Computer Science and Artificial Intelligence Laboratory

Technical Report

MIT-CSAIL-TR-2005-045
AIM-2005-022
CBCL-253

July 6, 2005

Ultra-fast Object Recognition from Few Spikes

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

massachusetts institute of technology, cambridge, ma 02139 usa — www.csail.mit.edu

Abstract

Understanding the complex brain computations leading to object recognition requires quantitatively characterizing the information represented in inferior temporal cortex (IT), the highest stage of the primate visual stream. A read-out technique based on a trainable classifier is used to characterize the neural coding of selectivity and invariance at the population level. The activity of very small populations of independently recorded IT neurons (~100 randomly selected cells) over very short time intervals (as small as 12.5 ms) contains surprisingly accurate and robust information about both object 'identity' and 'category', which is furthermore highly invariant to object position and scale. Significantly, selectivity and invariance are present even for novel objects, indicating that these properties arise from the intrinsic circuitry and do not require object-specific learning. Within the limits of the technique, there is no detectable difference in the latency or temporal resolution of the IT information supporting so-called 'categorization' (a.k. basic level) and 'identification' (a.k. subordinate level) tasks. Furthermore, where information, in particular information about stimulus location and scale, can also be read-out from the same small population of IT neurons. These results show how it is possible to decode invariant object information rapidly, accurately and robustly from a small population in IT and provide insights into the nature of the neural code for different kinds of object-related information.

[*]The authors, Chou Hung and Gabriel Kreiman, contributed equally to this work.

Supplementary Material is available at http://ramonycajal.mit.edu/kreiman/resources/ultrafast/.

This report describes research done at the Center for Biological & Computational Learning, which is in the McGovern Institute for Brain Research at MIT, as well as in the Dept. of Brain & Cognitive Sciences, and which is affiliated with the Computer Sciences & Artificial Intelligence Laboratory (CSAIL).

This research was sponsored by a Whiteman Fellowship and the Government grant from the Office of Naval Research (DARPA) Contract No. MDA972-04-1-0037, Office of Naval Research (DARPA) Contract No. N00014-02-1-0915, National Science Foundation (ITR/SYS) Contract No. IIS-0112991, National Science Foundation (ITR) Contract No. IIS-0209289, National Science Foundation-NIH (CRCNS) Contract No. EIA-0218693, National Science Foundation-NIH (CRCNS) Contract No. EIA-0218506, and National Institutes of Health (Conte) Contract No. 1 P20 MH66239-01A1.

Additional support was provided by:
Central Research Institute of Electric Power Industry (CRIEPI), Daimler-Chrysler AG, Compaq/Digital Equipment Corporation, Eastman Kodak Company, Honda R&D Co., Ltd., Industrial Technology Research Institute (ITRI), Komatsu Ltd., Eugene McDermott Foundation, Merrill-Lynch, NEC Fund, Oxygen, Siemens Corporate Research, Inc., Sony, Sumitomo Metal Industries, and Toyota Motor Corporation.

Primates can recognize and categorize objects as quickly as 200 ms after stimulus onset (*1, 2*). This remarkable ability underscores the high speed and efficiency of the object recognition circuitry thought to be mediated by the ventral visual pathway, running from primary visual cortex (V1) to inferior temporal cortex (IT) in primates (*3-5*). At the end of the ventral stream, single cells in IT show selectivity for complex visual stimuli such as faces and other objects, with some tolerance to changes in object size and position (*5-8*). It is thus a reasonable hypothesis that small groups of neurons in IT cortex tuned to different objects and object parts could provide sufficient information for several visual recognition tasks, including tasks traditionally distinguished (though equivalent from a computational point of view, see (*4*)) as identification, categorization, expression estimation, etc. This information could then be 'read out' by circuits in other areas, such as prefrontal cortex, which receive input from tuned IT neurons and may combine them with weights appropriate for different tasks.

Although a body of physiological and functional imaging data (*3-19*) suggests that visual object identity and category could in principle be *read out* from a population of IT neurons, fundamental aspects of this code remain under debate and are not sufficiently characterized at the population level to provide quantitative constraints for models of visual object recognition. First, the format of the population coding of object selectivity within IT remains uncertain (in terms of discriminative power in relation to population size, temporal resolution, time course and synergy between neurons), as previous studies have examined stimulus sets limited in size or scope, or were based on selected subsets of neurons. Second, it remains unclear how the response properties of individual IT neurons translate to the combination of selectivity *and* invariance at the neural population level, whose trade-off is the fundamental requirement for object recognition. Third, although visual experience has been shown to alter neuronal selectivity at multiple levels of the visual system, the role of visual experience in invariance properties is unclear. Finally, although some have suggested that different visual recognition tasks are supported by distinct neuronal populations and codes, this has not been directly tested at the population level.

In the present study, we examine these issues by obtaining independent recordings from a large unbiased sample of IT neuronal sites and using a population read out technique based on state-of-the-art classifiers (see Supplementary Material). With this approach (N1) the effectiveness of different possible neural codes can be compared in a direct and quantitative manner (*11, 20-24*). The performance of the classifier constitutes a lower bound on the information content available in the neural activity, but it is a meaningful measure that could be directly implemented by neuronal hardware. The results provide an unprecedented view of the population code available within IT, based on the responses of over 367 neuronal sites (sequentially collected) and 190 well isolated neurons to 77 objects (common to all recordings) and 130 novel objects.

We began by using the classifier approach to determine the ability of the IT population to 'categorize' the 77 visual stimuli as belonging to one of eight possible groups (Supplemental Methods). To do this, we recorded the IT responses to 77 grayscale objects presented to passively viewing monkeys at a standard object position (center of gaze) and size (3.4 deg). The eight object categories (N2) were defined prior to the experiment and consisted of foods, toys, monkey faces, human faces, vehicles, hand/body parts, cats/dogs, and box outlines (Figure 1A, all images are shown in Supplementary Figure 1B). Figure 1B (red curve) shows the cross-validated performance of classifiers in performing this categorization task as a function of the number of recording sites (N3). The performance is remarkably good: for example, the spiking activity of 128 randomly selected MUA sites is sufficient to categorize the objects with 86±6% accuracy (mean±s.d., chance=12.5%). Similarly, we tested the ability of the IT population to *identify* each

of the 77 objects (i.e. which specific object had been presented). Figure 1B shows that even small populations of IT neurons are capable of performing this task at high accuracy (for 128 sites, 58±7% correct, chance = 1.3%), although at lower performance than categorization for the same number of sites. Very similar levels of performance were obtained when only isolated single unit activity (SUA) was considered (N9 and Figure 1C). Notably, even the local field potentials (LFP) in IT contain some information about object category (N9 and Figure 1C).

Importantly, the performance values plotted in Figure 1B, C are based on the responses of *single stimulus presentations* that were not included in the training of the classifiers. Thus, the level of recognition performance is what a real downstream neuron could, in theory, perform on a single trial by simply counting spikes over a short time interval (100-300 ms interval and 50 ms time bins in this case, and using those counts in a nonlinear -- or even a linear -- classifier (see Supplementary Figure 14 and Methods). This is remarkable considering the supposedly high trial-to-trial variability of cortical neurons (*25, 26*). In addition to robustness to spiking variability, the IT population performance is also very robust to both significant deletions of neurons during testing and to large deletions of spikes simulating failures in neurotransmitter release (N4). The results shown in Figure 1 (and in other studies decoding neuronal responses) assume precise knowledge about stimulus onset. Interestingly, we could also accurately read out the stimulus onset information based on the neuronal responses (Supplementary Figure 15). In sum, whereas previous studies have demonstrated that responses of selected subsets of IT neurons provide some information about a stimulus (*3, 6, 7, 9-11, 21*), Figure 1 directly demonstrates that even small, randomly-selected groups of independent IT neurons can convey substantial, single-trial information about object identity and category.

A key computational difficulty of object recognition, however, is that it requires *both* selectivity and generalization (invariance) to image transformations that do not alter object identity (*5, 6, 9*). The main achievement of mammalian vision and one reason why it is still so much better than artificial vision is the combination of high selectivity and robust invariance – indeed, the main computational goal of the processing steps between V1 and IT is its superb trade-off between stimulus selectivity and invariance (*4*).

The results presented in Figure 1 correspond to training and testing on the same visual images (although not the same trials). The performance on the categorization task shows that the IT population is capable of supporting generalization over objects within pre-defined categories (Figure 1B, N6). We explored the ability of the IT population to generalize recognition of object identity over changes in position and scale by testing an additional 71 sites with the original 77 images and four transformations in position or scale (two scales and two positions). We found that we could reliably classify (with less than 10% reduction in performance) the objects across transformations in scale and position upon training the classifier at only one particular scale and position (Figure 2). This shows that the representation in IT is both selective and invariant in highly a non-trivial manner. In particular, although neuronal population selectivity for objects could be obtained from areas like V1, this selectivity would not generalize over changes in (e.g.) position (N5) where the object image occupies a completely non-overlapping position in the visual field (as in Figure 2). We also tested generalization across different exemplars of the same class (N6).

In most models of visual object recognition that have been quantitatively implemented (in the computer vision literature, see (*27-30*) and among the more biological models see (*31*)), the ability to generalize across (e.g.) changes in size and position is 'learned' from temporally-correlated exposure to different views of the same object (or via biologically less plausible

mechanisms (*32*), see however (*4, 33, 34*)). An extreme -- and commonly accepted -- version of this hypothesis predicts that such learning is required for each particular object at essentially each position and scale. 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 where previously unseen by the animal. We compared the generalization across position and scale with novel objects (10 novel objects per recording site, 130 objects total) with that observed for familiar objects. At a typical site (Figure 3A), the generalization of selectivity was similar for both familiar and novel objects in that, for both novel and familiar 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 (Figure 3B): 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 (supporting the findings in (9) and the model in (4)). Whether such invariance derives from a lifetime of previous experience with other related objects (e.g. shared features) or from genetic 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 (4).

A key advantage of the read-out approach is that it allows us to directly examine a range of potential neuronal object codes. In this paper we focus on temporal properties of such codes, by examining the temporal resolution of the object information conveyed by IT. We studied the temporal resolution of the classification performance using the activity within a time window from 100 to 300 ms after stimulus onset (the latency of IT neurons is well established to be ~100 ms (*12*), see also Figure 4B). The classification performance obtained from simply counting all spikes in that 200 ms window was first computed. Performance improved when the data in this window were cut into smaller bins (bin sizes ranging from 12.5 to 200 ms; Figure 4A) but we observed little improvement in performance for bin sizes smaller than ~50 ms. This implies that neurons do not need to integrate information over long time periods of several hundred ms (N7).

We next examined the time course of the object representation both for 'categorization' and 'identification' tasks (*35, 36*). We evaluated the identification and categorization performance of the population on individual 12.5 ms bins at a range of latencies from stimulus onset (Figure 4B). The time course did not depend strongly on stimulus grouping (N2) or on the difficulty of the classification task (Figure 4B, see also Supplementary Figure 11). Results were similar when bins sizes of 25 ms, 50 ms and 100 ms were used and also for the read out of stimulus category or identity under different scales or positions (See Supplementary Figure 9).

Strikingly, we could decode object category at 70±3% accuracy with the information from 256 sites in only *one* bin of 12.5 ms at 125 ms latency, which we will call the maximally informative latency (MIL) (Figure 4B). Given the firing rate of these neurons, this bin size typically contained no spike or at most two spikes (0.18±0.26 spikes/bin, mean±s.d.). This suggests that a few spikes from a small number of neurons (essentially a binary vector with either 1 or zero components) may be enough to encode 'what' information (N7, N8).

AIT is generally regarded as the brain area at the top of the ventral stream, whose main goal is to describe *what* is in the image, irrespectively of position and scale and other transformations. The trade-off between selectivity and invariance confirmed by our data thus may appear to suggest that detailed position information about the image may be lost in IT. Surprisingly, it turns out that it is also possible to read out both object size and position ('where' information) based on the

activity of the same population of neurons (Supplementary Figure 10A). Reading out object position or scale had a similar time-course to the read-out of object category (Supplementary Figure 10B) but there was little correlation between the signal to noise ratio of each site towards decoding scale/position vs. decoding object category suggesting that the nature of the underlying code is based on individual neurons all encoding but with different weights the two types of information (Supplementary Figure 10C).

Our observations characterize the available information in IT for object recognition but they do not necessarily imply that the brain utilizes the same coding schemes that we have used or the same algorithms for decoding. However, a linear classifier could easily be implemented by setting appropriate synaptic weights for the synapses in single neurons. Thus, neurons in the areas to which AIT units project, such as prefrontal cortex, could decode information over brief time intervals, using inputs from small neuronal populations (e.g. ~ 100 neurons) in IT. It is for instance conceivable that dynamic setting of the synaptic weights from AIT to prefrontal cortex (PFC) may switch between different tasks in PFC, reading out information from the neuronal population in inferior temporal cortex. In this perspective, tuned neurons in IT would be similar to "centers" in a learning network (25), supporting a range of different recognition tasks including 'categorization' and 'identification' in PFC (24).

The approach described here is a natural and powerful one to characterize properties of the information represented in a cortical area. A classifier can be trained on any stimulus property of interest and then tested to decide whether the relevant information is available in the neural activity and what is its neural code. Our results quantitatively show how neurons which are targets of IT cortex can rapidly, accurately and robustly perform tasks of categorization, identification and read-out of scale and position based on the activity of small neuronal populations.

Figures

Figure 1 Accurate read out of object category and identity from IT population activity

(A) Example of multi-unit spiking responses of 3 independently recorded sites to 5 stimuli. Rasters below each image show spikes in the 200 ms after stimulus onset for 10 repetitions (rows). (B) Performance of a Support Vector Machine (SVM) classifier with linear kernel as a function of the number of sites for reading out category (red) or identity (blue). The classification performance shown on the y-axis indicates the one-versus-all (8 groups for categorization and 77 objects in identification) performance of the classifier on test data (not used for training). The input from each site was the spike count in consecutive 50 ms bins from 100 to 300 ms after stimulus onset. Sequentially recorded sites were combined by assuming independence and concatenating the corresponding response vectors (see the Supplementary Material). Chance levels are 1/8 for categorization and 1/77 for identification (dashed lines). The error bars next to the dashed lines show the range of performances obtained using the 200 ms before stimulus onset (control). Error bars show SD for n=20 random choices of the sites used to train the classifier. (C) Categorization performance (*n*=64 sites) for different data sources used as input to the classifier: multi-unit activity (MUA) as shown in part B, single-unit activity (SUA), local field potentials (LFP), and MUA&LFP (N9).

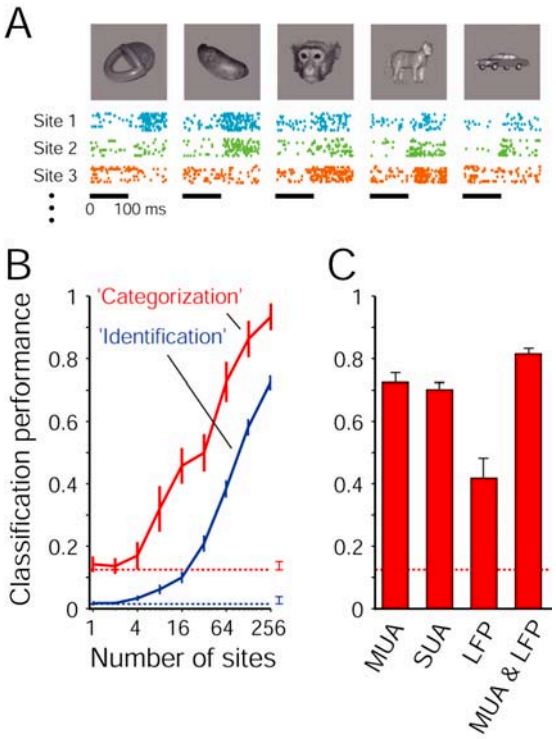


Figure 2 Invariance to scale and position changes

Read-out classifier performance for categorization (**A**) or identification (**B**) when the classifier was trained on the 77 objects at a given scale and position and performance was evaluated with spatially shifted or scaled versions of those pictures. Number of sites used to train the classifier = 64, time interval = 100 to 300 ms after stimulus onset, bin size = 50 ms. The dashed lines indicate chance performance (1/8 in A, 1/77 in B). The error bars to the right of the dashed lines show the performance of the classifier using the 200 ms before stimulus onset (control). Error bars show

SD for n=20 random choices of the sites used to train the classifier. Below each bar, we schematically indicate which stimuli were used to train the classifier ("TRAIN") and which stimuli were used to test its performance ("TEST"). The left-most column show the performance for training and testing on separate repetitions of the objects at the same standard position and scale (as in Figure 1). The second bar shows the performance when training on the standard set (size = 3.4°, center of gaze) and testing on the shifted and scaled images of the 77 objects. Subsequent columns use different image scales and positions for training.

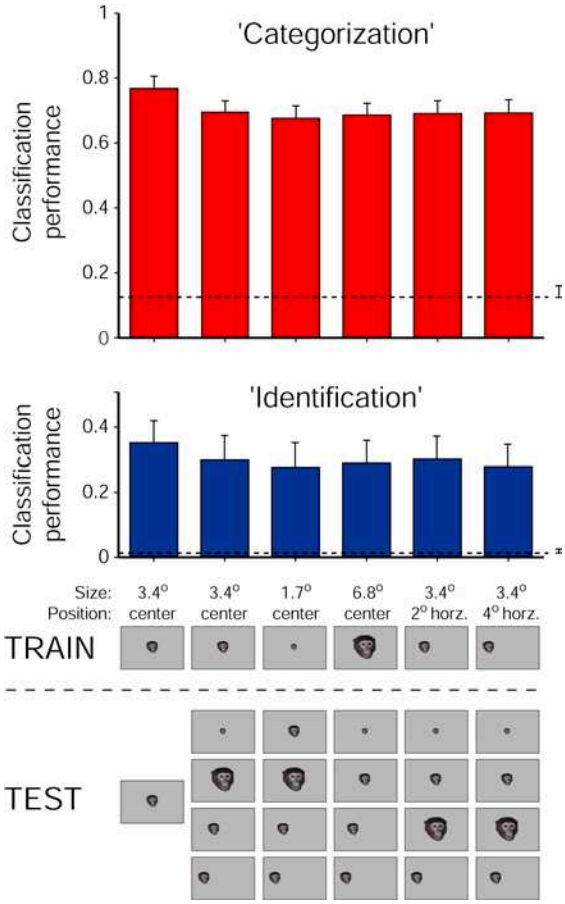


Figure 3 Invariance to novel objects

(**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° size; S1=center of gaze, 1.7° size; S2 = center of gaze, 6.8° size; P1 = 2° shift, 3.4° size; P2 = 4° shift, 3.4° 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. The invariance index was calculated by averaging the Spearman correlation coefficient for the top 10 objects across all possible condition pairs (1 indicates complete invariance and 0 indicates no invariance). Results are based on 13/19 sites selective among the 10 novel stimuli (ANOVA, $p<0.05$).

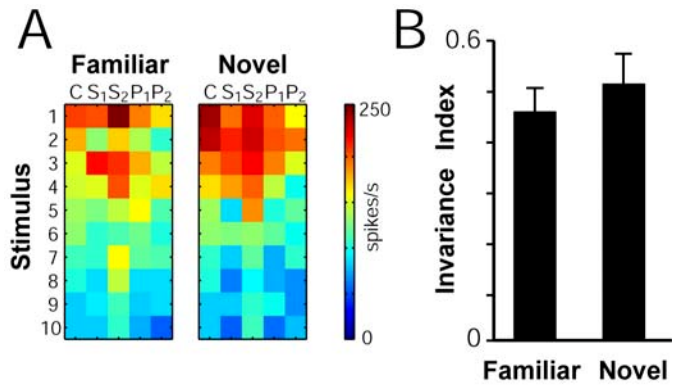
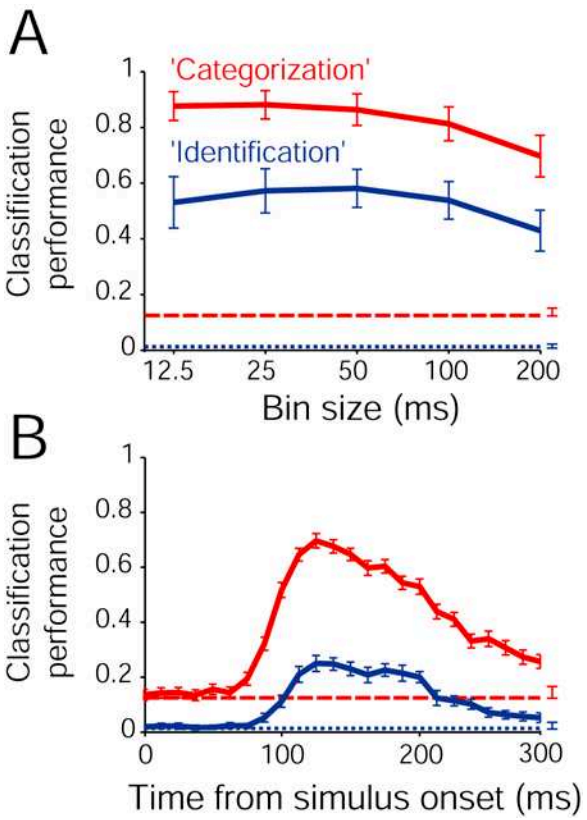


Figure 4 Latency and time resolution of the neural code

(**A**) Classification performance (*n*=128 sites) as a function of the bin size (log scale) used to count spikes across the 100 to 300 ms interval after stimulus onset for categorization (red) and identification (blue). The red and blue dashed lines show change performance for categorization and identification respectively. The error bars next to the dashed lines show the range of performances using the 200 ms before stimulus onset (control).
(**B**) Classification performance using a single bin of 12.5 ms to train and test the classifier. The colors and conventions are as in part (**A**). The time axis indicates the beginning of the bin where spikes are considered (which lasts for 12.5 msec).



Notes

N1. The read out approach used here consists of training a state-of-the-art classifier -- such as a radial basis function (RBF) or regularization classifier (*37, 38*) , for instance a support vector machine (SVM) -- to learn the map from the neural activity recorded from several neurons to the label of the objects presented to the monkey (see Supplementary Material). After training, the classifier can be used to decode the activity of the neurons for novel stimuli and novel recordings. The performance of the classifier constitutes a lower bound on the information content available in the neural activity for the specific task used in training. Unlike recent decoding experiments in the motor system (*39, 40*) our most important goal is not the direct application of this technique for prosthetic applications, which are possible, but the characterization of the information represented by the activity of IT neurons in terms of its underlying neuronal codes. The fact that a certain type of information can be read out from neural activity by the classifier does not necessarily imply that this information is actually used by the brain. Similarly, a coding of neural activity which is optimal for our classifier may not be the code used by the brain. Classifiers, however, can be easily implemented by neurons. In particular, even a simple linear classifier (see Supplementary Figure 14), which may be readily implemented by appropriate synaptic weights to the responses of a single neuron in, say, prefrontal cortex, can effectively read out information from a population of neurons in AIT to perform one of several object recognition tasks. Notice that the last stage of a quantitative model of visual cortex (4) is a RBF-like network capable of generalization across categories and view-points (*41*). In particular, the last layer of the model corresponds to a linear classifier from the RBF-like units in IT to 'output' neurons in PFC.

N2. The 8 groups used in the categorization task were arbitrarily defined before the experiments. However, unsupervised clustering of neuronal similarities yielded a natural grouping of the stimuli into similar classes (Supplementary Figure 1). Classification performance for categorization became significantly worse upon arbitrarily defining these groups as sets of random objects (with the same group sizes, Supplementary Figure 2). Furthermore, examination of the confusion matrix for the identification and categorization performances (Supplementary Figure 3) suggests that some categories were easier to discriminate than others. Stimulus discriminability in IT (and thus classifier performance) depends on similarity (it is harder to separate a face from another face than a face from a car). To further quantify this observation, we evaluated the classification performance in identification within each of the individual groups (Supplementary Figure 4). This showed that individual pictures within the toys or foodstuff groups were easier to discriminate than pictures within the cars group or the monkey faces group. Not surprisingly, the set of white box-like shapes showed the worst within-group classification performance. The classification performance within a group was higher for identification within arbitrarily defined groups. It is important to observe that in general the computational difficulty of any visual classification task depends on image similarity *but also on the recognition architectur*e (for instance on the dictionary of features that may be used by it). Stimuli were not normalized in terms of their contrast or other basic image properties and it is possible to partially read out object category based on some of these simple image properties (Supplementary Figure 13).

N3. The most glaring weakness of the study here is the lack of simultaneous recording from the population of 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 (see however (*42*)). In the approach described here, we have characterized information from a neuronal population assuming independence among the firing of different neurons. This is a strong assumption that will need to be revisited upon recording simultaneously from large numbers of neurons. Our estimate is thus likely to be a *lower bound* on the information represented by small populations of neurons in IT. However, it is interesting to observe that even under these conditions, we obtain such a high degree of accuracy in decoding visual information, suggesting that independent neurons (using a rate code over short time bins) already represent with high precision the information relevant for recognition, without the need for more sophisticated codes based on temporal synchrony among neurons.

N4. Multiple sources of noise can affect the encoding of information in the nervous system. We evaluated the robustness of the representation to two important biological potential noise sources. The performance of the classifier turned out to be very robust to significant deletions of neurons during testing, simulating neuronal or synaptic death (Supplementary Figure 6A) and also to large proportions of deleted spikes (simulating failures in spike transmission or neurotransmitter release, Supplementary Figure 6B). There is a trade-off between wiring specificity and robustness. Using a specific set of neurons for decoding yielded less robustness. The read-out experiments described in the paper correspond to randomly selecting a given number of sites for decoding. In principle, it is conceivable that the brain could be wired in a very specific manner such that neurons which are "looking" at IT activity are specialized in specific discriminations and receive stronger input from the most relevant features. We therefore performed a very simple feature selection step prior to the input to the classifier to select the sites with the highest signal to noise ratio (see Supplementary Material, Methods). This showed that high classification performance levels could be achieved with a much smaller number of pre-selected sites (Supplementary Figure 7). It is conceivable that attentional or feedback based mechanisms could bias the selection of different types of classifiers depending on the task.

N5. We performed a computational 'sanity check' using a quantitative. model of visual cortex (*4*). Performance of S1 units in the model suggests that neurons in early cortical areas such as V1 do not have the invariance properties that we observed whereas C2 units (possibly corresponding to the inputs to IT from V4) show strong scale invariance together with position invariance (M. Kouh).

N6. Here we trained the classifier with 70% of the pictures and then tested on the remaining 30% of the pictures. The test is for categorization, in other words we ask which group the new picture belongs to. The performance is quite good and only slightly below the performance levels reported above (see Supplementary Figure 8, cf. Figure 1).

N7. With 12.5 ms (from 100 to 112.5 after stimulus onset), and 256 neurons it is possible to achieve > 50% categorization performance where chance is 1/8. In this situation, a given cell typically showed zero or one spikes. To directly evaluate whether a few spikes within a short time window could constitute an important element of neural coding (*43*), we compared the classifier performance for bursts of spikes against isolated spikes (Supplementary Figure 12). We observed that bursts of spikes showed significantly better performance than isolated spikes ($p<0.01$). Because bursts largely occurred at the beginning of the response, this suggests that the initial burst conveys most of the information.

N8. A related question, in addition to the MIL, is the practically informative time window (PITW). That is, is there a time point beyond which signals are increasingly redundant with earlier signals and provide only marginal improvement in performance? The results show that read out works well within behaviorally relevant time scales: well before 150 ms after stimulus onset the classifier can already categorize and identify the stimulus with high accuracy, and beyond 150 ms performance gains begin to decrease with additional window length (Figure 4B, see also Supplementary Figure 11). This result has practical biological significance, as it suggests a critical temporal window, between 87.5 and ~150 ms, for ecologically-relevant information from IT to be passed to subsequent stages of processing.

N9. We compared the classification performance for spiking signals and local field potentials (LFPs) obtained by low-pass filtering the raw signal from 1 to 300 Hz (*44*). Both multi-unit activity (MUA) and single unit activity (SUA) obtained by spike sorting of the MUA (*45*) performed better than LFPs for the same number of recording sites (Figure 1C). This supports our previous observations than spikes represent local signals that are more selective than the LFPs which show a broader spatial resolution (*44*). Combining MUA and LFP by concatenating the spike counts and LFP power yielded a slightly better performance than MUA alone (MUA&LFP, Figure 1C). The performance for MUA was comparable to the performance of SUA. This is due to a trade-off between two factors: on the one hand MUA has about 3 times more spikes than SUA and on the other hand SUA shows sharper selectivity (see also Supplementary Figure 5).

N10. **Acknowledgements** We thank Minjoon Kouh for the comparisons against the object recognition model, Rodrigo Quiroga and Alexander Kraskov for spike sorting, Nancy Kanwisher for discussion, the Computation and Systems Biology Initiative at MIT for usage of their computer cluster and Christof Koch and Thomas Serre for comments on the manuscript. This research was sponsored by grants from DARPA, ONR , NIH, and National Science Foundation. Additional support was provided by Eastman Kodak Company, Daimler Chrysler, Honda Research Institute, NEC Fund, Siemens Corporate Research, Toyota, Sony, a Whiteman fellowship (G.K.) and the McDermott chair (T.P.).

References

1. S. Thorpe, D. Fize, C. Marlot, *Nature* **381**, 520 (1996).
2. M. Mace, G. Richard, A. Delorme, M. Fabre-Thorpe, *Neuroreport* **16**, 349 (2005).
3. D. Perrett, J. Hietanen, M. Oeam, P. Benson, *Philosophical Transactions of the Royal Society* **355**, 23 (1992).
4. M. Riesenhuber, T. Poggio, *Nature Neuroscience* **2**, 1019 (1999).
5. N. K. Logothetis, D. L. Sheinberg, *Annual Review of Neuroscience* **19**, 577 (1996).
6. E. Rolls, *Current Opinion in Neurobiology* **1**, 274 (1991).
7. K. Tanaka, *Annual Review of Neuroscience* **19**, 109 (1996).
8. H. Op De Beeck, R. Vogels, *J Comp Neurol* **426**, 505 (Oct 30, 2000).
9. N. K. Logothetis, J. Pauls, T. Poggio, *Current Biology* **5**, 552 (1995).
10. R. Desimone, T. Albright, C. Gross, C. Bruce, *Journal of Neuroscience* **4**, 2051 (1984).
11. P. M. Gochin, M. Colombo, G. A. Dorfman, G. L. Gerstein, C. G. Gross, *J Neurophysiol* **71**, 2325 (1994).
12. B. Richmond, R. Wurtz, T. Sato, *Journal of Neurophysiology* **50**, 1415 (1983).
13. N. Kanwisher, J. McDermott, M. M. Chun, *Journal of Neuroscience* **17**, 4302 (1997).
14. A. Ishai, L. G. Ungerleider, A. Martin, J. L. Schouten, J. V. Haxby, *Proceedings of the National Academy of Science* **96**, 9379 (1999).
15. D. Y. Tsao, W. A. Freiwald, T. A. Knutsen, J. B. Mandeville, R. B. Tootell, *Nat Neurosci* **6**, 989 (Sep, 2003).
16. L. M. Optican, B. J. Richmond, *J Neurophysiol* **57**, 162 (Jan, 1987).
17. H. Tamura, H. Kaneko, I. Fujita, *Neurosci Res* (May 11, 2005).
18. L. G. Ungerleider, J. V. Haxby, *Curr Opin Neurobiol* **4**, 157 (Apr, 1994).
19. M. Missal, R. Vogels, G. A. Orban, *Cereb Cortex* **7**, 758 (Dec, 1997).
20. R. deCharms, A. Zador, *Annual Review of Neuroscience* **23**, 613 (2000).
21. E. T. Rolls, A. Treves, M. J. Tovee, *Exp Brain Res* **114**, 149 (1997).
22. L. F. Abbott, *Quarterly Review of Biophysics* **27**, 291 (1994).
23. A. Tolias, A. Siapas, S. Smirnakis, N. Logothetis, paper presented at the Society for Neuroscience Annual Meeting, Washington 2002.
24. G. Kreiman, *Physics of Life Reviews* **2**, 71 (2004).
25. C. Koch, *Biophysics of Computation*. M. Stryker, Ed., Computational Neuroscience (Oxford University Press, New York, 1999), pp.
26. M. N. Shadlen, W. T. Newsome, *Current Opinion in Neurobiology* **4**, 569 (1994).
27. E. Trucco, A. Verri, *Introductory Techniques for 3D Computer Vision* (Prentice-Hall, New York, 1998), pp.
28. H. Rowley, S. Baluja, T. Kanade, *IEEE PAMI* **20**, 23 (1998).
29. K. Sung, T. Poggio, *IEEE PAMI* **20**, 39 (1998).
30. P. Foldiak, *Neural Computation* **3**, 194 (1991).
31. G. Wallis, E. T. Rolls, *Prog Neurobiol* **51**, 167 (Feb, 1997).
32. B. A. Olshausen, C. H. Anderson, D. C. Van Essen, *J Neurosci* **13**, 4700 (Nov, 1993).
33. K. Fukushima, *Biol Cybern* **36**, 193 (1980).
34. B. Mel, *Neural Computation* **9**, 777 (1997).
35. Y. Sugase, S. Yamane, S. Ueno, K. Kawano, *Nature* **400**, 869 (1999).
36. R. Vogels, *European Journal of Neuroscience* **11**, 1239 (1999).
37. T. Hastie, R. Tibshirani, J. Friedman, *The elements of statistical learning*, Springer Series in Statistics (Springer-Verlag, Basel, 2001), pp.
38. T. Poggio, S. Smale, *Notices of the AMS* **50**, 537 (2003).

39. J. K. Chapin, K. A. Moxon, R. S. Markowitz, M. A. L. Nicolelis, *Nature Neuroscience* **2**, 664 (1999).
40. K. V. Shenoy *et al.*, *Neuroreport* **14**, 591 (Mar 24, 2003).
41. T. Poggio, E. Bizzi, *Nature* **431**, 768 (2004).
42. N. C. Aggelopoulos, L. Franco, E. T. Rolls, *J Neurophysiol* **93**, 1342 (Mar, 2005).
43. R. Krahe, G. Kreiman, F. Gabbiani, C. Koch, W. Metzner, *Journal of Neuroscience* **22**, 2374 (2002).
44. G. Kreiman, C. Hung, T. Poggio, J. DiCarlo, *AI Memo* **2004-020** (2004).
45. R. Quiroga, N. Nadasdy, Y. Ben-Shaul, *Neural Computation* **16**, 16161 (2004).

See also http://ramonycajal.mit.edu/kreiman/resources/ultrafast/

Appendix: Supplementary Material

Methods

Here we summarize the recordings and stimulus presentation and describe the classifiers and data analysis methods in detail. For further detail about the recordings, stimulus presentation and preprocessing, see (*1*).

Recordings and stimulus presentation

Recordings were made from two monkeys (Macaca mulatta) weighing 6.0 kg (monkey K, female) and 5.0 kg (monkey N, male). We used a set of 77 complex grayscale stimuli. Stimuli were arbitrarily divided prior to the experiment into 8 sets: toys, foodstuffs, human faces, hand/body parts (from monkey K), monkey faces (monkey K and animals from adjacent cages), vehicles, white boxes, and synthetic images of cats and dogs. Visual stimuli (3.4 deg) were presented on a monitor were during passive fixation. Each stimulus was on the screen for 100 msec, interleaved by a 100 msec blank matching the background gray (28 Cd/m^2). Stimuli ranged in luminance from 0.5 to 57 Cd/m^2. To preserve the approximate physical appearance of the objects, stimuli were not normalized for mean gray level or contrast. Stimuli were presented in pseudorandom order, randomly presenting one entire set of 77 stimuli before beginning a second randomized set, until each stimulus had been shown 10 times. Recordings were made from both hemispheres of monkey K and the right hemisphere of monkey N. Penetrations were made over a ~10x10 area of the ventral surface (Horsley-Clark AP: 10-20 mm, ML: 14-24 mm). Multi-unit recordings were made using glass-coated Pt/Ir electrodes (0.5-1.5 M Ω at 1 kHz). Spiking activity (400 Hz-6 kHz) and LFPs (1-300 Hz) were amplified, filtered, and stored using conventional equipment. Spike sorting was performed to obtain single unit activity (SUA) from the MUA using the algorithm of Quiroga et al (*2*).

Data analysis

For any given site s ($s=1,\ldots,N$; let ${}_s\boldsymbol{r}_{ij}$ denote the response during repetition i of picture j ($i=1,\ldots,n_{rep}$; n_{rep} = 10 in most cases; $j=1,\ldots,77$). The number of sites was N = 364 for MUA data, 71 for invariance study with MUA data, 315 for LFP data, 45 sites for invariance study with LFP data, 190 for SUA data, 20 for invariance study with SUA data. In order to compare different possible coding schemes, we explored different possible definitions of the response vector $\boldsymbol{r}$. For the spike data, we explored a family of codes based on counting spikes in successive windows of size w starting t_i ms after stimulus onset and ending t_f ms after stimulus onset. 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 Figure 4A). We describe results for different values of these parameters in the text; however, the default condition was w=50 ms, t_i=100 ms and t_f=300 ms. Let $c(t_i,w,b)$ denote the number of spikes in the interval $[t_i+bw;t_i+(b+1)w)$. For the local field potential data, we divide time in a similar fashion but $c(t_i,w,b)$ was defined as the total power in the corresponding time interval. The response $\boldsymbol{r}$ was defined as:

$\boldsymbol{r} = [c(t_i,w,0), c(t_i,w,1), \ldots, c(t_i,w,b)]$ where $b = 0, \ldots, (t_f-t_i)/w-1$

This vector was used as input to the statistical classifier (see below). When considering the responses of multiple sites, we concatenated the corresponding response vectors and used the concatenated vector (herein called $\boldsymbol{R}$) as input to the classifier. It should be noted that this assumes independence among different neurons, an assumption that needs to be revisited upon availability of simultaneous recordings from multiple neurons (*3*). The dimensionality of the input is therefore $(b+1)n$ where n indicates the number of sites. We used n=1, 2, .., 256 sites (except when N<256).

We also separated spikes into spike bursts and isolated spikes (Supplementary Figure 12). A spike burst was defined by at least two spikes with an interspike interval < 20 ms. For the spike bursts, $\boldsymbol{r}$ 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.

The data were always divided into a training set and a test set. In most cases, the training set comprised 70% of the repetitions for each picture while the test set included the remaining 30% set. The training set was randomly chosen from all available repetitions and n_{iter} iterations were performed (n_{iter}=10). For the invariance to scale and position changes (Figure 2), 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. In the case of studying the extrapolation to different pictures within a class, training was performed on all repetitions using 70% of the pictures and testing on all the repetitions of the remaining 30% of the pictures (Supplementary Figure 8).

We focused particularly on two different tasks, classification and identification. For classification, 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, and cats/dogs). Chance was therefore 1/8. For identification, the picture labels directly indicated the identity of the image (77 possible labels). Chance was therefore 1/77.

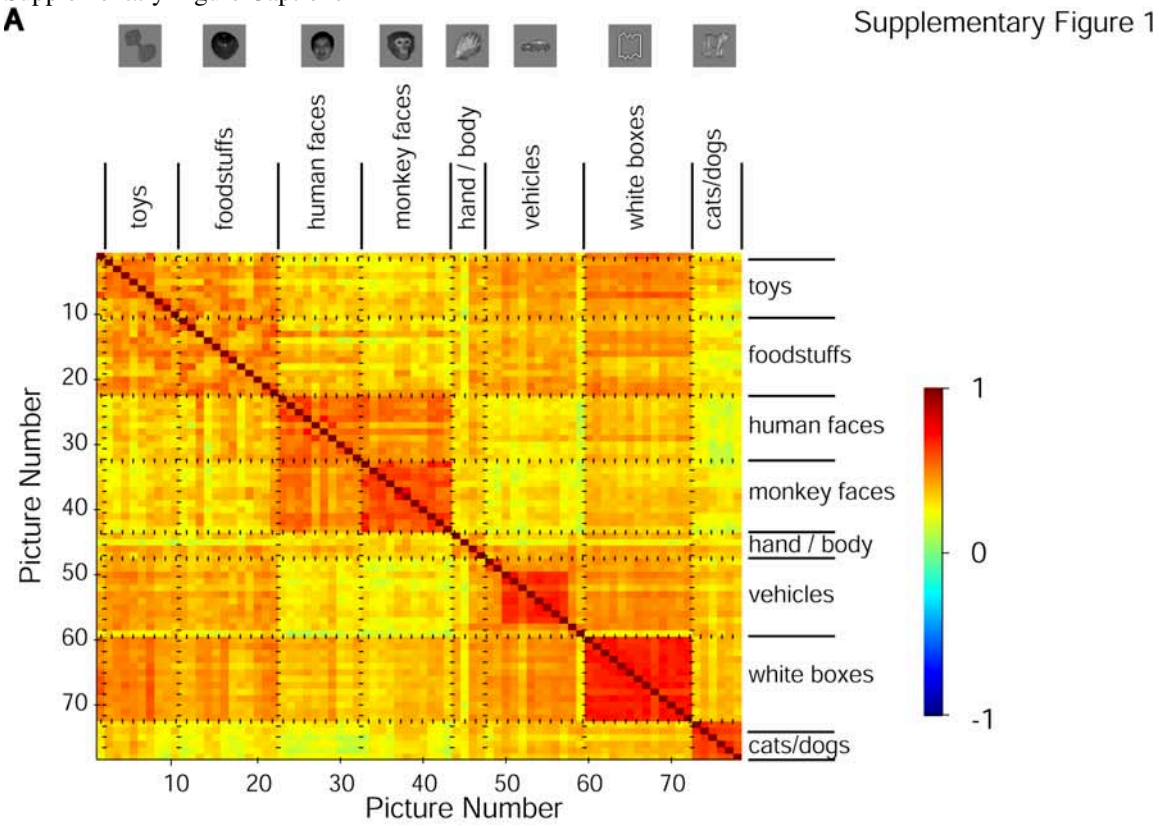
We compared the performance of different statistical classifiers including Fisher linear discriminant classifier (*4*), Support Vector Machine (SVM) using linear or Gaussian kernel (*5*), Regularized least squares classifier (*6*). Supplementary figure 14 compares the performances of these different classifiers. Throughout the text, we use SVM with linear kernel unless stated otherwise. We used the implementation of SVM by Ryan Rifkin (*7*). We initially tested the performance of the classifier on a small sub sample 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; chunk size = 1000; sigma for Gaussian kernel = 16. The regularized least squares classifier is described in (*6*).

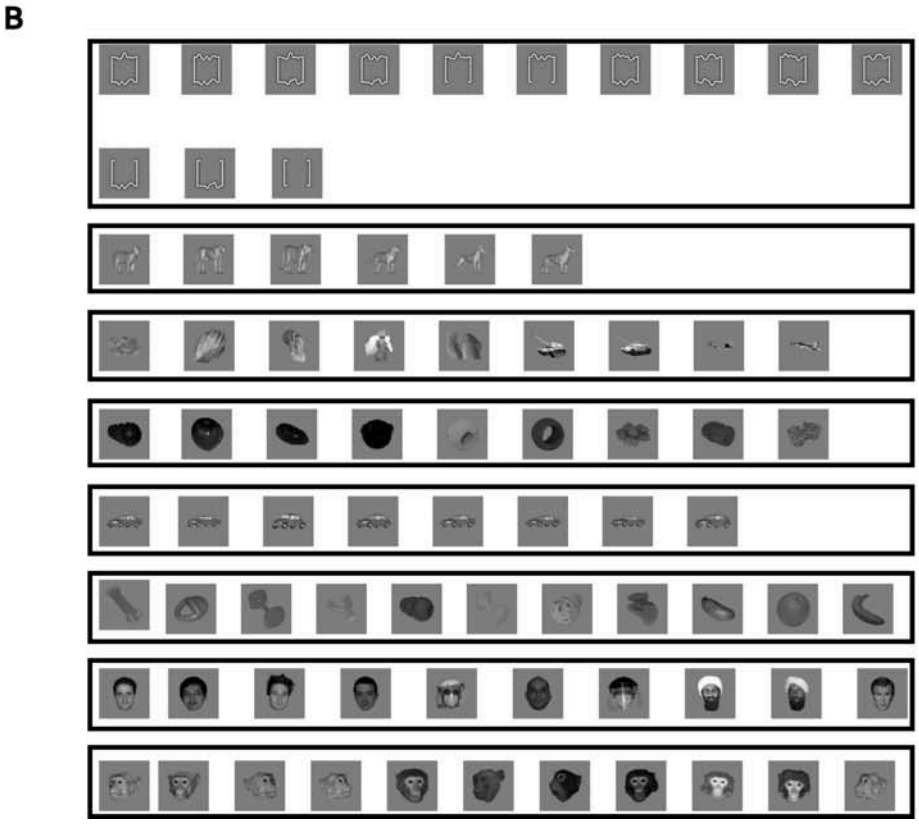
In most of the graphs described throughout the text the n sites used as input to the classifier were randomly chosen from the total set of N sites. This random choice of the sites was repeated at least 20 times (and in most cases 50 times). As a very simple approach to feature selection, we also considered 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 ($s=1,\ldots,N$) and a particular stimulus group g ($g=1,\ldots,G$), ${}_sSNR_g$ 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 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. We also compared the classification results against those obtained by arbitrarily assigning pictures to groups in a random fashion (Supplementary Figure 2 and 4).

We compared the degree of scale and position invariance observed in the neuronal recordings against the responses of units from the standard model of object recognition (*8*). S1 units in the model (set to model the responses of neurons in primary visual cortex) did not show scale and position invariance whereas C2 units showed robust invariance to changes in the scale and position of the units.

We asked whether we could determine whether a picture was presented or not based on the neuronal responses (Supplementary Figure 15). For this purpose, we trained a binary classifier using half of the repetitions and tested 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 $t+\tau+\iota$ 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 ι: 12.5, 25, 50 and 100 ms.

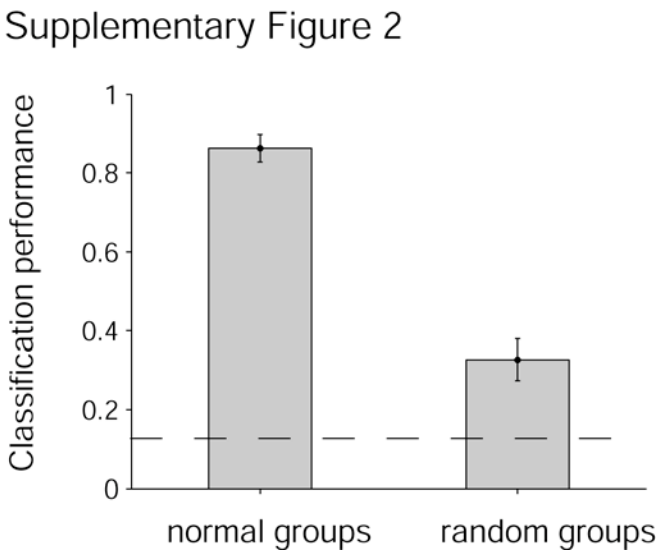
Supplementary Figure Captions





Supplementary Figure 1: Unsupervised clustering of neuronal responses yields categories similar to the pre-defined ones

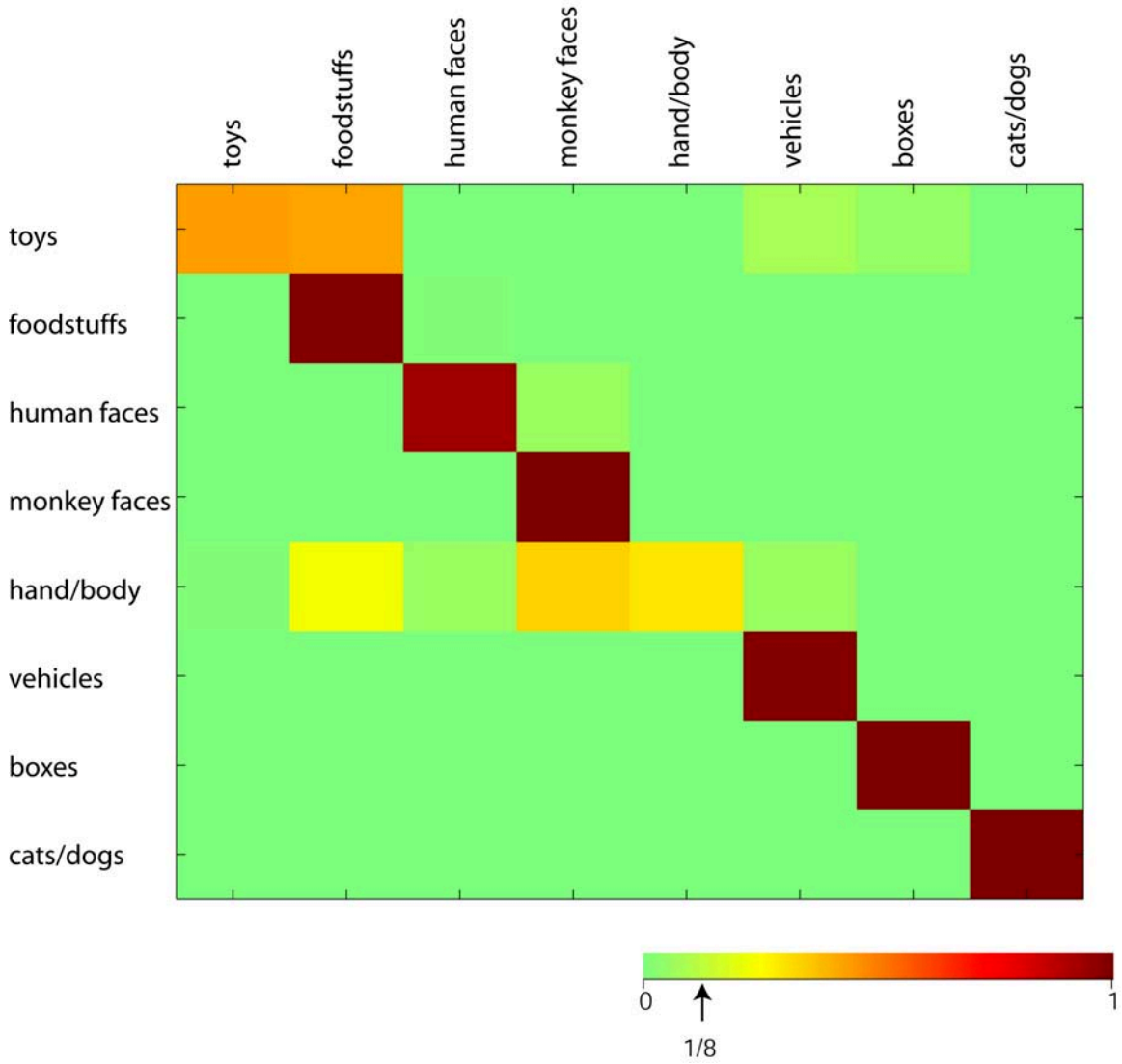
(**A**) We defined the neuronal similarity between two pictures *i* and *j* (*i,j*=1,…,77) based on the spiking population activity as the Pearson correlation coefficient between the vectors containing MUA responses of all sites (*N*=367) to picture *i* and picture *j*. The responses consisted of the spike counts averaged over repetitions in the [100;300) ms interval after stimulus onset. The dimensions of the resulting symmetric similarity matrix were 77 x 77 (the diagonal is trivially 1). The correlation coefficients are color coded (see scale on the right of the matrix). The horizontal and vertical lines divide the different groups of pictures (see Methods above for details). A representative example of each group is shown on top.
(**B**) Results of k-means clustering algorithm (with 10 iterations and random initial conditions) on the neuronal similarity matrix defined in (**A**). The rectangles delimit the pictures belonging to the same cluster. In the results illustrated here, the number of clusters was set to 8 (results for other numbers of clusters are shown on-line).



Supplementary Figure 2: Classifier performance is significantly worse for arbitrary group definition

Comparison of the classifier performance using the 8 groups of pictures defined in the text (see Methods above) versus using 8 arbitrary groups formed by random collections of pictures. The number and size of the random groups was matched to the corresponding values in the normal groups. Error bars show s.d. over 30 iterations using randomly selected sites. The horizontal dashed line shows change performance (1/8). Classifier parameters: MUA, *n*=256 sites, [100;300) ms interval, bin size = 50 ms.

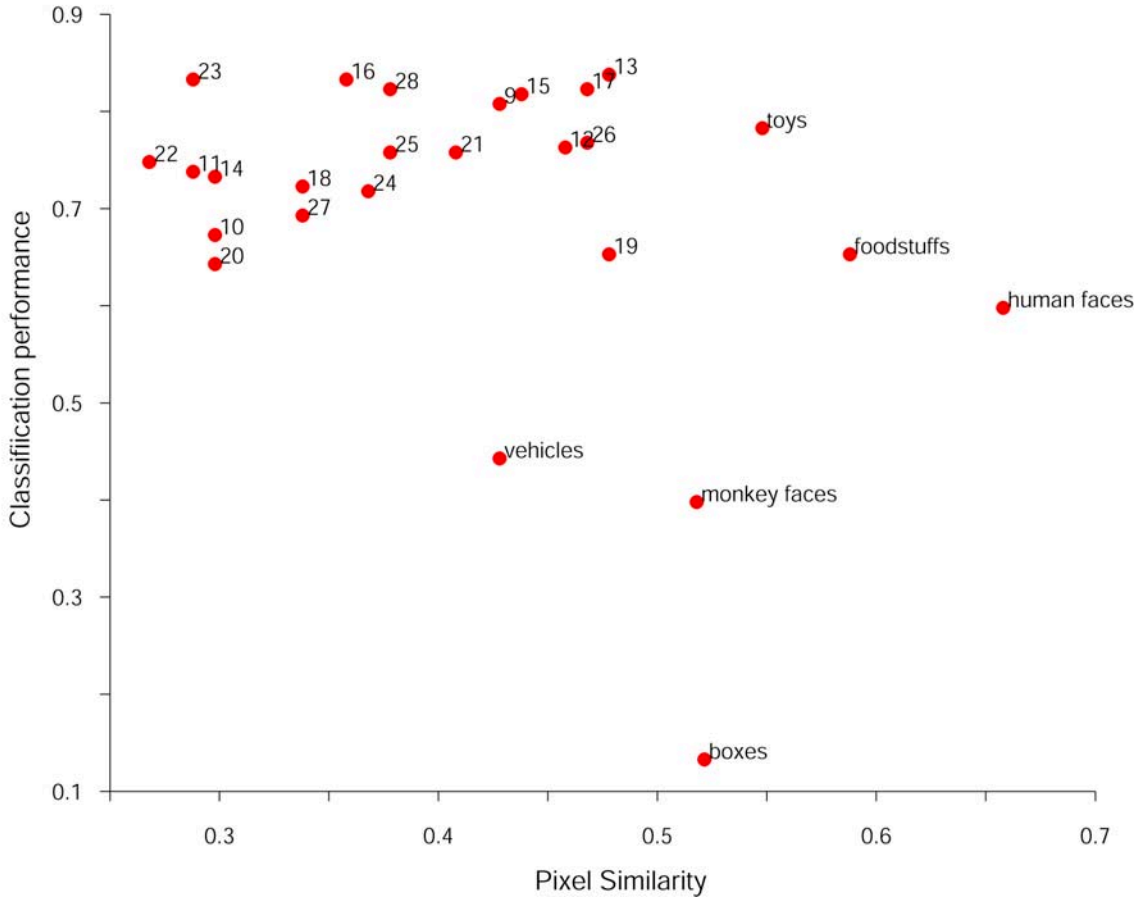
Supplementary Figure 3



Supplementary Figure 3: Pattern of mistakes made by the classifier

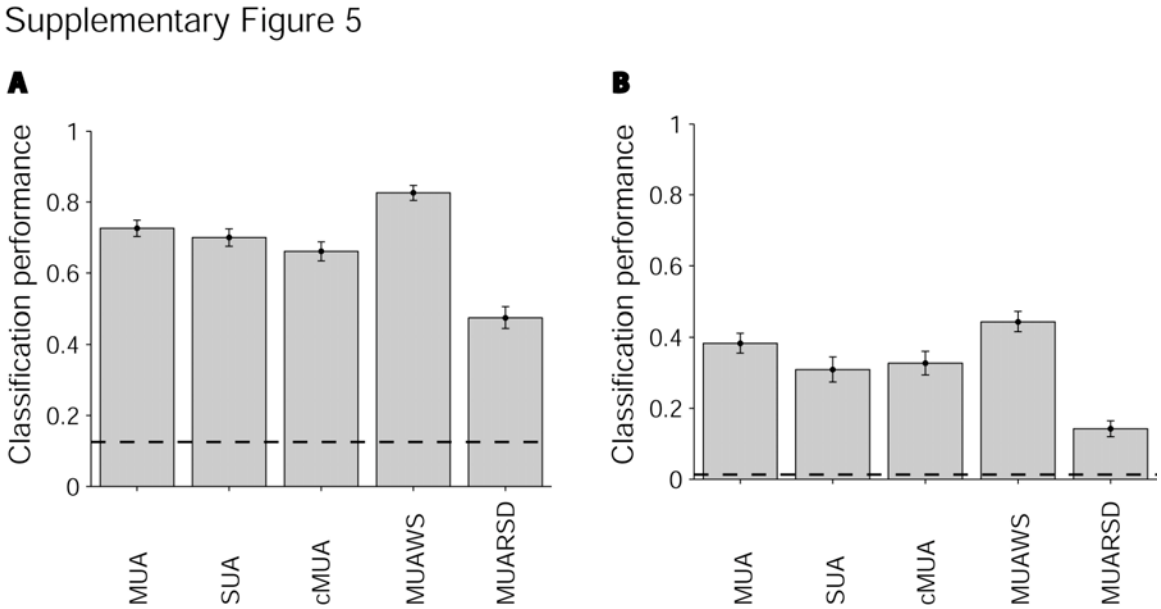
This confusion matrix describes the pattern of mistakes made by the classifier. The color table shows the actual group presented on the monitor to the monkey on the left and the classifier predictions on the top. The table shows (in color code) the proportion of cases where a given category was confused by another category (see color scale below main plot). If all entries were 1 along the diagonal and 0 elsewhere, this would mean that the performance of the classifier is perfect. If all entries showed a uniform color (corresponding to 1/8) then classifier performance would be at chance. Note that the performance of the classifier is better for some groups (e.g. monkey faces) than for others (e.g. toys). Classifier parameters: MUA, *n*=256 sites, [100;300) ms interval, bin size=50 ms.

Supplementary Figure 4



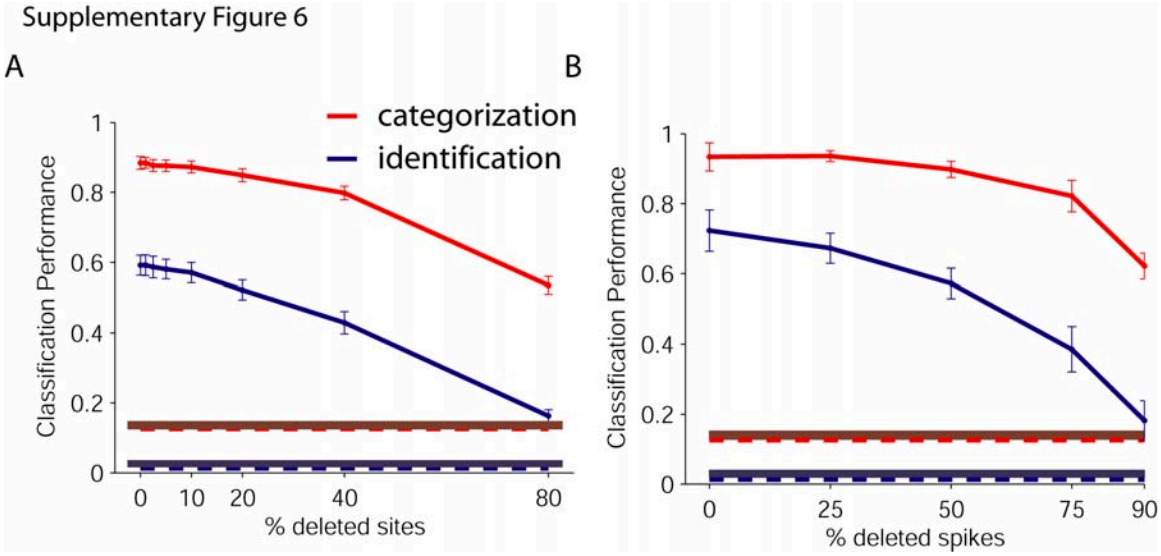
Supplementary Figure 4: Read-out performance as a function of the pixel similarity among pictures within a given group

Pixel similarity between two images was defined by shifting a square box of 30x30 pixels in steps of 10 pixels, computing the correlation coefficient between the two image patches and then reporting the median correlation for the top 10 patches (several other parameters were explored and this is the set of parameters that yielded the strongest agreement with the pre-defined image categories). The x-axis shows the pixel similarity of a group of pictures defined as the average of all pair comparisons. The y axis shows the classifier performance in discriminating the different pictures within that group. The figure includes the pre-defined groups (see Supplementary Figure 1A) and also 20 arbitrary groups composed of 10 random pictures each.



Supplementary Figure 5: Spike sorting and classifier performance

Read-out classifier performance for classification (**A**) and identification (**B**) for different types of spike signals. MUA: multi-unit activity (signal is high-pass filtered with a corner frequency of 400 Hz and a threshold is imposed). SUA: single unit data (spike sorting by Alexander Kraskov and Rodrigo Quiroga). cMUA: MUA, multi-unit activity after spike sorting (for several sites, the spike sorting algorithm returned a cluster that was considered to be a valid neuronal signal but not a single unit; this is here labeled cMUA). MUAWS: MUA for sites that contain at least one discernable single unit in the spike sorting analysis. MUARSD: MUA with random spike deletion so that the average spike count matches the average spike count for SUA. Classifier parameters: n = 64 sites, time interval = [100,300) ms, bin size = 50 ms.



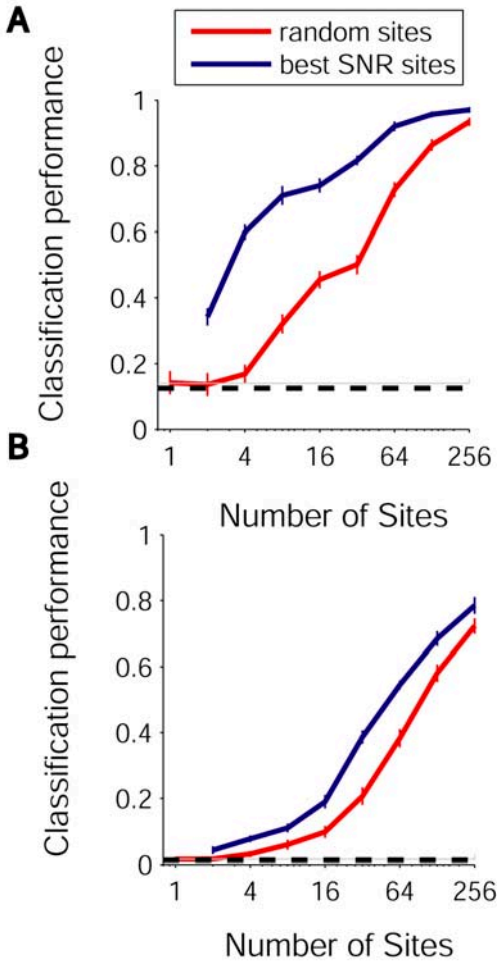
Supplementary Figure 6: The neural code is very robust to neuronal drop-out

A. Training of the classifier 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). Classifier parameters: MUA, n=256 sites, [100;300) ms time interval, bin size = 50 ms.

B. Here, a proportion of spikes (indicated in the x-axis) were removed from the classifier (simulating the process of failures in spike transmission or neurotransmitter release). Classifier parameters: MUA, n=256

sites, [100;300) ms time interval, bin size = 50 ms. Red=categorization, blue=identification. Dashed lines = chance performance. Red and blue rectangles show performance in the -100 to 0 ms interval.

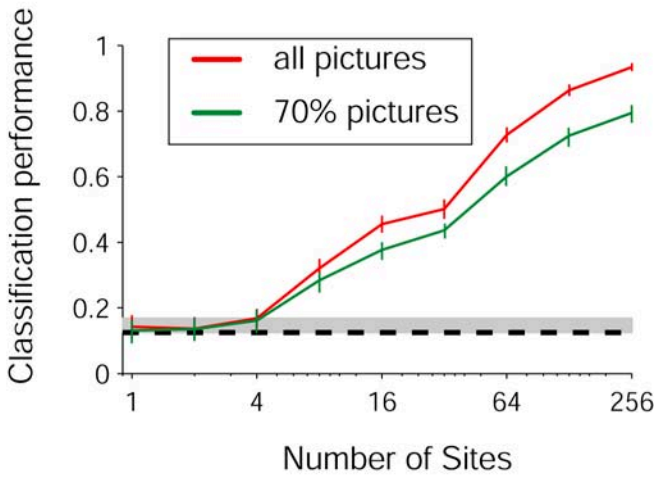
Supplementary Figure 7



Supplementary Figure 7: Specific wiring significantly improves classifier performance

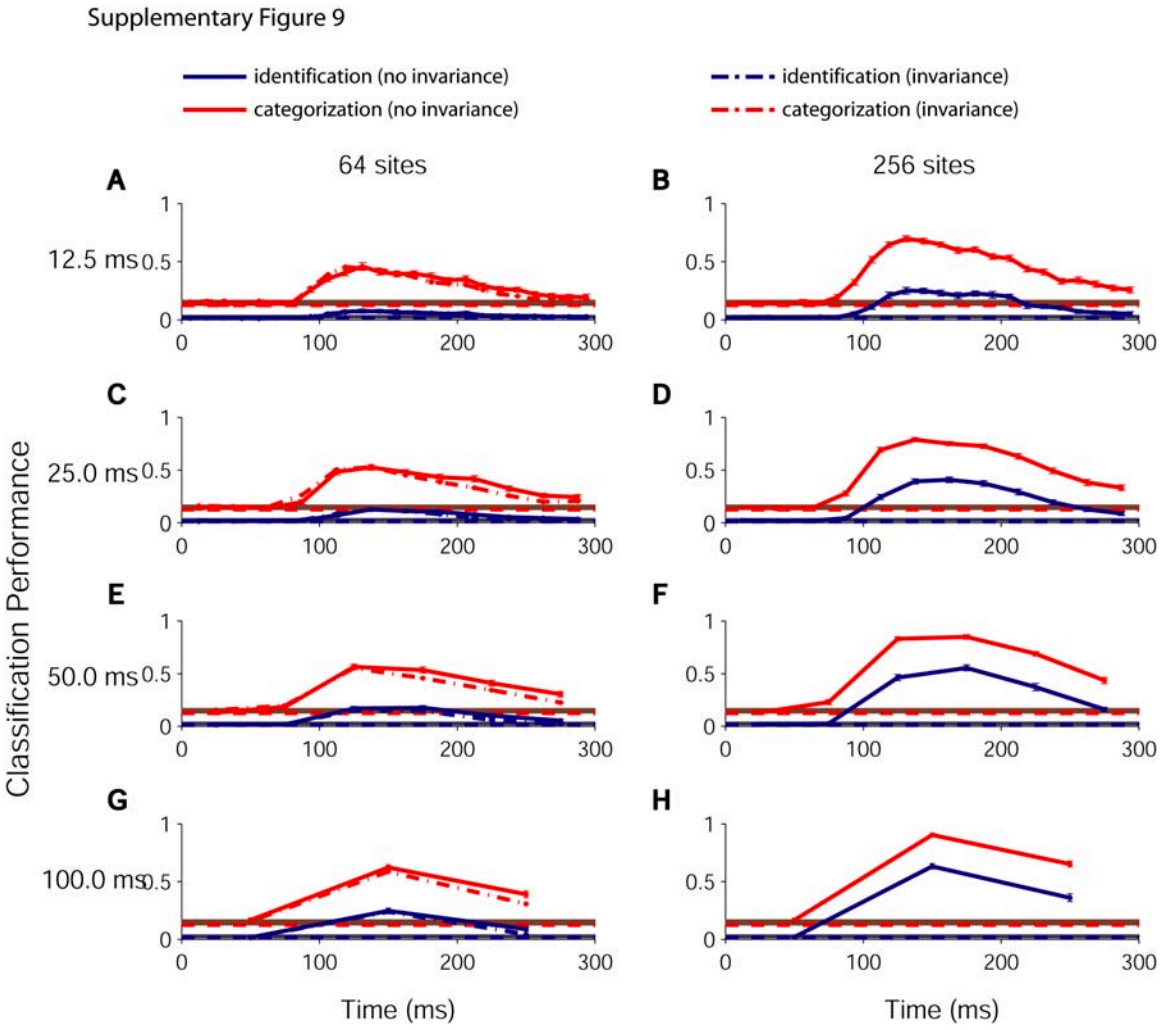
Specific wiring significantly improves performance. Read-out performance as a function of the number of sites for categorization (**A**) or identification (**B**) using randomly selected sites (red) or pre-selected sites (blue). For a given site and a given group, the SNR was defined as the difference 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 read-out task. Sites were ranked based on the SNR and the sites with the highest SNR were selected in the "best SNR sites" case. Classifier parameters: MUA, [100;300) ms interval, bin size = 50 ms.

Supplementary Figure 8



Supplementary Figure 8; Extrapolation to novel pictures within the same categories

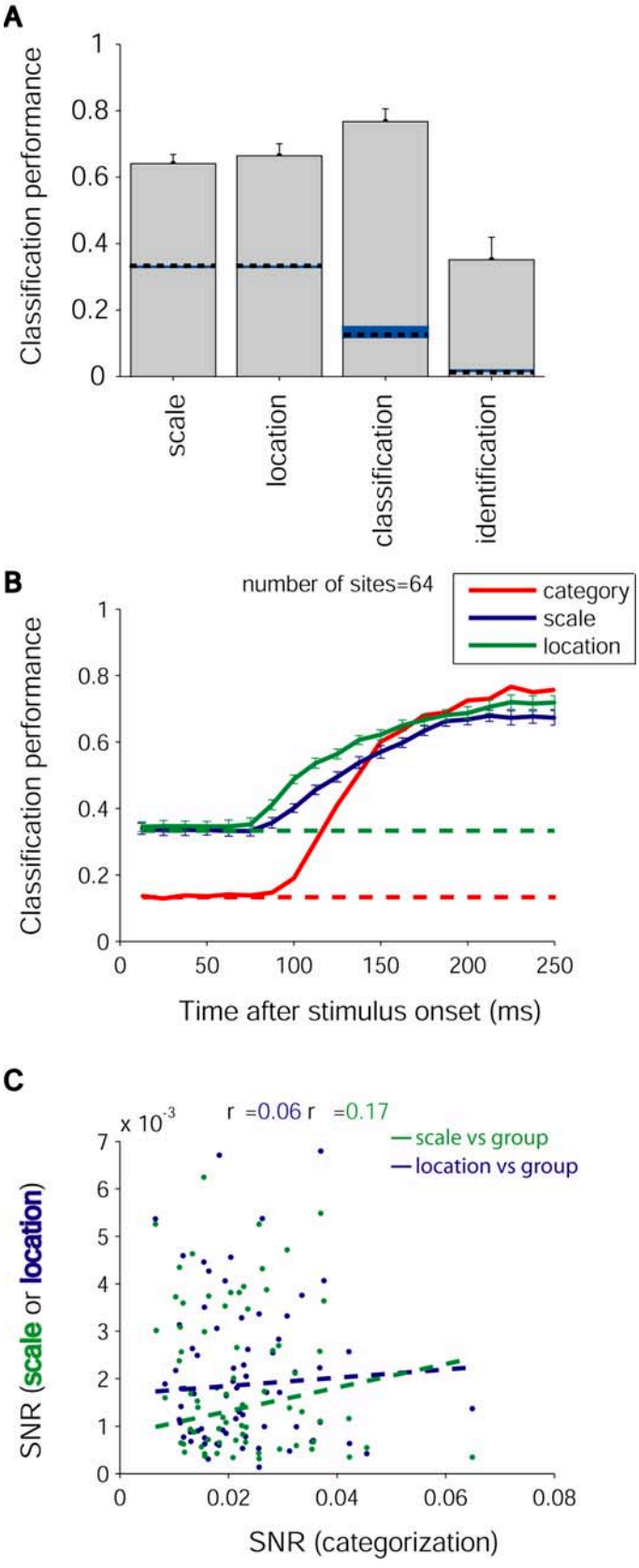
In this figure, the classifier was trained with all the repetitions for 70% of the pictures and testing for classification performance was done with all repetitions for the remaining 30% of the pictures (green). Thus, the classifier never saw the neuronal responses to the test pictures during the training 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 Figure 1). Classifier parameters: [100;300) ms interval, bin size = 50 ms.



Supplementary Figure 9: Latency and time resolution

Classification performance for categorization (red) and identification (blue) as a function of time using a single bin of 12.5 ms (**A-B**), 25 ms (**C-D**), 50 ms (**E-F**) and 100 ms (**G-H**) to train and test the classifier. The bins are centered at the time point indicated on the x axis. The format is the same as in Figure 4B. In A, C, E, G, we also show the performance for the invariance condition (dash-dot lines) and we used 64 sites. In B, D, F, H we used 256 sites. The horizontal dashed lines show the chance levels. The horizontal rectangles show the range of performances using the 100 ms before stimulus presentation.

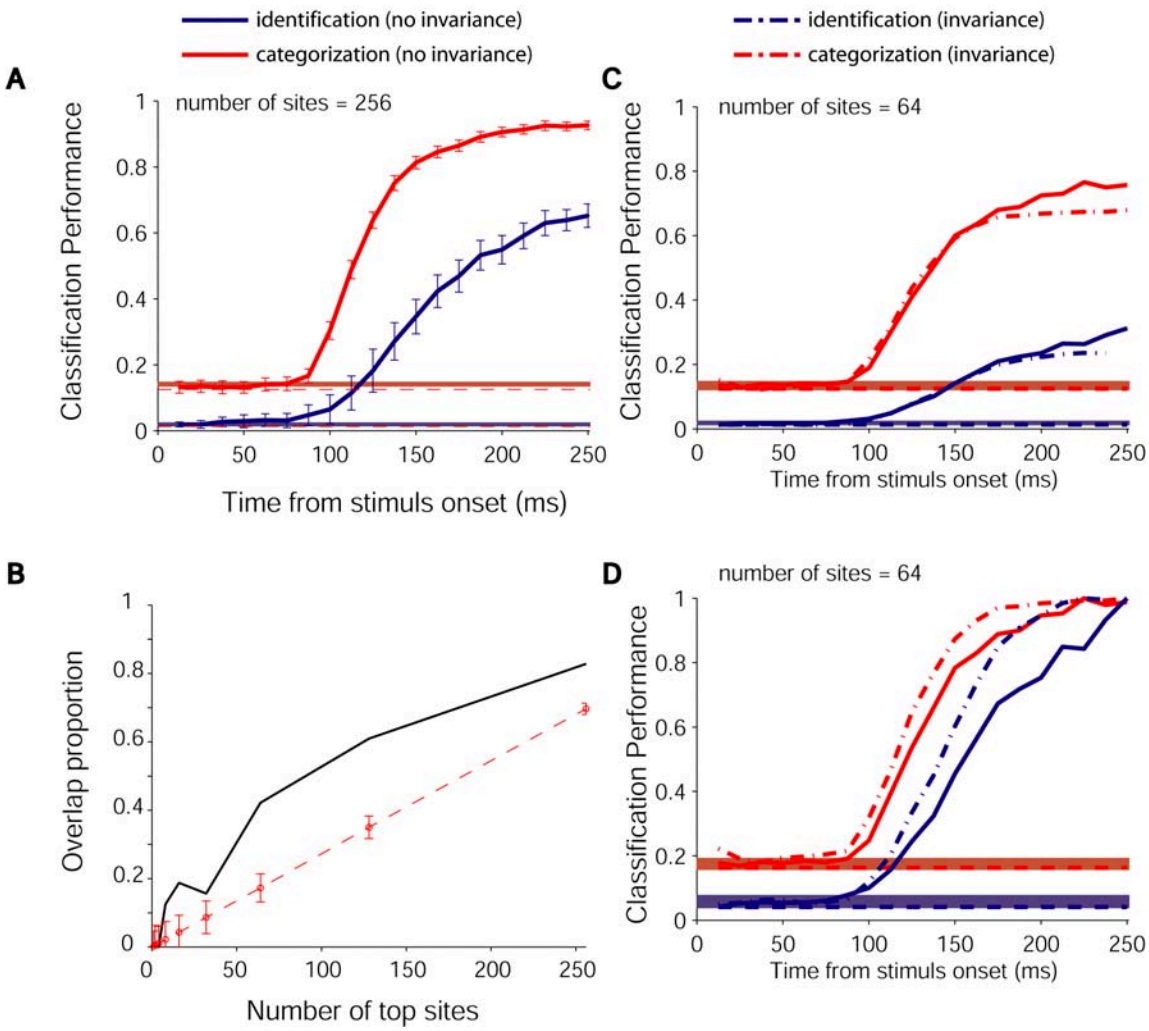
Supplementary Figure 10



Supplementary Figure 10: We can also read out image scale and position

A Read-out performance for scale, location, classification and identification. To read out 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 image (regardless of the image identity). Then the performance was evaluated by asking what the scale or position was for the remaining 30% of the repetitions. The dashed line indicates chance performance (which is different for each task). Chance was 1/3 for read-out of scale and position because there were 3 possible scales and positions. 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 location (green) as a function of the cumulative time from stimulus onset in bin sizes of 12.5 ms. The classifier is trained on all images as in part A (using 70% of the repetitions), the labels indicate the stimulus scale (3 possible values), location (3 possible values) or stimulus category (8 possible values). 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, location, 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 is a poor correlation between scale read-out and category read-out SNR values and between position read-out and category read-out SNR values. In contrast, there was a stronger correlation between the SNR values for identification and categorization (r=0.54, see Figure 11B).

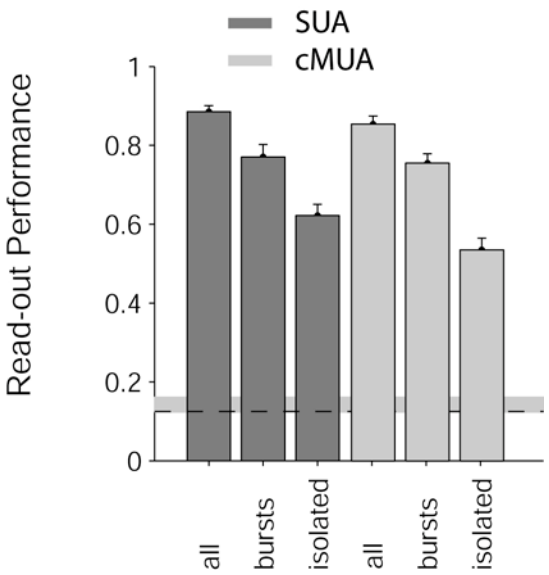
Supplementary Figure 11



Supplementary Figure 11: Latency of categorization versus identification

A. Read-out classifier performance as a function of cumulative time from stimulus onset for classification (red) and identification (blue). Classifier parameters: MUA, bin size = 12.5 ms, n = 256 sites. The dashed line shows the chance performance levels (1/8 for classification and 1/77 for identification). The red and blue rectangles near the dashed lines show the performance for the 100 ms time interval before stimulus onset. Error bars = s.d. for 20 random choices of sites used to train the classifier.
B. Sites were ranked according to the SNR for classification or identification (see Methods for definition of SNR). Here we indicate the proportion of overlapping sites among the best sites (x-axis) for classification vs. identification (black line). The red dashed line indicates the overlap proportion expected by chance.
C. Classification performance as a function of cumulative time from stimulus onset (as in part A) showing invariance to scale and position (dash-dotted lines). Classifier parameters: MUA, bin size=12.5 ms, n=64 sites.
D. Same data shown in C normalized by the maximum performance in each case.

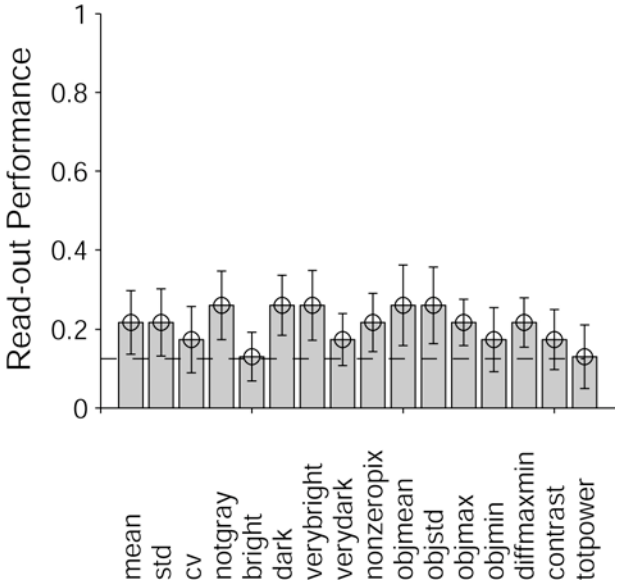
Supplementary Figure 12



Supplementary Figure 12: Spike bursts perform better than isolated spikes

Comparison of read-out performance for bursts versus isolated spikes. A burst event was defined by at least 2 spikes with an interspike interval of < 10 ms. All spikes were thus labeled as either burst spikes or isolated spikes. The plot below shows the read-out performance for SUA (dark gray) or cMUA (light gray, see above for definition of cMUA) for all spikes, spike bursts or isolated spikes. For spike bursts, we still counted the number of spikes in each bin.

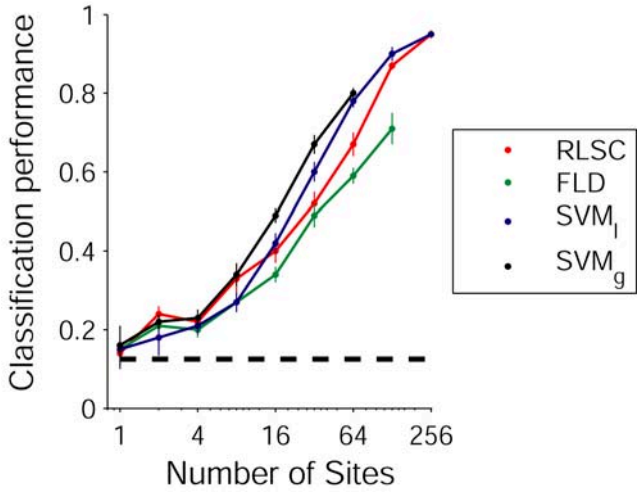
Supplementary Figure 13



Supplementary Figure 13: Classifier performance based on basic image properties

Performance in reading out object class from each of 16 different basic image properties: mean=mean pixel intensity, std = s.d. of pixel intensities, cv = s.d./mean, notgray = proportion of pixels different from background, bright = proportion of bright pixels, dark = proportion of dark pixels, very bright = proportion of very bright pixels, verydark = proportion of very dark pixels, objmean = mean pixel intensity within object, objstd = s.d. of pixel intensity within object, objmin,objmax = minimum object intensity, maximum object intensity, diffmaxmin = objmax-objmin, contrast = image contrast, totpower = total image power. Error bars denote 1 s.d. from 100 iterations. The dotted line indicates chance level.

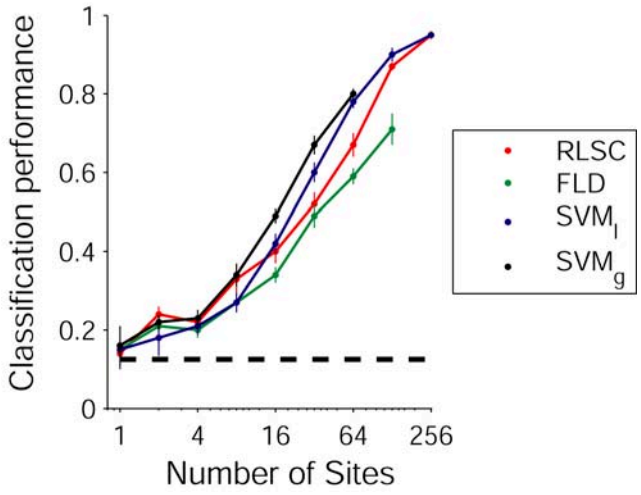
Supplementary Figure 14



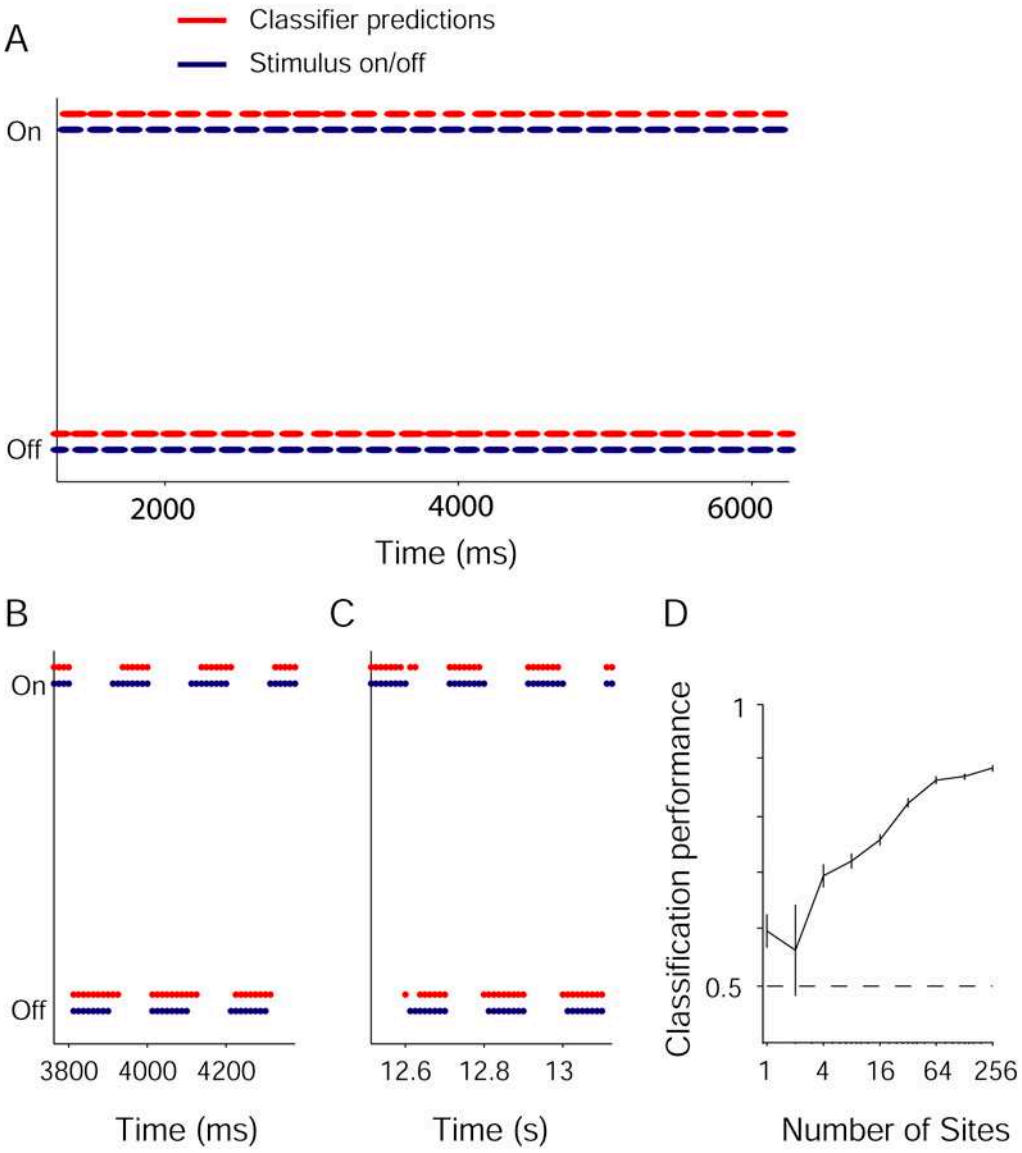
Supplementary Figure 14: Comparison among different classifiers

Comparison of the classification performance results using different statistical classifiers: regularized least squares classifier ('RLSC'), Fisher linear discriminant ('FLD'), support vector machine with linear kernel ('SVM_l'), support vector machine with gaussian kernel ('SVM_g'). Classifier parameters: MUA, bin size = 50 ms, time interval = 100 to 300 ms, SVMFu C=10, SVMFu sigma=16, normalizer=1.

Supplementary Figure 14



Supplementary Figure 15



Supplementary Figure 15: Read-out of picture presence

A classifier was trained to indicate whether a picture was presented or not at each time point (bin size = 12.5 ms, see Methods above). (A) Actual picture presentation scheme (blue) and classifier predictions (red). (B, C) Zoom-in showing in more detail the classifier predictions versus the actual presentation scheme. (D) Classification performance as a function of the number of sites used to train the classifier.

Further references

1. G. Kreiman, C. Hung, T. Poggio, J. DiCarlo, *AI Memo* **2004-020** (2004).
2. R. Quiroga, N. Nadasdy, Y. Ben-Shaul, *Neural Computation* **16**, 16161 (2004).
3. N. C. Aggelopoulos, L. Franco, E. T. Rolls, *J Neurophysiol* **93**, 1342 (Mar, 2005).
4. C. M. Bishop, *Neural Networks for Pattern Recognition* (Clarendon Press, Oxford, 1995), pp.
5. T. Poggio, S. Smale, *Notices of the AMS* **50**, 537 (2003).
6. 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.
7. R. Rifkin. (MIT, Cambridge, 2000).
8. M. Riesenhuber, T. Poggio, *Nature Neuroscience* **2**, 1019 (1999).

See also http://ramonycajal.mit.edu/kreiman/resources/ultrafast/