Supporting Online Material for

Fast Readout of Object Identity from Macaque Inferior Temporal Cortex

Chou P. Hung,* Gabriel Kreiman, Tomaso Poggio, James J. DiCarlo

*To whom correspondence should be addressed. E-mail: chouhung@mit.edu

This PDF file includes:

SOM Text
Figs. S1 to S7
References

Fast read-out of object identity from macaque inferior temporal cortex

Hung CP[*], Kreiman G[*], Poggio T, DiCarlo JJ

* These authors contributed equally to this work

McGovern Institute for Brain Research, Center for Biological and Computational Learning, Computation and Systems Biology Initiative, Department of Brain and Cognitive Sciences Massachusetts Institute of Technology, Cambridge, MA 02142, U.S.A.

Supplementary Material (see also (*1*))

Here we summarize the recording procedure and stimulus presentation and describe the classifiers and data analysis methods in detail. For further details about the recordings, stimulus presentation and data preprocessing, see (*2*).

Methods

Recordings and stimulus presentation: Recordings were made from two monkeys (*Macaca mulatta*). We used a set of 77 complex objects rendered in grayscale. Objects were divided prior to the recordings into 8 classes: toys, foodstuffs, human faces, monkey hand/body parts, monkey faces, vehicles, white boxes, and synthetic images of cats and dogs. Object images (3.4 deg) were presented at the center of gaze, unless otherwise stated (Fig. 2), during passive fixation of a central point. The monkeys were rewarded for maintaining fixation within a ± 2 deg window for up to 4 sec. The monkeys were not required to perform object discrimination in this task (although they were also trained on an active discrimination task based on the same stimuli; see below and Fig. S7). Stimulus positions relative to eye fixation were updated at the monitor's frame rate (75 Hz). To approximate stimulus presentation rates observed in free viewing animals (*3*), object images were presented on the screen for 93 ms followed by a 93 ms blank period.

Twenty object images were presented during each fixation period (behavioral trial) and only data from trials in which fixation was maintained for the entire trial were included in the final analyses. Object images were not normalized for mean gray level, contrast or other basic image properties. It is possible to partially read out object category based on some of these simple image properties (*1*). Objects were presented in pseudorandom order, randomly presenting one entire set of the 77 object images before beginning a second randomized set, until each object image had been shown at least 10 times. Penetrations were guided by structural MR images of each monkey and made over a ~10x10 mm area of the ventral surface of Anterior Inferior Temporal Cortex (Horsley-Clark AP: 10-20 mm, ML: 14-24 mm). Recording sites within the same penetration were separated by a few hundred microns (*2*). Multi-unit recordings were made using glass-coated Pt/Ir electrodes (0.5-1.5 MΩ at 1 kHz). Spiking activity (400 Hz-6 kHz) and local field potentials (LFPs, 1-300 Hz) were amplified, filtered, and stored using conventional equipment (*2, 3*). Spike sorting was performed to obtain single unit activity (SUA) from the MUA using the algorithm of Quiroga et al (*4*) with modifications by Alexander Kraskov (*5*).

Data analysis

For any given recording site s ($s = 1,...,N$); let ${}_s r_{ij}$ denote the response during repetition i of object image j, $i = 1,...,n_{rep}$, $n_{rep} \geq 10$, $j=1,...,77$ (for the study of invariance to scale and position $j=1,...,385$). The number of sites was $N = 364$ for MUA data, $N = 71$ for the invariance study with MUA data, $N = 315$ for LFP data, $N = 45$ sites for the invariance study with LFP data, $N = 190$ for SUA data, $N = 20$ for the invariance study with SUA data.

Coding schemes: To quantitatively compare different possible coding schemes, we explored different possible definitions of the response vector **r** (*6-9*). For the spike data, we explored a family of codes based on counting spikes in successive bins of size w within the interval starting t_i ms after stimulus onset and ending t_f ms after stimulus onset (*10-16*). We describe results for

different values of these parameters in the text and Supplementary Figures (*1*). The default condition was $w = 50$ ms, $t_i = 100$ ms and $t_f = 300$ ms (e.g. Fig. 1). The parameter w controls the time resolution of the code; we used w=12.5 ms, 25 ms, 50 ms, 100 ms and 200 ms (see Fig. 3A). We also specifically explored the rapid encoding of information using a single bin by setting $t_f=t_i+w$ (see Fig. 3B). Selective responses in short time windows by single neurons in IT have been observed previously (*17-21*).

Let $c(t_i,w,b)$ denote the number of spikes in the interval $[t_i+bw;t_i+(b+1)w)$ where b is an integer so that the entire interval goes from t_i to t_f. For the local field potential data (Fig. 1C), we divided time in a similar fashion but $c(t_i,w,b)$ was defined as the total power in the corresponding time interval (*22*). The response ***r*** was defined as:

$$\mathbf{r} = [c(t_i,w,0),\ c(t_i,w,1),\ ...,c(t_i,w,b)] \quad \text{where } b = (t_f\text{-}t_i)/w\text{-}1$$

This vector was used as input to the decoding classifier (see below). When considering the responses of multiple sites, we concatenated the corresponding response vectors and used the concatenated vector as input to the classifier. The dimensionality of the input is therefore $(b+1)N$ where N indicates the number of sites. We used $N = 1, 2, 4, ..., 256$ sites (except when we had less than 256 sites); regularization becomes critical, particularly for high dimensions (*23-25*).

Independence assumption: It should be noted that this concatenation step assumes independence among different neurons, an assumption that needs to be revisited upon availability of simultaneous recordings from multiple neurons. It is quite possible that correlations between neurons— which we cannot detect with our independent recordings – may contain additional information and reveal additional aspects of the neural codes ((*26*), see however (*27, 28*)). Our estimates are thus likely to constitute a lower bound on the information represented by small populations of neurons in IT. However, it is interesting that even under these conditions we obtain such a high degree of accuracy in decoding visual information.

Spike bursts: To directly evaluate whether a few spikes within a short time window could constitute an important element of neural coding, we compared the classifier performance for

bursts of spikes against isolated spikes (*28-30*). A spike burst was defined by at least two spikes with an interspike interval < 20 ms. For the spike bursts, *r* was also defined by counting the number of spikes within the burst. The above definitions of the input to the classifiers also apply for the spike bursts and isolated spikes except that we counted spikes only within the corresponding spike classes. Bursts of spikes indeed showed significantly better performance than isolated spikes ($p < 0.01$). Because bursts largely occurred at the beginning of the response, this suggests that initial bursts convey most of the information (*1*).

Single unit activity: Single unit activity (SUA) was obtained by spike sorting (*4*). Both MUA and SUA performed better than local field potentials (LFPs) for the same number of sites (Fig. 1C). This supports our previous observations that spikes are more selective than LFPs (*2*). Performance for MUA and SUA were similar due to a trade-off between sharper selectivity of SUA and more spikes in MUA (1).

Training and testing: The data were always divided into a training set and a test set. In most cases the training set comprised 70% of the repetitions of each object image while the test set included the remaining 30% (see below for other training/testing set combinations). The training set was randomly chosen from all available repetitions and n_{iter} iterations were performed ($n_{iter} = 10$). In the case of studying the extrapolation to different pictures within a class, training was performed on all repetitions using 70% of the objects and testing on all the repetitions of the remaining 30% of the objects (Fig. S3).

Technical details on the experiments described in the main text

Classification and identification: We focused particularly on two tasks which are usually considered different (even if the difference is mostly semantic and they are completely equivalent from a computational point of view, see (*31*)): classification and identification. For *categorization*, the picture labels indicated which out of 8 possible classes the picture belonged to

(toys, foodstuffs, human faces, monkey faces, hand/body, vehicles, white boxes, cats/dogs). Chance was therefore 12.5%. For identification, the picture labels directly indicated the identity of the image (77 possible labels) and chance was therefore 1.3%. We used a one-versus-all classification scheme (*32*). We trained one binary classifier for each class against the rest of the stimuli. For a novel input, the prediction was given by the classifier with the maximum output. Classification performance as shown on all the plots indicates the proportion of correct decoding for *test* data (i.e. data not seen by the classifier during training). As a control, we used data before stimulus onset by evaluating the performance of the classifier using $t_i = -200$ and $t_f = 0$. We also compared the classification results against those obtained by arbitrarily assigning pictures to groups in a random fashion (*1*).

State-of-the-art classifiers: We compared the performance of different statistical classifiers (*23, 33*) including Fisher linear discriminant (FLD) classifier (*34*), Support Vector Machine (SVM) using linear or Gaussian kernels (*23*), Regularized least squares classifier (RLSC, which is a linear classifier, see (*24*)). The SVM classifiers yielded the best performance. Throughout the main text, we used the SVM with linear kernel because its architecture can be easily implemented in cortex as a thresholded sum of weighted synaptic inputs. A quantitative comparison of the performances of these different classifiers is shown in our supplementary web material (*1*). We used the implementation of SVM by Ryan Rifkin (*35*). We initially tested the performance of the classifier on a small subsample of the data exploring a large set of parameters (including different kernel types, different parameters for the kernels, etc.) and then used the optimized parameters for the analysis of the full dataset. The parameters for *SvmFu* were C = 10; N = 10; sigma for Gaussian kernel = 16.

Input vectors and feature selection: In most of the graphs described throughout the text the *N* sites used as input to the classifier were randomly chosen from the total set of *N* sites. This random choice of the sites was repeated at least 20 times (and in most cases 50 times) and we report the average obtained from all of these random neuronal subpopulations. As a very simple

approach to feature selection, we also considered the situation where sites were chosen if they were particularly "good" for the classification task. For this purpose, we defined the signal to noise ratio for a site s ($s=1,...,N$) and a particular stimulus group g ($g=1,...,G$), as:

$$_sSNR_g = \frac{<_s r_g> - <_s r_{notg}>}{\sqrt{_s\sigma_g^2 + _s\sigma_{notg}^2}}$$ where $<.>$ denotes the average over pictures and repetitions and

σ denotes the standard deviation over pictures and repetitions. Sites were ranked for each group based on their *SNR* values. To select *N* "good" sites, we iteratively chose the one with the highest *SNR,* then a different site with the highest *SNR* for a different group and so on. The results of this analysis are shown in Fig. S2.

Invariance of IT representation: Although previous studies reported different ranges of scale and position invariance for individual neurons to specific preferred objects (*36-43*), the combination of highly accurate decoding that is selective and invariant to a significant degree across a large collection of objects can only be achieved at the population level as shown here. To assess the degree of invariance of the population activity we studied the performance of the classifier in extrapolating to shifted or scaled versions of the images (Fig. 2). Each picture was presented in 5 different conditions: (1) center of gaze, 3.4^o size (standard condition); (2) center of gaze, 1.7^o size; (3) center of gaze, 6.8^o size; (4) 2° shift to the contralateral side, 3.4^o size; (5) 4° shift to the contralateral side, 3.4^o size. For the invariance to scale and position changes, the training set consisted of all repetitions at a particular scale and position while the testing set consisted of all repetitions at all other scales and positions. For example, we trained the classifier to learn the map between the neuronal responses and the category or identity of all images presented at the standard condition (condition 1 above) and then we tested its performance using the neuronal responses to all other conditions (conditions 2 through 5 above). This case corresponds to the second bar in Fig. 2. Similarly, we trained the classifier using the neuronal responses from condition 2 and tested on the rest, etc. (Fig. 2).

Invariance of IT representation for novel objects: In most models of visual object recognition and in many computer vision algorithms (*41, 44-52*) the ability to generalize across (e.g.) changes in scale and position may be 'learned' from visual experience of a specific object (see however (*48, 49, 51, 53, 54*)). To examine this possibility, we tested the ability of individual IT neurons to generalize their selectivity across the same changes in position and scale for novel objects that were previously unseen by the animal. We directly compared the degree of position and scale invariance for novel objects (never viewed by the animal until the time of recording, 10 novel objects per recording site) vs. familiar objects (each previously viewed by the animal ~500 times at each of the transformed scales and positions and ~30,000 times at the standard scale and position). This comparison could not be made at the population level because new objects had to be used for each single-electrode recording session. Instead, we made a direct comparison of position and scale invariance of novel versus familiar objects at each site. To make this comparison, 10 of the 77 familiar objects were chosen at each site that had response rates (in the standard position and scale) that matched the response rates produced by the 10 novel objects tested at that site (within 30% in each case). At a typical site (Fig. S6A), the generalization of selectivity was similar for both familiar and novel objects in that, for both familiar and novel sets of 10 objects, the same pattern of selectivity was found for all tested positions and scales. We calculated an invariance index for all selective sites (Fig. S6B): the invariance index was not different between familiar and novel objects (means 0.46 and 0.51 respectively, $p > 0.2$, paired t-test, $N = 13$ sites). This shows that scale and position invariance for arbitrary sets of visual objects does not require previous exposure to those particular objects. Whether such invariance derives from a lifetime of previous experience with other related objects (e.g. shared features) or from innate properties of the visual system or both, remains to be determined. In any case, the observation that the adult IT population has significant position and scale invariance for arbitrary 'novel' objects provides a strong constraint for any explanation of the computational architecture and function of the ventral visual stream.

Comparison between passive and active viewing: In order to achieve a sufficient number of trials and sites across a large set of stimuli (e.g. the test of invariance with novel stimuli required 4400 presentations per site), the results were based on data collected while the animal passively fixated the rapid sequence of stimuli. However, it is possible that the results may differ during active discrimination of the objects. To address this question, we analyzed the responses of 104 sites tested under both passive fixation and an active task. Both animals were also trained on an active discrimination task using pairs of objects chosen from the set of 77 objects. Stimulus size and position were identical across the two tasks. The mean ranks of the responses to the two objects used during the active task were 2.2 and 65.3 (1 being the strongest and 77 being the weakest; box stimuli were avoided in this task). We hereafter refer to these two objects as stimulus A (preferred) and stimulus B (non-preferred). This 'active' task (which we call Multiple Choice Match to Sample) required the animals to discriminate a 200 ms cue object from two possible choices. After a 500 ms delay period, two targets appeared at 8 deg eccentricity left and right of the cue along the horizontal meridian, and reward was given for choosing (saccading to and holding fixation on) the matching object. To minimize the effect of differences in the duration of object presentation, responses were analyzed within the [100:200) ms window post stimulus onset, instead of the [100:300) ms window. We observed overall a lack of change in the discriminability ($d_a = \dfrac{\overline{A} - \overline{B}}{\sqrt{\dfrac{\sigma_A^2 + \sigma_B^2}{2}}}$) of the two objects across the two tasks (Fig. S7; $n = 98$, $p > 0.9$). Although in general the differences between passive and active viewing may be strongly task-dependent, these results suggest that, at least to a first order approximation, the performance of the classifier would not be substantially different if the animal were performing an active discrimination task instead of passive viewing.

A computational sanity check: We compared the degree of scale and position invariance observed in the neuronal recordings against the responses of units from a model of

object recognition in the ventral stream (*31*). S1 units in the model (corresponding to simple cells in primary visual cortex) did not show scale and position invariance whereas C2 units (which are a proxy for IT units since they are the inputs to IT units in the model) showed robust invariance to changes in the scale and position in the visual field of the stimulus (*1*).

Read out of scale and position: The question of whether IT populations also carry at least coarse information about scale and position – irrespective of category and identity – was addressed as follows. A classifier was trained with 70% of the repetitions of all pictures at 3 different positions (see above and Fig. 2); performance was then evaluated on the remaining 30% of the repetitions. We tested read-out of scale at a fixed position and read-out of position at a fixed scale. Thus, the training set consisted of 70% of the repetitions for all pictures and the labels of each object consisted of the 3 possible scales for scale read-out or positions for position read-out (irrespective of the object identity or category). The small number of scales and positions testes implies that the resolution of our analysis is quite limited. Scale and position could be read-out from the same small population of neurons. The results of this analysis are shown in Fig. S4.

Read out of stimulus onset: The question of whether IT populations also carry information about the time of onset of a visual stimulus (Fig. S5) was approached by training a binary classifier using half of the repetitions and testing its performance on the remaining half. The labels were +1 (picture on) and -1 (picture off). For any time point t, the input to the classifier was a vector containing the responses of N sites from time $t+\tau$ until time $t+\tau+\eta$ using a bin size of 12.5 ms. Responses from multiple sites were concatenated assuming independence as discussed above. We explored the following values for the time lag τ: 12.5, 25, 50, 100 and 200 ms and we explored the following values for the integration time η: 12.5, 25, 50 and 100 ms.

Robustness of the IT representation: Fig. S1 shows robustness to neuronal drop-out and spike deletion (*55*). Neuronal drop-out is meant to simulate neuronal or synaptic death while spike deletion is mean to simulate failures in spike transmission or neurotransmitter release. To

study the robustness to neuronal drop-out, we trained the classifier as explained above. During testing, a given percentage of the sites were randomly removed and the classifier was forced to reach a decision using the remaining sites (Fig. S1A). To study robustness to spike deletion, a random percentage of spikes were removed from both the training and testing phases (Fig. S1B).

Supplementary Figures Captions

Fig. S1: The neural code is robust to neuronal drop-out and spike deletion

A. Training of the classifier (red = categorization, blue = identification) was performed as described in the Methods. Here, before testing the performance of the classifier, a proportion of sites (indicated in the x-axis) were removed from the classifier (simulating the process of neuronal death or axonal death). Dashed lines = chance performance. The error bars next to the dashed lines show the range of performances obtained using the 100 ms before stimulus onset (control). **B**. Here, a proportion of spikes (indicated in the x-axis) was removed from the classifier (red = categorization, blue = identification) both for training and testing (simulating the process of failures in spike transmission or neurotransmitter release). Dashed lines = chance performance. The error bars next to the dashed lines show the range of performances obtained using the 100 ms before stimulus onset (control).

Classifier parameters (for parts **A** and **B**): MUA, [100;300) ms time interval, bin size = 50 ms. Error bars show SD over 20 iterations of random choices of $N = 256$ sites.

Fig. S2: Specific wiring significantly improves classifier performance

Classification performance on test data as a function of the number of sites for categorization (**A**) or identification (**B**) using randomly selected sites (solid lines) or pre-selected sites (dotted lines). For a given site and a given group, the SNR was defined as the difference between the mean response to that group and the mean response to other groups divided by the sum of the standard deviations of the group and non-group responses (see Methods). This yielded a list of the "top" sites for each classification task, using only the training set. Sites were ranked based on the SNR and the sites with the highest SNR (on the training set) were selected in the "best SNR sites" case.

Classifier parameters: MUA, [100;300) ms interval, bin size = 50 ms. Dashed lines = chance performance. Error bars next to the dashed lines show range of performance in the -100 to 0 ms interval.

Fig. S3: Extrapolation to novel pictures within the same categories

In this Figure, the classifier was trained with all the repetitions for 70% of the pictures and testing for categorization performance was done with all repetitions for the remaining 30% of the pictures (green). Thus, the classifier never saw the neuronal responses to the test pictures during the training phase. For comparison, we also show the performance upon training on 70% of the repetitions with all pictures and testing on the remaining 30% of the repetitions for all pictures (red, as shown in Fig. 1). Classifier parameters: [100;300) ms interval, bin size = 50 ms. Error bars show standard deviations over 20 iterations of random choices of sites.

Fig. S4: Scale and position ('where') can be read out independently of object identity ('what')

A Read-out performance for the scale read-out task, position read-out task, categorization task and identification task. To read out object scale and position the classifier was trained on 70% of the repetitions of the 77 pictures at different scales and positions (using the standard set and the two additional scales or the standard set and the two additional positions). The example labels indicated only the scale or the position of the object, *regardless* of the object identity or category. Performance was evaluated by asking what the object scale or position was for the remaining 30% of the repetitions. The dashed line indicates chance performance (which is different for each task). Read-out of scale was tested at a fixed position and read-out of position was tested at a fixed scale; chance was 1/3 for read-out of scale and position. The blue rectangles indicate the

range of performances for the interval [-100;0) ms with respect to stimulus onset (control). Classifier parameters: MUA, $N = 64$ sites, [100;300) ms interval, bin size = 50 ms.

B. Classifier performance in reading out stimulus category (red), scale (blue) or position (green) as a function of the cumulative time from stimulus onset in bin sizes of 12.5 ms. The classifier was trained on all images as in part A (using 70% of the repetitions). The dashed lines show chance levels (1/3 for scale and position, 1/8 for category).

C. To explore the neural code underlying different types of information (identity, position, scale) expressed by the same population of neurons we used a standard variable selection technique (see Methods above). Here we show the SNR for scale read-out (green) or position read-out (blue) as a function of the SNR for categorization for each site. The dashed lines represent a linear fit to the data ($r = 0.06$ for scale, $r = 0.17$ for position). There was little correlation between category read-out SNR and either scale read-out SNR or position read-out SNR. This implies that different neurons represent distinct kinds of information (about identity, position, scale) but to a different extent. Thus an individual neuron's spike rate may, for example, depend more on identity than scale and more on scale than position. In contrast, there was a stronger correlation between the SNR values for identification and categorization ($r = 0.54$, (*1*)).

Fig. S5: Read-out of stimulus onset

A classifier was trained to indicate whether any object was present or not at each time point (bin size = 12.5 ms, see Methods above). (**A**, **B**) Two different epochs showing actual object image presentation scheme (black) and classifier predictions (red). (**C**) Classification performance as a function of the number of sites used to train the classifier. Error bars show S.D. over 20 random choices of sites. The dashed line shows chance performance (50%).

Fig. S6: Invariance to novel objects

(**A**) Example of a site's response to 10 familiar (left) and novel (right) objects at the 5 different scales and positions (C = center of gaze, 3,4^o size; S1=center of gaze, 1.7^o size; S2 = center of gaze, 6.8^o size; P1 = 2° shift, 3.4^o size; P2 = 4° shift, 3.4^o size. The response of the unit (spike count in the 100 to 200 ms interval) is color-coded (axis next to response plot in spikes/s).

(**B**) Summary showing the degree of invariance for novel objects versus familiar objects. In each site, the invariance index was calculated as the average Spearman correlation coefficient for the 10 object responses across all possible pairings of the five position and scale conditions that were tested (i.e. 1 indicates complete invariance and 0 indicates no invariance). Results are based on the 13/19 sites which were selective (ANOVA $p < 0.05$) among the 10 novel stimuli in the standard condition. There was no significant difference between the invariance index for novel and familiar objects (paired t-test, $p > 0.2$).

Fig. S7: Comparison of responses during active versus passive tasks

During passive fixation (used for all experiments described in this work except this one), the animals were rewarded for maintaining fixation. In the active task, two of these objects were chosen based on the ranking of the responses to the 77 objects. The animal was required to perform a Multiple Choice Match-to-Sample task in which the animal discriminated between the two objects during a 200 ms cue period. Reward was given for choosing the matching target object after a 500 ms delay period. Response rates were compared within the [100:200) ms window post stimulus onset in both tasks. (**A**) The scatter plot compares the discriminability (d$_a$) of the two objects across the two tasks. The dashed line is the identity line and the solid line is a linear fit to the data. (**B**) Distribution of the log$_2$ ratio of the discriminability across the two tasks (mean ratio = -0.0023, two-tailed t test $p > 0.9$, see Supplementary Material).

References and Notes

1. http://cbcl.mit.edu/software-datasets/kreiman/fastreadout.html
2. G. Kreiman, C. Hung, T. Poggio, J. DiCarlo, *AI Memo* **2004-020** (2004).
3. J. DiCarlo, H. Maunsell, *Nat. Neurosci.* **3**, 814 (2000).
4. R. Quian Quiroga, N. Nadasdy, Y. Ben-Shaul, *Neural Comp.* **16**, 16161 (2004).
5. A. Kraskov, personal communication
6. G. Kreiman, *Physics of Life Reviews* **2**, 71 (2004).
7. R. deCharms, A. Zador, *Ann. Rev. Neurosci.* **23**, 613 (2000).
8. L. F. Abbott, *Quart. Rev. Biophys.* **27**, 291 (1994).
9. A. Pouget, P. Dayan, R. Zemel, *Ann. Rev. Neurosci.* **26**, 381 (2003).
10. W. Bair, *Curr. Op. Neurobiol.* **9**, 447 (1999).
11. C. Koch, *Biophysics of Computation* (Oxford University Press, New York, 1999).
12. M. N. Shadlen, W. T. Newsome, *Curr. Op. Neurobiol.* **4**, 569 (1994).
13. F. Rieke, D. Warland, R. van Steveninck, W. Bialek, *Spikes* (The MIT Press, Cambridge, Massachusetts, 1997).
14. L. M. Optican, B. J. Richmond, *J. Neurophys.* **57**, 162 (1987).
15. N. C. Aggelopoulos, L. Franco, E. T. Rolls, *J Neurophysiol* **93**, 1342 (2005).
16. L. F. Abbott, E. T. Rolls, M. J. Tovee, *Cereb. Cortex* **6**, 498 (1996).
17. M. Tovee, E. Rolls, *Visual cognition* **2**, 35 (1995).
18. C. Keysers, D. K. Xiao, P. Foldiak, D. I. Perret, *J. Cog. Neurosci.* **13**, 90 (2001).
19. G. Kovacs, R. Vogels, G. A. Orban, *Proc Natl Acad Sci U S A* **92**, 5587 (1995).
20. E. T. Rolls, M. J. Tovee, S. Panzeri, *J Cogn Neurosci* **11**, 300 (1999).
21. G. Kovacs, R. Vogels, G. A. Orban, *Proc. Natl. Acad. Sci. USA* **92**, 5587 (1995).
22. C. Mehring *et al.*, *Nat. Neurosci.* **6**, 1253 (2003).
23. T. Poggio, S. Smale, *Notices of the AMS* **50**, 537 (2003).
24. R. Rifkin, G. Yeo, T. Poggio, in *Advances in Learning Theory: Methods, Model and Applications* Suykens, Horvath, Basu, Eds. (VIOS Press, Amsterdam, 2003), vol. 190, pp. Chapter 7, 131-154.
25. T. Hastie, R. Tibshirani, J. Friedman, *The elements of statistical learning* (Springer-Verlag, New York, 2001).
26. P. Steinmetz *et al.*, *Nature* **404**, 187 (2000).
27. N. C. Aggelopoulos, L. Franco, E. T. Rolls, *J. Neurophys.* **93**, 1342 (2005).
28. R. Krahe, G. Kreiman, F. Gabbiani, C. Koch, W. Metzner, *J. Neurosci.* **22**, 2374 (2002).
29. N. A. Lesica, G. B. Stanley, *J. Neurosci.* **24**, 10731 (2004).
30. R. Krahe, F. Gabbiani, *Nat. Rev. Neurosci.* **5**, 13 (2004).
31. M. Riesenhuber, T. Poggio, *Nat. Neurosci.* **3 Suppl**, 1199 (2000).
32. R. Rifkin, A. Klautau, *Journal of Machine Learning Research* **5**, 101 (2004).
33. T. Hastie, R. Tibshirani, J. Friedman, *The elements of statistical learning*. Springer-Verlag, Ed., Springer Series in Statistics (Basel, 2001).
34. C. M. Bishop, *Neural Networks for Pattern Recognition* (Clarendon Press, Oxford, 1995).
35. http://cbcl.mit.edu/software-datasets/
36. D. Perrett, J. Hietanen, M. Oeam, P. Benson, *Phil. Trans. Roy. Soc.* **355**, 23 (1992).
37. E. Rolls, *Curr. Op. Neurobiol.* **1**, 274 (1991).
38. R. Desimone, T. Albright, C. Gross, C. Bruce, *J. Neurosci.* **4**, 2051 (1984).
39. N. K. Logothetis, J. Pauls, T. Poggio, *Curr. Biol.* **5**, 552 (1995).
40. M. Ito, H. Tamura, I. Fujita, K. Tanaka, *J Neurophysiol* **73**, 218 (1995).
41. G. Wallis, E. T. Rolls, *Prog. Neurobiol.* **51**, 167 (1997).

42.   K. Tanaka, *Ann. Rev. Neurosci.* **19**, 109 (1996).
43.   N. K. Logothetis, D. L. Sheinberg, *Ann. Rev. Neurosci.* **19**, 577 (1996).
44.   J. Rowley, S. Baluja, T. Kanade, *IEEE PAMI* **20**, 23 (1998).
45.   K. Sung, T. Poggio, *IEEE PAMI* **20**, 39 (1998).
46.   D. I. Perrett, M. W. Oram, E. Ashbridge, *Cognition* **67**, 111 (1998).
47.   B. Tjan, *Advances in NIPS* **13**, 66 (2001).
48.   C. Wallraven, H. Bulthoff, *Proceedings of CVPR* (2001).
49.   G. Wallis, *Network* **9**, 265 (1998).
50.   B. A. Olshausen, C. H. Anderson, D. C. Van Essen, *J Neurosci* **13**, 4700 (1993).
51.   M. Riesenhuber, T. Poggio, *Nat. Neurosci.* **2**, 1019 (1999).
52.   P. Foldiak, *Neural Comp.* **3**, 194 (1991).
53.   K. Fukushima, *Biol. Cybern.* **36**, 193 (1980).
54.   B. Mel, *Neural Comp.* **9**, 777 (1997).
55.   G. Kreiman, R. Krahe, W. Metzner, C. Koch, F. Gabbiani, *J. Neurophys.* **84**, 189 (2000).

Supplementary Figure 1

Supplementary Figure 2

A   'Categorization'

best SNR sites

random sites

Classification Performance

Number of Sites

B   'Identification'

Number of Sites
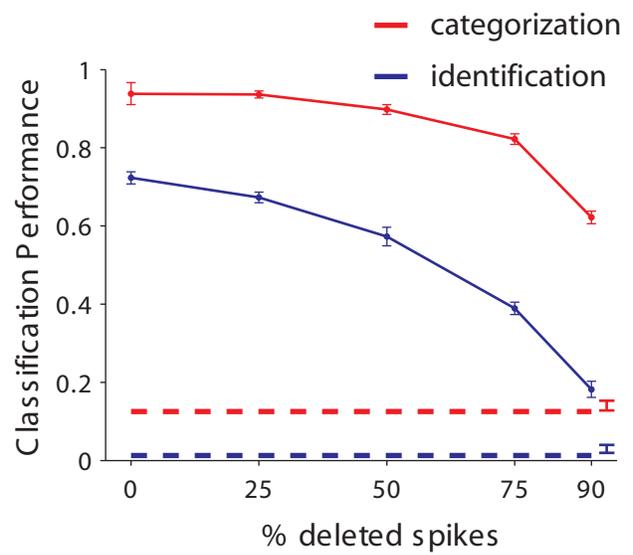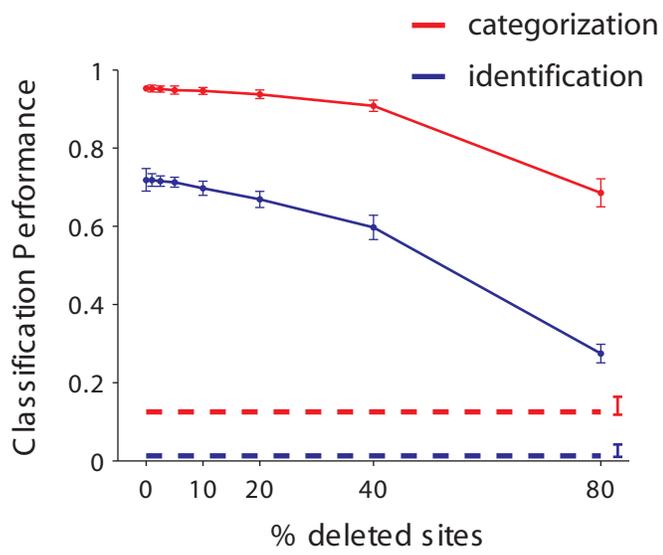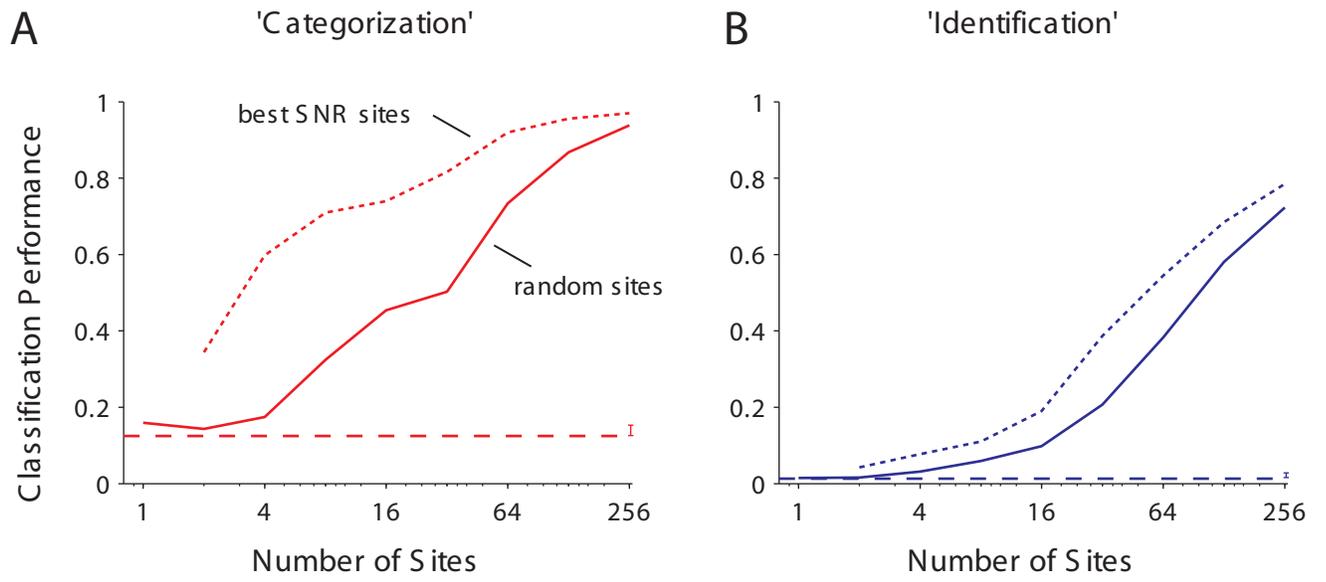
Supplementary Figure 3
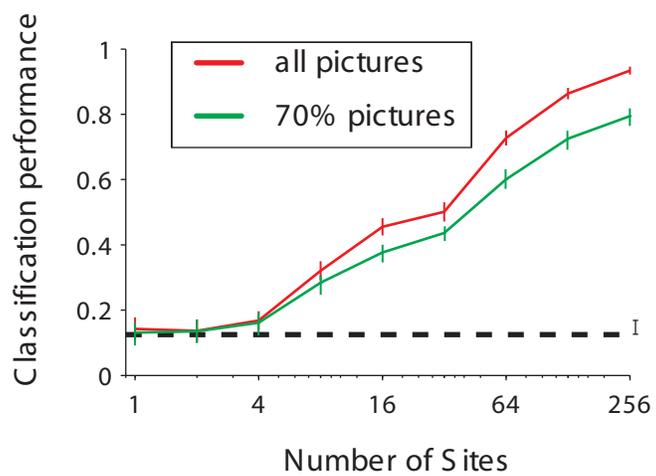
Supplementary Figure 4

Supplementary Figure 5

A

Classifier predictions

Stimulus on/off

C

On

Off

3.8   4.0   4.2

100 ms

B

On

Off

12.6   12.8   13

Time (s)

Classification performance

1

0.5

1   4   16   64   256

Number of Sites

Supplementary Figure 6

A

B

Supplementary Figure 7

A

B