Published in final edited form as:
Curr Biol. 2010 May 11; 20(9): 872–879. doi:10.1016/j.cub.2010.03.050.

Robust selectivity to two-object images in human visual cortex

Yigal Agam[1], **Hesheng Liu**[4], **Alexander Papanastassiou**[6], **Calin Buia**[1], **Alexandra J. Golby**[5], **Joseph R. Madsen**[6], and **Gabriel Kreiman**[1,2,3,*]

[1] Department of Ophthalmology and Kirby Neurobiology Center, Children's Hospital, Harvard Medical School

[2] Center for Brain Science, Harvard University

[3] Swartz Center for Theoretical Neuroscience, Harvard University

[4] Massachusetts General Hospital

[5] Department of Neurosurgery, Brigham and Women's Hospital

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

SUMMARY

We can recognize objects in a fraction of a second in spite of the presence of other objects [1–3]. The responses in macaque areas V4 and inferior temporal cortex [4–15] to a neuron's preferred stimuli are typically suppressed by the addition of a second object within the receptive field (see however [16,17]). How can this suppression be reconciled with rapid visual recognition in complex scenes? One option is that certain "special categories" are unaffected by other objects [18] but this leaves the problem unsolved for other categories. Another possibility is that serial attentional shifts help ameliorate the problem of distractor objects [19–21]. Yet, psychophysical studies [1–3], scalp recordings [1] and neurophysiological recordings [14,16,22–24], suggest that the initial sweep of visual processing contains a significant amount of information. We recorded intracranial field potentials in human visual cortex during presentation of flashes of two-object images. Visual selectivity from temporal cortex during the initial ~200 ms was largely robust to the presence of other objects. We could train linear decoders on the responses to isolated objects and decode information in two-object images. These observations are compatible with parallel, hierarchical and feed-forward theories of rapid visual recognition [25] and may provide a neural substrate to begin to unravel rapid recognition in natural scenes.

RESULTS

We recorded intracranial field potentials (IFPs) from 672 electrodes (296 in different parts of visual cortex) in 9 subjects implanted with subdural intracranial electrodes. Subjects were presented with contrast-normalized grayscale images (100 ms duration) containing one or two objects. The single-electrode analyses focus on 24 visually selective electrodes.

*To whom correspondence should be addressed: gabriel.kreiman@tch.harvard.edu.

Spatial summation

Figure 1 illustrates the IFP signals from a visually selective electrode in the left fusiform gyrus. Consistently with previous studies (e.g. [24,26]), this electrode showed an enhanced response to human faces compared to other categories (Figure 1A). The activity elicited by two objects from the same category was almost indistinguishable from the activity in one-object images (Figure 1A). There was only a small attenuation when the preferred category was paired with a non-preferred category (Figure 1B). This robustness was largely independent of the non-preferred category (Figure S1A). We defined the IFP "response magnitude" as the signal range, $R=max(IFP)-min(IFP)$, in the [50;300] ms interval after stimulus onset (Figure 1F). Because lack of visual selectivity could be confused with robustness [12], the single-electrode analyses were restricted to 24 electrodes that showed selectivity in one-object images (Experimental Procedures; see Figure S1B–E for more examples and Figure 2A for a normalized average plot). Some electrodes (e.g. Figure 1, S1C–D) showed two peaks in the responses to one-object or two-object images (also Figure 2A).

Object positions were randomized. If robustness to the second object were due to a small IFP receptive field surrounding only one position, we would expect to observe a bimodal response distribution. We did not observe any evidence for such bimodal distributions (Figure 1F, S1B4–E4). Furthermore, the IFPs to the preferred category were consistently stronger across different positions (Figure 1C–D). The position tolerance index (defined in Figure 2B) was 0.09±0.08 for single objects and 0.08±0.07 for two-object images (mean±SD), indicating only a modest response drop across positions. There was no clear preference for the top or bottom positions (Figure 2C). There was a weak correlation between the position tolerance index and the suppression index defined below (Pearson correlation coefficient (ρ)=0.26; $p>0.05$). These observations suggest that robustness to two-object images cannot be ascribed to a small IFP receptive field surrounding only the preferred category. However, the IFP response in both positions still allows for pooling over neurons with smaller receptive fields.

Single neuron responses in macaque area V4 [4,5] or inferior temporal cortex [6–15] to two-object images are significantly attenuated in the presence of non-preferred objects within the receptive field (see however [16,17]). To quantify the degree of suppression, let R_1 (R_2) indicate the response to category 1 (category 2) alone and R_{12} indicate the response to both categories. There was a correlation between R_{12} and $max(R_1,R_2)$ even though R_{12} was consistently below $max(R_1,R_2)$ (Figure 1E,S2A). R_{12} was also correlated with $max(R_1,R_2)$ at the population level (Figure 2D,S2B). When considering individual exemplars, the mean suppression index (SI, defined in Figure 2E) was −0.09±0.16 (24 electrodes, n=80 exemplar pairs). When considering categories, the mean SI was −0.02±0.09 (24 electrodes, n=140 category pairs). We did not observe differences in SI between those electrodes that preferred human faces versus other categories (Figure 2D–E;S2B–C). The SI values were also similar in the [50;200] ms interval and in the target-present trials (Figure S2C). We considered several typical models for estimating the response to object pairs from the responses to the individual objects: maximum, average, unscaled power, scaled linear, normalization, scaled power and generalized linear (Figure S2D–I). The best fits were obtained using a two-parameter model: $\alpha max(R_1, R_2) + \beta min(R_1, R_2)$ (Figure S2F–H; see also [4,27]). There was a stronger contribution from the first term (max) compared to the second term (min) ($<\alpha>=0.74\pm0.18$, $<\beta>=0.13\pm0.28$, mean±SD, Figure S2G).

Single-trial decoding in two-object images

We asked whether we could decode visual information in single presentations of two-object images from the activity of individual electrodes or electrode ensembles using a machine-learning approach [23,24,28]. Figure S3A–D illustrates single-trial responses from the electrode in Figure 1. We extracted three parameters from the single-trial responses ([50;300]

ms): the minimum voltage time, the maximum voltage time and the response magnitude (Figure S3). We used the responses to one-object images to train a binary support vector machine classifier (SVM) with a linear kernel to indicate the presence or absence of the preferred category. The classification performance (CP) was evaluated using the responses to two-object images (CP=50% indicates chance levels whereas CP=100% indicates perfect performance; see also Figure S5F). In Figure S3, the classification performance was 71±1% (mean±SD). To assess the statistical significance of the CP values, we computed the distribution of CP values in 100 iterations where we randomly shuffled the object labels. The CP values ranged from 51% to 77% (60±7%; mean±SD, n=24 electrodes; Figure 3A). Of the 24 visually selective electrodes (based on single-object images), 21 electrodes (88%) showed a significant CP in the two-object condition (training on single objects and testing on two-object images; $_{single}CP_{2\text{-}object}$).

We previously examined single-trial responses to one-object images [23,24] by training a classifier with a fraction of the repetitions (70%) and evaluating CP with the remaining repetitions ("$CP_{selectivity}$"). $CP_{selectivity}$ was correlated with $_{single}CP_{2\text{-}object}$ (Figure S4B). The points in Figure S4B were below the diagonal, indicating a drop in CP when extrapolating from one-object images to two-object images. Of the 18 locations with >10 electrodes (Table S1), four locations yielded significant CP in two-object images: the inferior occipital cortex, the lateral fusiform gyrus, the parahippocampal gyrus and the inferior temporal cortex [29] (Figure S4C–D). Yet, we note that our sampling is far from exhaustive.

We extended the machine learning approach by considering a "pseudopopulation" defined by combining electrodes across the entire data set (e.g. [23,24]). We concatenated the responses from multiple electrodes (we did not consider interactions among electrodes). Electrode selection for the pseudopopulation was based on selectivity to one-object images (Figure 3B) or electrode location (Figure 3D–G). The performance of a pseudopopulation consisting of 45 electrodes is shown in Figure 3B. The ensemble of electrodes yielded a stronger extrapolation to two-object images than the individual electrodes (cf. Figure 3A vs. 3B). The main locations that yielded significant classification performance were the inferior occipital gyrus, the lateral fusiform gyrus and inferior temporal cortex (Figure 3D–G).

We also examined the performance of the classifier when it was trained using the responses to two-object images. In Figure S5B–F we compare different ways of training and testing the classifier's performance. Overall, the CP values for these different variations were similar (Figure 3C and S5). These results suggest that an algorithm that learns to recognize object categories from the neural signals in human temporal cortex can be trained in the presence or in the absence of another object in the image.

Temporal dynamics and latencies

Physiological signals in the human temporal cortex show a latency of ~100–150 ms (e.g. [24]; see also similar latencies in macaques [22,23,30] and human scalp signals [1,31]). The selectivity in the IFPs in two-object images was apparent from the beginning of the evoked IFP signal (Figure 1B,S1B–E). We computed the latency of the responses to two-object images (Figure 4A). If the selective responses were due to attentional shifts or fast saccades to one of the objects, we might expect that the responses to two-object images would show a longer latency compared to one-object images [32–34]. In contrast, we did not observe significant differences between the response latencies to one-object images (167±45 ms, mean±SD, n=24) and two-object images (157±37 ms, mean±SD, n=24) (Figure 4A; two-tailed t-test p>0.3). Furthermore, there was a weak but significant correlation between the latencies to one-object and two-object images (ρ=0.67; Figure 4B). To further examine the temporal evolution of the IFP responses, we computed the suppression index (SI) as a function of time in 25 ms bins. Overall, SI remained close to 0 in the [50;300] ms interval (e.g. compare thin and thick traces

in Figure S1B3–E3) and there were no consistent monotonic changes in SI over time (Figure 4C).

DISCUSSION

We can rapidly recognize objects within 100–200 ms of seeing a complex scene [1,2,25]. Object recognition in multi-object images poses a challenging problem due to difficulties in segmentation, increased processing time and response attenuation [4–15]. Given these challenges, what are the neural mechanisms that underlie rapid recognition in multi-object images? Attention may help filter out "irrelevant" information enhancing certain locations/features/objects. While attention plays an important role in crowded images [19–21], it remains difficult to explain the high performance during brief presentation of a novel image by serial attentional shifts [1–3,14,25]. Alternatively, the first sweep of information through the ventral stream may contain sufficient information to account for recognition in multi-object images. We evaluated this possibility by quantifying how well we can decode information from IFP recordings in human visual cortex in response to two-object images. We report that the representation in inferior occipital gyrus, fusiform gyrus and inferior temporal cortex can support object recognition even in the presence of a second object in the image. The rapid onset of the selective responses suggests that recognition in two-object images may not require additional computational steps.

The degree of response suppression reported here is lower than in previous studies [4–15] (see however [16,17,35]). Several non-exclusive factors may account for these differences including the species (macaques versus humans), brain areas (it remains difficult to establish one-to-one homologies between macaques and humans), recording techniques (field potentials versus action potentials), tasks and stimulus characteristics (particularly distance between objects and whether the two objects appear in the same hemifield [10,13,16]). The biophysical nature underlying the IFPs remains only poorly understood. IFPs may reflect synaptic potentials averaged over many neurons [36]. We *speculate* that the IFPs may provide a "population view" that shows enhanced robustness to two-object images compared to individual neurons.

A possible mechanism to account for robustness to two-object images would be rapid attentional shifts and/or saccades to the electrode's preferred category. While this possibility cannot be entirely ruled out here, it seems to be an unlikely account of our observations. (i) Subjects could not predict where to saccade before image onset (positions were randomized). Additionally, the response distributions were unimodal (Figure S1) and behavioral performance was indistinguishable across positions. These observations render it unlikely that the results could be accounted by pre-onset fixation or spatial attention to one location. (ii) Adding saccade times of 200–300 ms [37, 38] and latencies of 100–150 ms [24] to the 100 ms stimulus flash, physiological responses elicited by saccades would take place after ~300 ms. (iii) In one subject where we monitored eye position, we did not observe any differences in the responses or suppression index (SI) that could be explained by eye movements. (iv) We observed similar SI when the analysis interval was restricted to [50;200] ms (Figure S2C). (v) Similar SIs were observed for electrodes that preferred faces or other categories (Figure 2D–E, S2B–C). Furthermore, in some cases, there were different electrodes in the same subject that preferred different categories; a category-specific attentional account would necessarily fail to explain the responses in some electrodes (e.g. Figure S1B–C). (vi) The SIs during the initial 300 ms were unaffected by target presence (Figure S2C). (vii) The latencies to one-object images and two-object images were similar (Figure 4). Taken together, observations (i)–(vii) do not rule out an attentional account of our findings but delimit the possible roles of attention. The physiological characterization of the spatial summation properties (Figures 2, 3, S2), category preferences (Figures 3, S4, S5), task demands (Figure S2C) and timing (Figure

4) places strong constraints on how attentional shifts should be incorporated into biophysically-plausible computational circuits for visual recognition (e.g. [25]).

We used two relatively large objects surrounded by a gray background. Visual recognition becomes more challenging and reveals serial attentional shifts upon increasing the amount of "clutter" in the image. Therefore, we do not claim that the initial physiological signals in temporal cortex can account for visual recognition under *all* possible visual conditions. Our work does suggest, however, that the presence of two objects and modest response suppression do not completely disrupt visual recognition by the initial sweep of visually selective signals.

EXPERIMENTAL PROCEDURES

Subjects

Subjects were nine patients (10–47 years old, 6 right-handed, 3 males) with epilepsy admitted into either Children's Hospital Boston (CHB) or Brigham and Women's Hospital (BWH) to localize the seizure foci for potential surgical resection [39,40]. The tests were approved by the IRBs at both Hospitals and were performed under the subjects' written consent.

Recordings

Subjects were implanted with 64 to 88 intracranial subdural grid (64%) or strip (36%) electrodes (8 subjects) or intracortical depth electrodes (1 subject) as part of the surgical approach to treat epilepsy. The grid/strip electrodes were 2 mm in diameter, with 1 cm separation and impedances below 1kΩ (Ad-Tech, Racine, WI). The signal from each electrode was amplified (×2500), filtered between 0.1 and 100 Hz and sampled at 256 Hz at CHB (XLTEK, Oakville, ON) and 500 Hz at BWH (Bio-Logic, Knoxville, TN). A notch filter was applied at 60Hz to remove line noise artifacts (5-th order bandstop Butterworth filter between 58 and 62 Hz implemented in MATLAB's *butter* function). We refer to the voltage signal as "intracranial field potential" (IFP). Subjects stayed in the hospital 6 to 9 days. The number and location of the electrodes were determined by clinical criteria (Table S1). In one subject, we monitored eye movements using a non-invasive system from ISCAN (DTL-300, Woburn, MA) which provides a spatial resolution of ~1 deg and a temporal scanning frequency of 60 Hz. We excluded from the analyses those electrodes that were considered to be part of the epileptogenic focus according to clinical criteria.

Stimulus presentation and task

Subjects were presented with grayscale images containing one or two objects. Objects belonged to one of five possible categories: animals, chairs, human faces, cars and houses. There were 5 exemplar objects per category and each exemplar object was contrast normalized. Images were presented for 100 ms, with a 1000 ms gray screen in between images. Images included one object (30%) or two objects (70%). The two objects were presented either above and below the fixation point (50%) or to the left and right of the fixation point (50%). In the one-object images, the object was randomly presented at one of the possible locations (above/below or left/right with respect to the fixation point) at the same size and eccentricity as in the two-object images. In the first three subjects, there were four possible positions (above, below, right, left of the fixation point). In the remaining six subjects, the task was restricted to two positions (above/below) to increase the number of repetitions at each position. Objects subtended ~3.4 degrees of visual angle and were presented with their center ~3.8 degrees from the fixation point. Subjects were asked to fixate on the fixation point. Object order and positions were randomized. The duration of each session (and therefore the number of repetitions) depended on clinical constraints and subject fatigue (min duration = 6 mins, max duration = 29 mins, mean=14.8±8.0 mins). In many cases we ran several sessions per subject (min=1 session, max=4 sessions, mean=2.9±1.1 sessions). The first two presentations within each block were

not considered for analyses to avoid potential non-stationary effects. Data from all sessions for a given subject were pooled together for analyses. On average, the total number of presentations was 1156±451; 338±131 one-object images (67±13 per category) and 650±242 two-category images (64±66 per category pair). At the onset of each block (50 images per block) a target category was announced by a written word presented on the screen. Subjects had to indicate by pressing designated "yes" and "no" keys whether or not each image included an object from the target category. The overall performance was 92±12% correct (range=75–100 % correct; one-object images 92±10%; two-object images: 91±13%). The average reaction time was 630 ±90 ms (one-object images: 625±103 ms; two-object images: 640±96 ms).

Data Analyses

Electrode localization—To localize the electrodes, we integrated anatomical information from preoperatively acquired MRI and spatial information of electrode positions provided by postoperatively acquired CT. For each subject, the 3-D brain surface was reconstructed and an automatic parcellation was performed using Freesurfer [24]. CT images were first registered to the MRI using a 3-D affine transform based on multiple fiducial marks. After the co-registration, electrodes were projected onto the nearest brain surface (Table S1). Electrodes were superimposed on the reconstructed brain surface for visualization purposes in the figures. Talairach coordinates and brain renderings for all 672 electrodes are available upon request.

IFP response definition—We focused on the initial part of the IFP response (50 to 300 ms after stimulus onset) because (i) it is more strongly correlated with the visual stimuli; (ii) it is less affected by potential effects of eye movements or attentional shifts [37]; (iii) we can directly compare the responses against the same intervals used in macaque studies (e.g. [15, 23]) and (iv) we can more readily compare the initial sweep of the response with feed-forward models of object recognition [25]. We have previously characterized IFP signals based on different response definitions [24]. We define the "*IFP response magnitude*", *R*, as the voltage range (*max(V)−min(V)*) in the [50;300] ms time interval. An electrode was defined to show *visual selectivity* if a one-way ANOVA across object categories based on the IFP response magnitude to the one-object images yielded $p<0.01$ [24]. Visually selective electrodes responded to an average of 1.45 categories (ranging from 1 to 3 categories). Unless stated otherwise (Figure S2C), the analyses focus on those trials where the target category was absent to remove the possible influence of the target on the spatial summation properties. The initial IFP response magnitude was not significantly affected by the presence or absence of the target category (Figure S2C). Many electrodes did show significant differences between target and non-target trials beyond 300 ms after stimulus onset. However, the physiological responses beyond 300 ms are beyond the scope of the manuscript.

Spatial summation properties—We compared the responses to one-object images against two-object images (e.g. Figure 2 and S2). We also considered several biophysically-plausible simple models [4] to explain the response to two-object images based on the responses to the constituent single objects (Figure S2). When fitting these models, we used the function *nlinfit* in MATLAB.

Classifier analysis—We used a machine-learning approach [23,28] to read out visual information from the IFP responses in single trials. We considered the [50;300] ms interval and we defined three features of the IFP signal: the minimum voltage time (t_{min}), the maximum voltage time (t_{max}) and the response magnitude *R* (Figure S3A–D). For each electrode *i*, we constructed a response vector: [$t^i_{\min}, t^i_{\max}, R^i$]. Several other ways of defining the response vector for each electrode were described previously [24]. This response vector is defined for each individual trial (there is no averaging of responses across trials). The classifier approach allows us to consider each electrode independently or to examine the encoding of information

by an ensemble of multiple electrodes. When considering a set of N electrodes, we assumed independence across electrodes and concatenated the responses to build the ensemble response vector: [$t^1_{\min}, t^1_{\max}, R^1, \ldots, t^N_{\min}, t^N_{\max}, R^N$]. The results shown throughout the manuscript correspond to binary classification between a given category and the other categories (see Figure S5F for multiclass classification). In a binary classifier, chance corresponds to 50% (horizontal dashed line in the plots). We used a support vector machine (SVM) classifier with a linear kernel to learn the map between the ensemble response vectors and the object categories. In all cases, the data were divided into two non-overlapping sets, a training set and a test set. We examined different ways of separating the data into a training set and a test set (Figure S5). Throughout the text, we report the proportion of test repetitions correctly labeled as "Classification performance" (CP). To assess the statistical significance of the classification performance values, we compared the results against those obtained after performing 100 iterations where we randomly shuffled the object labels. We considered CP to be significant if performance was more than 3 standard deviations above the null hypothesis.

Latency—We used two different definitions to compute the response latency as described in Figure 4.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank the patients for their cooperation. We also thank Nuo Li, David Cox, Geoffrey Ghose and Rufin Vogels for comments on the manuscript and Sheryl Manganaro and Paul Dionne for technical assistance. We acknowledge financial support from the Epilepsy Foundation, the Klingenstein Fund, the Whitehall Foundation, NIH grant 1R21EY019710 and an NIH New Innovator Award (1DP2OD006461).

References

1. Thorpe S, Fize D, Marlot C. Speed of processing in the human visual system. Nature 1996;381:520–522. [PubMed: 8632824]
2. Potter M, Levy E. Recognition memory for a rapid sequence of pictures. Journal of Experimental Psychology 1969;81:10–15. [PubMed: 5812164]
3. Li FF, VanRullen R, Koch C, Perona P. Rapid natural scene categorization in the near absence of attention. Proc Natl Acad Sci U S A 2002;99:9596–9601. [PubMed: 12077298]
4. Ghose GM, Maunsell JH. Spatial summation can explain the attentional modulation of neuronal responses to multiple stimuli in area V4. J Neurosci 2008;28:5115–5126. [PubMed: 18463265]
5. Connor CE, Preddie DC, Gallant JL, Van Essen DC. Spatial attention effects in macaque area V4. J Neurosci 1997;17:3201–3214. [PubMed: 9096154]
6. Sheinberg DL, Logothetis NK. Noticing familiar objects in real world scenes: the role of temporal cortical neurons in natural vision. J Neurosci 2001;21:1340–1350. [PubMed: 11160405]
7. Rolls ET, Tovee MJ. The responses of single neurons in the temporal visual cortical areas of the macaque when more than one stimulus is present in the receptive field. Exp Brain Res 1995;103:409–420. [PubMed: 7789447]
8. Miller EK, Gochin PM, Gross CG. Suppression of visual responses of neurons in inferior temporal cortex of the awake macaque by addition of a second stimulus. Brain Res 1993;616:25–29. [PubMed: 8358617]
9. Chelazzi L, Duncan J, Miller EK, Desimone R. Responses of neurons in inferior temporal cortex during memory-guided visual search. Journal of Neurophysiology 1998;80:2918–2940. [PubMed: 9862896]
10. Missal M, Vogels R, Li CY, Orban GA. Shape interactions in macaque inferior temporal neurons. J Neurophysiol 1999;82:131–142. [PubMed: 10400942]

11. Zoccolan D, Cox DD, DiCarlo JJ. Multiple object response normalization in monkey inferotemporal cortex. J Neurosci 2005;25:8150–8164. [PubMed: 16148223]
12. Zoccolan D, Kouh M, Poggio T, DiCarlo JJ. Trade-off between object selectivity and tolerance in monkey inferotemporal cortex. J Neurosci 2007;27:12292–12307. [PubMed: 17989294]
13. Sato T. Interactions of visual stimuli in the receptive fields of inferior, temporal neurons in awake macaques. Exp Brain Res 1989;77:23–30. [PubMed: 2792266]
14. Li N, Cox DD, Zoccolan D, DiCarlo JJ. What response properties do individual neurons need to underlie position and clutter "invariant" object recognition? J Neurophysiol 2009;102:360–376. [PubMed: 19439676]
15. De Baene W, Premereur E, Vogels R. Properties of shape tuning of macaque inferior temporal neurons examined using rapid serial visual presentation. J Neurophysiol 2007;97:2900–2916. [PubMed: 17251368]
16. Gawne TJ, Martin JM. Responses of primate visual cortical V4 neurons to simultaneously presented stimuli. J Neurophysiol 2002;88:1128–1135. [PubMed: 12205134]
17. 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]
18. Reddy L, Kanwisher N. Category selectivity in the ventral visual pathway confers robustness to clutter and diverted attention. Curr Biol 2007;17:2067–2072. [PubMed: 17997310]
19. Reynolds JH, Chelazzi L. Attentional modulation of visual processing. Annu Rev Neurosci 2004;27:611–647. [PubMed: 15217345]
20. Desimone R, Duncan J. Neural mechanisms of selective visual attention. Annual Review of Neuroscience 1995;18:193–222.
21. Kastner S, Ungerleider LG. Mechanisms of visual attention in the human cortex. Annu Rev Neurosci 2000;23:315–341. [PubMed: 10845067]
22. Keysers C, Xiao DK, Foldiak P, Perret DI. The speed of sight. Journal of Cognitive Neuroscience 2001;13:90–101. [PubMed: 11224911]
23. 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]
24. Liu H, Agam Y, Madsen JR, Kreiman G. Timing, timing, timing: Fast decoding of object information from intracranial field potentials in human visual cortex. Neuron 2009;62:281–290. [PubMed: 19409272]
25. Serre T, Kreiman G, Kouh M, Cadieu C, Knoblich U, Poggio T. A quantitative theory of immediate visual recognition. Progress In Brain Research 2007;165C:33–56. [PubMed: 17925239]
26. 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]
27. Reddy L, Kanwisher NG, Vanrullen R. Attention and biased competition in multi-voxel object representations. Proc Natl Acad Sci U S A 2009;106:21447–21452. [PubMed: 19955434]
28. Quian Quiroga R, Panzeri S. Extracting information from neural populations: Information theory and decoding approaches. Nature Reviews Neuroscience 2009;10:173–185.
29. Grill-Spector K, Malach R. The human visual cortex. Annual Review of Neuroscience 2004;27:649–677.
30. Logothetis NK, Sheinberg DL. Visual object recognition. Annual Review of Neuroscience 1996;19:577–621.
31. Johnson JS, Olshausen BA. Timecourse of neural signatures of object recognition. J Vis 2003;3:499–512. [PubMed: 14507255]
32. Lamme VA, Roelfsema PR. The distinct modes of vision offered by feedforward and recurrent processing. Trends Neurosci 2000;23:571–579. [PubMed: 11074267]
33. Treisman AM, Gelade G. A feature-integration theory of attention. Cognit Psychol 1980;12:97–136. [PubMed: 7351125]
34. Buschman TJ, Miller EK. Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. Science 2007;315:1860–1862. [PubMed: 17395832]

35. Reddy L, Quiroga RQ, Wilken P, Koch C, Fried I. A single-neuron correlate of change detection and change blindness in the human medial temporal lobe. Curr Biol 2006;16:2066–2072. [PubMed: 17055988]

36. Logothetis NK. The neural basis of the blood-oxygen-level-dependent functional magnetic resonance imaging signal. Philos Trans R Soc Lond B Biol Sci 2002;357:1003–1037. [PubMed: 12217171]

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

38. Kirchner H, Thorpe SJ. Ultra-rapid object detection with saccadic eye movements: visual processing speed revisited. Vision Res 2006;46:1762–1776. [PubMed: 16289663]

39. 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]

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

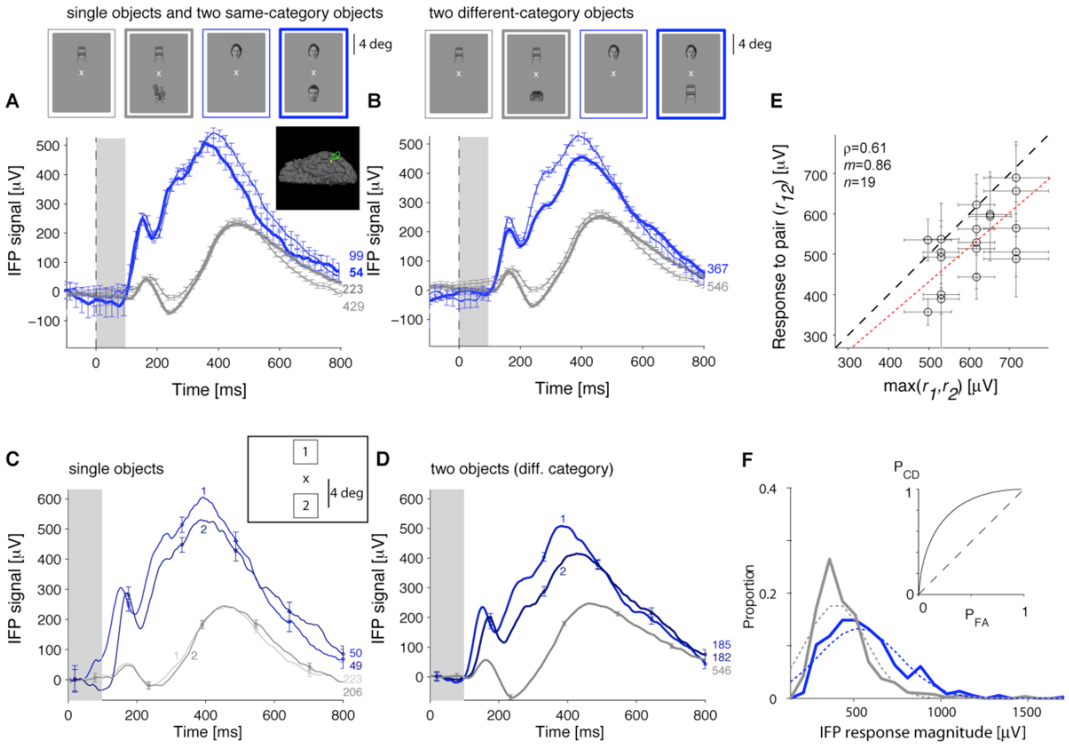

Figure 1. Example of robustness to the presence of a second object in intracranial field potential (IFP) recordings from human visual cortex

Responses of an electrode in the left fusiform gyrus (Talairach coordinates = [−41 −43 −24]; see inset) showing an enhanced response to the "human faces" category ("preferred category", blue) versus other categories (gray) (see more examples in Fig. S1B–E). The stimulus was presented for 100 ms (gray rectangle); responses are aligned to stimulus onset (t=0). The number of repetitions is shown next to each curve. Error bars denote one SEM (shown every 10 time points for clarity). **A**. Responses to one-object images (thin lines) or two objects from the same category (thick lines) averaged over different exemplars and positions. **B**. Responses to two objects from different categories (thick lines) compared to one-object images (thin lines) for images that contain the preferred category (blue) or non-preferred categories (gray). These curves show the responses averaged over different exemplars and positions; Fig. S1A provides the responses for all category pairs. **C**. Responses to one-object images separated by object position (positions "1" and "2"; see inset and Experimental Procedures). Blue indicates the preferred category and gray indicates the non-preferred categories. **D.** Responses to two-object images (objects from different categories) separated by the position of the preferred category. "1": preferred category above the fixation point; "2": preferred category below the fixation point. For the non-preferred categories (gray line), we show the average over all positions (both positions "1" and "2" contained non-preferred category objects). **E**. IFP response magnitude (*max(IFP)−min(IFP)* in [50;300] ms) to two-object images (r_{12}) versus response to the preferred one-object image (max(r_1,r_2)). Each point represents the responses to a pair of exemplars including the preferred-category. The dashed line shows the identity line and the dotted line shows the linear fit (ρ = Pearson correlation coefficient; m=slope). **F**. Distribution of IFP response magnitudes based on the data in **B** for the two-object images containing the preferred category (blue) or non-preferred categories (gray). The dashed curves show a Gaussian fit (two parameters: mean and SD). The inset shows the ROC curve indicating the probability of correctly identifying the preferred category. Departure from the diagonal (dashed line) indicates increased correct detection (P_{CD}) for a given probability of false alarms (P_{FA}).

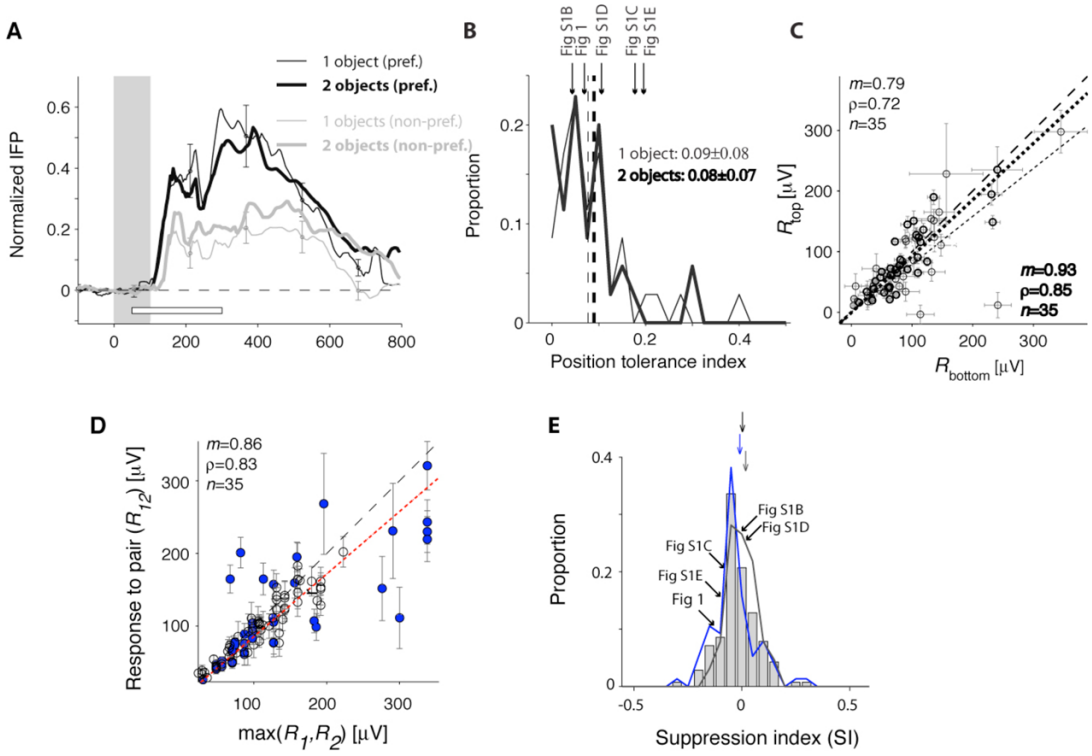

Figure 2. Comparison of the responses across positions and number of objects in the image

A. Average normalized IFP responses to one-object images (thin lines) and two-object images (thick lines) for preferred categories (black) and non-preferred categories (gray). The responses were normalized by subtracting the baseline and dividing by the maximum response to the one-object images containing the preferred category. The white horizontal bar shows the [50;300] ms interval used to define the IFP response magnitude. This figure only included visually selective electrodes, images where the target category was absent and where we had at least 5 repetitions (24 electrodes, 35 categories, Experimental Procedures). **B**. The position tolerance index was defined as $\frac{|R_{pos1} - R_{pso2}|}{\max(R_{pos1}, R_{pos2})}$ where R_{posi} is the response magnitude for the preferred category at position i (i=1,2). Distribution of the position tolerance index for one-object images (thin line) and two-object images (thick line). Bin size=0.025. The dashed line indicates the mean and the arrows denote the examples in Figure 1 and S1 (for the 2-object images). **C**. Comparison of the response magnitudes when the preferred category was in the top position versus the bottom position for one-object images (thin circles) and two-object images (thick circles). The dashed line is the diagonal line and the dotted line is the linear fit. **D**. Comparison of the responses to two-object images (R_{12}) against the maximum of the response to the two corresponding one-object images (max(R_1,R_2)). To compare the responses across electrodes, we subtracted the minimum response in each electrode. The dashed line shows the diagonal line and the dotted line is the linear fit. Blue circles indicate cases where the preferred category was a human face. **E**. Distribution of the response suppression index, defined as $SI = \frac{R_{12} - \max(R_1, R_2)}{\max(R_1, R_2)}$. Negative (positive) values of SI indicate response suppression (enhancement) with respect to the preferred category. Gray bars include all selective electrodes; the blue curve includes those electrodes that showed enhanced responses to human faces and the gray curve includes electrodes with other preferences. The arrows denote the mean of the

distributions (offset vertically for clarity). We indicate the position of the example electrodes from Fig. 1 and S1. Bin size = 0.05.

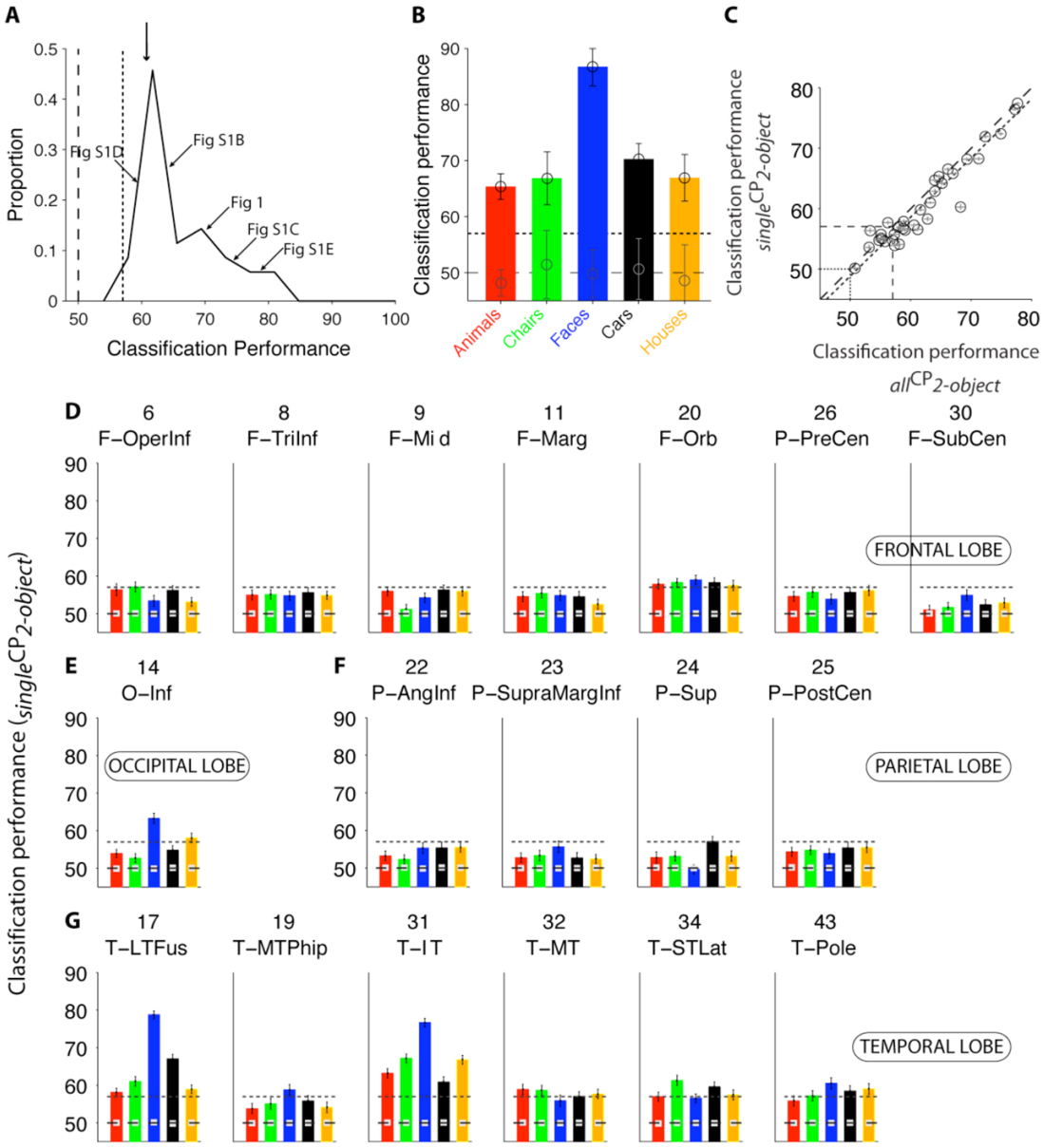

Figure 3. Decoding visual information in single presentations of two-object images

A. Distribution of single-electrode classification performance (CP) values to decode category information in two-object images for all the electrodes that showed visually selective responses (*n*=24). The classifier was trained with single-object images and CP was evaluated using two-object images. The vertical dashed line indicates chance performance (50%) and the vertical dotted line indicates the significance criterion threshold based on shuffling object labels (Experimental Procedures). The arrows point to the examples from Fig. 1 and S1. The downward arrow shows the mean (60±4%). Bin size=2.5%. **B**. Classification performance for each category using a pseudopopulation containing 45 electrodes. Error bars denote one standard deviation over 20 iterations with random choices of units and repetitions for training the classifier (Experimental Procedures). The dashed line indicates chances levels and the dotted line shows the significance threshold. Gray circles show the mean CP for the shuffled labels condition. **C.** Comparison of the CP obtained upon training the classifier using one-object images and evaluating CP with the responses to two-object images ("$_{single}$CP$_{2\text{-}object}$",

y-axis) versus training the classifier with the responses to two-object images and evaluating CP with different repetitions of two-object images ("$_{all}CP_{2\text{-}object}$", x-axis; see also Fig. S5). We only included visually selective electrodes and selective categories. Error bars denote one SEM (20 randomizations of the repetitions used for training). The dashed line is the identity line and the dotted line is the linear fit to the data (35 categories, n=24 electrodes, ρ = 0.96). To compare $_{single}CP_{2\text{-}object}$ against $_{all}CP_{2\text{-}object}$ independently of the number of training points, we randomly downsampled the examples with two-object images to match the examples with one-object images. **D–G**. For each location with ≥10 electrodes (Table S1), we built a pseudopopulation based on the entire data set. The responses of up to 45 electrodes in each location were concatenated. The format is the same as in Fig. 3B. The classifier was trained using one-object images and CP was evaluated using two-object images. Error bars indicate one SEM. The dashed line indicates chance levels (50%) and the dotted line indicates the statistical significance threshold. The gray squares indicate the average CP obtained from 100 random shuffles of the object labels. Table S1 describes the electrode locations and location abbreviations.

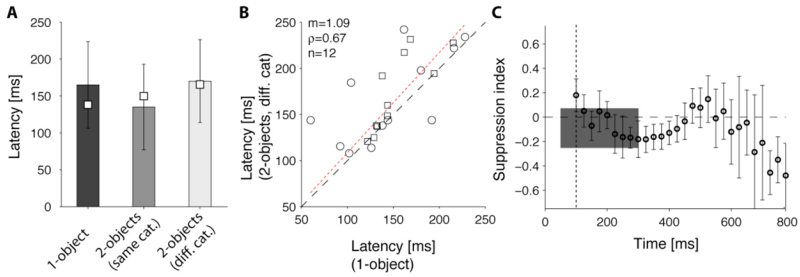

Figure 4. Response latencies and temporal evolution of the response suppression in two-object images

We compared two possible definitions for the latency. *Def1*=first time point where the response to the preferred category exceeded by >20% the response to the non-preferred categories during at least 75 ms. *Def2*=first time point where a one-way ANOVA across object categories yielded $p<0.01$ for 15 consecutive time points. In both cases, the number of repetitions was randomly subsampled so that the number of 1-object repetitions was the same as the number of 2-object repetitions. **A**. Mean response latencies (bars=*Def1*; squares=*Def2*) for the visually selective electrodes. Error bars represent one standard deviation (*Def1*). There was no significant difference between the response latencies to one-object images (black bar) compared to two-object images (light bar) (two-tailed t-test $p>0.3$). **B**. There was a weak but significant correlation between the response latencies for one-object images (x-axis) and 2-object images (y-axis). Circles=*Def1*; Squares=*Def2*. The dotted line indicates a linear fit to the data for *Def1*. **C**. Mean response suppression as a function of time from stimulus onset. For each visually selective electrode, we computed the average IFP in bins of 25 ms and computed the suppression index in each bin as defined in Figure 2. Here we show the mean suppression index in each bin (error bars = SEM). The vertical dotted line denotes the image offset time (t=100 ms). The gray rectangle shows the mean±SD suppression index computed by considering the IFP response magnitude between 50 and 300 ms.