkins Medical Institution (JHMI). They were implanted with intracranial electrodes to localize seizure foci for potential surgical resection. All studies described here were approved by each hospital's institutional review boards and were carried out with the subjects' informed consent. Electrode locations were driven by clinical considerations; the majority of the electrodes were not in the visual cortex.

Recordings

Subjects were implanted with 2-mm-diameter intracranial subdural electrodes (Ad-Tech, Racine, WI, USA) that were arranged into grids or strips with 1 cm separation. Each subject had between 44 and 144 electrodes (94 ± 25 , mean \pm SD), for a total of 1,699 electrodes. The signal from each electrode was amplified and filtered between 0.1 and 100 Hz with sampling rates ranging from 256 Hz to 1,000 Hz at CHB (XLTEK, Oakville, ON, Canada), BWH (Bio-Logic, Knoxville, TN, USA), and JHMI (Natus, San Carlos, CA and Nihon Kohden, Tokyo, Japan). A notch filter was applied at 60 Hz. All the data were collected during periods without any seizure events. In two subjects, eye positions were recorded simultaneously with the physiological recordings (ISCAN, Woburn, MA).

Neurophysiology Experiments

After 500 ms of fixation, subjects were presented with an image (256×256 pixels) of an object for 150 ms, followed by a 650 ms gray screen, and then a choice screen (Figure 1A). The images subtended 5° of visual angle. Subjects performed a 5-alternative forced choice task, categorizing the images into one of five categories (animals, chairs, human faces, fruits, or vehicles) by pressing corresponding buttons on a gamepad (Logitech, Morges, Switzerland). No correct/incorrect feedback was provided. Stimuli consisted of contrast-normalized grayscale images of 25 objects, 5 objects in each of the aforementioned 5 categories. In $\sim 30\%$ of the trials, the images were presented unaltered (the Whole condition). In 70% of the trials, the visual features were presented through Gaussian bubbles of SD 14 pixels (the Partial condition, see example in Figure 1B) (Gosselin and Schyns, 2001). The more bubbles, the more visibility. Each subject was first presented with 40 trials of whole objects, then 80 calibration trials of partial objects, where the number of bubbles was titrated in a staircase procedure to set the task difficulty at $\sim 80\%$ correct rate. The number of bubbles was then kept constant throughout the rest of the experiment. The average percentage of the object shown for each subject is reported in Figure 1C. Unless otherwise noted (below), the positions of the bubbles were randomly chosen in each trial. The trial order was pseudorandomized.

Six subjects performed a variant of the Main experiment with three key differences. First, contrast was normalized between the Whole and Partial conditions by presenting all objects in a phase-scrambled background (Figure 1B). Second, in 25% of the Partial condition trials, the spatial distribution of the bubbles was fixed to a single seed (the Partial Fixed condition). Each of the images in these trials was identical across repetitions. Third, because experimental time was limited, only objects from three categories (animals, human faces, and vehicles) were presented to collect enough trials in each condition.

Data Analyses

Electrode Localization

Electrodes were localized by coregistering the preoperative MRI with the post-operative computed tomography (CT) (Destrieux et al., 2010; Liu et al., 2009). For each subject, the brain surface was reconstructed from the MRI and then assigned to 1 of 75 regions by Freesurfer. Each surface was also coregistered to a common brain for group analysis of electrode locations. The location of electrodes selective in both Whole and Partial conditions is shown in Table S2. In Figure 6A, we computed the Spearman's correlation coefficient between the latency differences (Partial-Whole) and distance along the posterior-anterior axis of the temporal lobe. In Figure 4F, we partitioned the electrodes into three groups: fusiform gyrus, inferior occipital gyrus, and other. We used the Fisher's exact test to assess whether the proportion of electrodes selective in both conditions is greater in the fusiform gyrus versus other, and in inferior occipital gyrus versus other.

Visual Response Selectivity

All analyses in this manuscript used correct trials only. Noise artifacts were removed by omitting trials where the IFP amplitude exceeded five times the SD. The responses from 50 to 500 ms after stimulus onset were used in the analyses.

ANOVA. We performed a nonparametric one-way ANOVA of the IFP responses. For each time bin, the F-statistic (ratio of variance across object categories to variance within object categories) was computed on the IFP response (Keeping, 1995). Electrodes were denoted selective in this test if the F-statistic crossed a threshold (described below) for 25 consecutive milliseconds (e.g., Figure 2D). The latency was defined as the first time of this threshold crossing. The number of trials in the two conditions (Whole and Partial) was equalized by random subsampling; 100 subsamples were used to compute the average F-statistic. A value of 1 in the F-statistic indicates no selectivity (variance across categories comparable to variance within categories), whereas values above 1 indicate increased selectivity.

Decoding. We used a machine learning approach to determine if and when sufficient information became available to decode visual information from the IFP responses in single trials (Bishop, 1995). For each time point t , features were extracted from each electrode by using principal component analysis (PCA) on the IFP response from (50 t) ms, and then keeping those components that explained 95% of the variance. The features set also included the IFP range (max-min), time to maximum IFP, and time to minimum IFP. A multiclass linear discriminant classifier with diagonal covariance matrix was used to either categorize or identify the objects. Ten-fold stratified cross-validation was used to separate the training sets from the test sets. The proportion of trials where the classifier was correct in the test set is denoted the "decoding performance" throughout the text. In the Main experiment, a decoding performance of 20% (1/5) indicates chance for categorization, and 4% (1/25) indicates chance for identification. The number of trials in the Whole and Partial conditions was equalized by subsampling; we computed the average decoding performance across 100 different subsamples. An electrode was denoted selective if the decoding performance crossed a threshold (described below) at any time point t , and the latency was defined as the first time of this threshold crossing.

Pseudopopulation. Decoding performance was also computed from an ensemble of electrodes across subjects by constructing a pseudopopulation, and then performing the same analyses described above (Figure 6C). The pseudopopulation pooled responses across subjects (Hung et al., 2005; Mehring et al., 2003; Pasupathy and Connor, 2002). The features for each trial in this pseudopopulation were generated by first randomly sampling exemplar-matched trials without replacement for each member of the ensemble, and then concatenating the corresponding features. The pseudopopulation

size was set by the minimum data set size of the subject, which in our data was 100 trials (4 from each exemplar). Because of the reduced data set size, 4-fold cross-validation was used.

d-Prime. We compared the above selectivity metrics against d' (Green and Swets, 1966). The value of d' was computed for each electrode by comparing the best category against the worst category, as defined by the average IFP amplitude. d' measures the separation between the two groups normalized by their SD. The latency of selectivity for d' was measured using the same approach as the ANOVA (Figures S4A and S4B).

Significance Thresholds. The significance thresholds for ANOVA, decoding, and d' were determined by randomly shuffling the category labels 10,000 times and using the value of the 99.9th percentile (ANOVA: $F = 5.7$; decoding: 23%, $d' = 0.7$). This represents a false discovery rate of $q = 0.001$ for each individual test. As discussed in the text, we further restricted the set of electrodes by considering the conjunction of the ANOVA and decoding tests. We evaluated this threshold by performing an additional 1,000 shuffles and measuring the number of selective electrodes that passed the same selectivity criteria by chance. In Table 1, we present the number of electrodes that passed each significance test and the number of electrodes that passed the same tests after randomly shuffling the object labels. The conclusions of this study did not change when using a less strict criterion of $q = 0.05$ (median latency difference for ANOVA: 123 ms, $n = 45$ electrodes selective in both conditions, Figure S6).

Latency Measures

We considered several different metrics to quantify the selectivity latency (i.e., the first time point when selective responses could be distinguished) and the visual response latency (i.e., the time point when a visual response occurred). These measures are summarized in Figures 6D and S6.

Selectivity Latency

The selectivity latency represented the first time point when different stimuli could be discriminated and was defined above for the ANOVA, decoding, and d' analyses.

Response Latency

Latency of the visual response was computed at a per-trial level by determining the time, in each trial, when the IFP amplitude exceeded 3 SDs above the baseline activity. Only trials corresponding to the preferred category were used in the analysis. To test the multimodality of the distribution of response latencies, we used Hartigan's dip test. In 27 of the 30 electrodes, the unimodality null hypothesis could not be rejected ($p > 0.05$).

Frequency Band Analyses

Power in the gamma frequency band (70–100 Hz) was evaluated by applying a fifth-order Butterworth filter bandpass and computing the magnitude of the analytical representation of the signal obtained with the Hilbert transform. The same analyses (ANOVA, decoding, per-trial latency) were applied to the responses from all electrodes in different frequency bands.

Bubble Overlap Analyses

For each pair of partial object trials, the percent of overlap was computed by dividing the number of pixels that were revealed in both trials by the area of the object (Figure 4D). Because a low degree of object overlap would be expected in trials with weak physiological responses, we focused on the most robust responses for these analyses by considering those trials when the IFP amplitude was greater than the 90th percentile of the distribution of IFP amplitudes of all the nonpreferred category trials.

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures and two tables and can be found with this article online at <http://dx.doi.org/10.1016/j.neuron.2014.06.017>.

AUTHOR CONTRIBUTIONS

H.T. and C.B. performed the experiments and analyzed the data; H.T., C.B., and G.K. designed the experiments; R.M., N.E.C., W.S.A., and J.R.M. assisted

in performing the experiments; W.S.A. and J.R.M. performed the neurosurgical procedures; H.T. and G.K. wrote the manuscript. All authors commented and approved the manuscript.

ACKNOWLEDGMENTS

We thank all the patients for participating in this study; Laura Groomes for assistance with the electrode localization and psychophysics experiments; Sheryl Manganaro, Jack Connolly, Paul Dionne, and Karen Walters for technical assistance; and Ken Nakayama and Dean Wyatte for comments on the manuscript. This work was supported by NIH (DP2OD006461) and NSF (0954570 and CCF-1231216).

Accepted: June 2, 2014

Published: July 17, 2014

REFERENCES

- Ahissar, M., and Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends Cogn. Sci.* *8*, 457–464.
- Allison, T., Puce, A., Spencer, D.D., and McCarthy, G. (1999). Electrophysiological studies of human face perception. I: potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cereb. Cortex* *9*, 415–430.
- Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychol. Rev.* *94*, 115–147.
- Biederman, I., and Cooper, E.E. (1991). Priming contour-deleted images: evidence for intermediate representations in visual object recognition. *Cognit. Psychol.* *23*, 393–419.
- Bishop, C.M. (1995). *Neural Networks for Pattern Recognition*. (Oxford: Clarendon Press).
- Brown, J.M., and Koch, C. (2000). Influences of occlusion, color, and luminance on the perception of fragmented pictures. *Percept. Mot. Skills* *90*, 1033–1044.
- Buschman, T.J., and Miller, E.K. (2007). Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science* *315*, 1860–1862.
- Callaway, E.M. (2004). Feedforward, feedback and inhibitory connections in primate visual cortex. *Neural Netw.* *17*, 625–632.
- Chen, J., Zhou, T., Yang, H., and Fang, F. (2010). Cortical dynamics underlying face completion in human visual system. *J. Neurosci.* *30*, 16692–16698.
- Connor, C.E., Brincat, S.L., and Pasupathy, A. (2007). Transformation of shape information in the ventral pathway. *Curr. Opin. Neurobiol.* *17*, 140–147.
- Davidesco, I., Harel, M., Ramot, M., Kramer, U., Kipervasser, S., Andelman, F., Neufeld, M.Y., Goelman, G., Fried, I., and Malach, R. (2013). Spatial and object-based attention modulates broadband high-frequency responses across the human visual cortical hierarchy. *J. Neurosci.* *33*, 1228–1240.
- Deco, G., and Rolls, E.T. (2004). A neurodynamical cortical model of visual attention and invariant object recognition. *Vision Res.* *44*, 621–642.
- Desimone, R., Albright, T.D., Gross, C.G., and Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *J. Neurosci.* *4*, 2051–2062.
- Destrieux, C., Fischl, B., Dale, A., and Hagren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage* *53*, 1–15.
- Felleman, D.J., and Van Essen, D.C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* *1*, 1–47.
- 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.
- Gosselin, F., and Schyns, P.G. (2001). Bubbles: a technique to reveal the use of information in recognition tasks. *Vision Res.* *41*, 2261–2271.
- 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., Grady, C.L., Horwitz, B., Ungerleider, L.G., Mishkin, M., Carson, R.E., Herscovitch, P., Schapiro, M.B., and Rapoport, S.I. (1991). Dissociation of object and spatial visual processing pathways in human extrastriate cortex. *Proc. Natl. Acad. Sci. USA* *88*, 1621–1625.
- Hopfield, J.J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA* *79*, 2554–2558.
- Hung, C.P., Kreiman, G., Poggio, T., and DiCarlo, J.J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science* *310*, 863–866.
- Issa, E.B., and Dicarlo, J.J. (2012). Precedence of the eye region in neural processing of faces. *J. Neurosci.* *32*, 16666–16682.
- 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 recognition of partially visible natural objects in the presence and absence of their occluders. *Vision Res.* *45*, 3262–3276.
- 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.
- Keeping, E.S. (1995). *Introduction to Statistical Inference*. (New York: Dover).
- Keysers, C., Xiao, D.K., Földiák, P., and Perrett, D.I. (2001). The speed of sight. *J. Cogn. Neurosci.* *13*, 90–101.
- Kirchner, H., and Thorpe, S.J. (2006). Ultra-rapid object detection with saccadic eye movements: visual processing speed revisited. *Vision Res.* *46*, 1762–1776.
- Kovacs, G., Vogels, R., and Orban, G.A. (1995). Selectivity of macaque inferior temporal neurons for partially occluded shapes. *J. Neurosci.* *15*, 1984–1997.
- Lamme, V.A., and Roelfsema, P.R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci.* *23*, 571–579.
- Lee, T.S., and Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* *20*, 1434–1448.
- Lerner, Y., Harel, M., and Malach, R. (2004). Rapid completion effects in human high-order visual areas. *Neuroimage* *21*, 516–526.
- Liu, H., Agam, Y., Madsen, J.R., and Kreiman, G. (2009). Timing, timing, timing: fast decoding of object information from intracranial field potentials in human visual cortex. *Neuron* *62*, 281–290.
- 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.
- Mehring, C., Rickert, J., Vaadia, E., Cardoso de Oliveira, S., Aertsen, A., and Rotter, S. (2003). Inference of hand movements from local field potentials in monkey motor cortex. *Nat. Neurosci.* *6*, 1253–1254.
- Meyers, E.M., and Kreiman, G. (2011). Tutorial on pattern classification in cell recordings. In *Understanding Visual Population Codes*, N. Kriegeskorte and G. Kreiman, eds. (Boston: MIT Press).
- Missal, M., Vogels, R., and Orban, G.A. (1997). Responses of macaque inferior temporal neurons to overlapping shapes. *Cereb. Cortex* *7*, 758–767.
- Miyashita, Y., and Chang, H.S. (1988). Neuronal correlate of pictorial short-term memory in the primate temporal cortex. *Nature* *331*, 68–70.
- Murray, R.F., Sekuler, A.B., and Bennett, P.J. (2001). Time course of amodal completion revealed by a shape discrimination task. *Psychon. Bull. Rev.* *8*, 713–720.
- Nakayama, K., He, Z., and Shimojo, S. (1995). Visual surface representation: a critical link between lower-level and higher-level vision. In *Visual Cognition*, S. Kosslyn and D. Osherson, eds. (Cambridge: MIT press).

- Nielsen, K.J., Logothetis, N.K., and Rainer, G. (2006). Dissociation between local field potentials and spiking activity in macaque inferior temporal cortex reveals diagnosticity-based encoding of complex objects. *J. Neurosci.* *26*, 9639–9645.
- Optican, L.M., and Richmond, B.J. (1987). Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. III. Information theoretic analysis. *J. Neurophysiol.* *57*, 162–178.
- O'Reilly, R.C., Wyatte, D., Herd, S., Mingus, B., and Jilk, D.J. (2013). Recurrent processing during object recognition. *Front. Psychol.* *4*, 124.
- Pasupathy, A., and Connor, C.E. (2002). Population coding of shape in area V4. *Nat. Neurosci.* *5*, 1332–1338.
- 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.* *6*, 6234–6242.
- Reich, D.S., Mechler, F., and Victor, J.D. (2001). Temporal coding of contrast in primary visual cortex: when, what, and why. *J. Neurophysiol.* *85*, 1039–1050.
- Reynolds, J.H., and Chelazzi, L. (2004). Attentional modulation of visual processing. *Annu. Rev. Neurosci.* *27*, 611–647.
- Richmond, B.J., Wurtz, R.H., and Sato, T. (1983). Visual responses of 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.
- Rutishauser, U., Tudusciuc, O., Neumann, D., Mamelak, A.N., Heller, A.C., Ross, I.B., Philpott, L., Sutherling, W.W., and Adolphs, R. (2011). Single-unit responses selective for whole faces in the human amygdala. *Curr. Biol.* *21*, 1654–1660.
- Schiltz, C., and Rossion, B. (2006). Faces are represented holistically in the human occipito-temporal cortex. *Neuroimage* *32*, 1385–1394.
- 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.
- Schyns, P.G., Petro, L.S., and Smith, M.L. (2007). Dynamics of visual information integration in the brain for categorizing facial expressions. *Curr. Biol.* *17*, 1580–1585.
- Sehatpour, P., Molholm, S., Schwartz, T.H., Mahoney, J.R., Mehta, A.D., Javitt, D.C., Stanton, P.K., and Foxe, J.J. (2008). A human intracranial study of long-range oscillatory coherence across a frontal-occipital-hippocampal brain network during visual object processing. *Proc. Natl. Acad. Sci. USA* *105*, 4399–4404.
- Shapley, R.M., and Victor, J.D. (1978). The effect of contrast on the transfer properties of cat retinal ganglion cells. *J. Physiol.* *285*, 275–298.
- Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annu. Rev. Neurosci.* *19*, 109–139.
- Taylor, J.C., Wiggett, A.J., and Downing, P.E. (2007). Functional MRI analysis of body and body part representations in the extrastriate and fusiform body areas. *J. Neurophysiol.* *98*, 1626–1633.
- Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. *Nature* *381*, 520–522.
- Vidal, J.R., Ossandón, T., Jerbi, K., Dalal, S.S., Minotti, L., Ryvlin, P., Kahane, P., and Lachaux, J.P. (2010). Category-specific visual responses: an intracranial study comparing gamma, beta, alpha, and ERP response selectivity. *Front. Hum. Neurosci.* *4*, 195.