Behavioral demand modulates object category representation in the inferior temporal cortex

Running title: Task dependent category representation in IT

Nazli Emadi1,2,3, Hossein Esteky1,2

1: School of Cognitive Sciences, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran,
2: Research Center for Brain and Cognition, School of Medicine, University of Shahid Beheshti, Tehran, Iran.
3: Howard Hughes Medical Institute and Department of Neurobiology, Stanford University School of Medicine, Stanford, California, USA.

To whom correspondence should be addressed:
E-mail: nazlie@stanford.edu (N.E.); esteky@ipm.ir (H.E.)

Keywords: object category; behavioral demand; visual perception; inferior temporal cortex
Visual object categorization is a critical task in our everyday life. Many studies have explored category representation in the inferior temporal (IT) cortex at the level of single neurons and population. However, it is not clear how behavioral demands modulate this category representation. Here we recorded from the IT single neurons in monkeys performing two different tasks with identical visual stimuli: passive fixation and body/object categorization. We found that category selectivity of the IT neurons was improved in the categorization compared to the passive task where reward was not contingent on image category. The category improvement was the result of larger rate enhancement for the preferred category and smaller response variability for both preferred and non-preferred categories. These specific modulations in the responses of IT category neurons enhanced signal to noise ratio of the neural responses to better discriminate between the preferred and non-preferred categories. Our results provide new insight into the adaptable category representation in the IT cortex which depends on behavioral demands.
INTRODUCTION

Visual object categorization organizes the infinite number of visual stimuli into meaningful groups; therefore, it is crucial to understanding the visual world and interacting with it. Understanding the cortical representation of the visual object categories is one of the most fundamental questions in object recognition studies. The inferior temporal (IT) cortex lies at the final stage of the ventral visual pathway and is the first visual area that contain neurons with selective category responses. (Desimone et al. 1984; Logothetis and Sheinberg 1996; Perrett et al. 1982; Tanaka 1996). It is heavily connected with the prefrontal cortex, memory related cortices and association areas (Axmacher et al. 2008; Miyashita 1993; Webster et al. 1994). These properties of the IT cortex have made it a suitable area where detailed bottom-up visual information could be transformed into more abstract category-related information (Logothetis and Sheinberg 1996). The IT cortex contains neurons with selective responses to visual object categories such as faces and bodies (Desimone et al. 1984; Kiani et al. 2007; Kobatake and Tanaka 1994; Logothetis and Sheinberg 1996; Perrett et al. 1982; Tanaka 1996; Tsao et al. 2006). Object categories are also represented in the activity of the population of IT neurons (Hung et al. 2005; Kiani et al. 2007; Meyers et al. 2008). Functional MRI (fMRI) studies of the inferior temporal lobe of monkeys and humans have shown cortical patches responsive to a specific category, such as faces or bodies (Downing et al. 2001; Kanwisher et al. 1997; Pinsk et al. 2005; Rajimehr et al. 2009; Tsao et al. 2003). We have previously shown that the activity of the category selective neural clusters in the IT is causally linked to the perceptual categorization performance (Afraz et al. 2006).

The position of the IT cortex at the top of the visual hierarchy and its connections with areas involved in cognitive functions such as the prefrontal cortex (Webster et al. 1994), has made it a good candidate to study the effects of top-down non-visual cues on the sensory visual processing. The strength of top-down signals enhances along the ventral visual pathway towards the IT cortex. It has been shown that attentional enhancement of firing rate gets larger and occurs earlier along the ventral stream; from V1 toward V2 and then V4 (Buffalo et al. 2010). Task demand is one of the non-visual cues that could potentially modulate the IT neural responses. It is shown that in the IT cortex, the same stimulus elicits different responses depending on the context and the task demands (Anderson and Sheinberg 2008; Fuster and Jervey 1981; Vogels et
al. 1995). Richmond and Sato showed that the IT neural responses enhance during the visual
discrimination compared to the stimulus attention tasks (Richmond and Sato 1987). Sato
reported a suppression in the IT neural responses in pattern discrimination compared to passive
viewing task (Sato 1988). Koida and Komatsu found that in the color categorization and
discrimination tasks IT neural activity increased and decreased, respectively, compared to the
passive viewing task (Koida and Komatsu 2007). In contrast, Suzuki et al. found no difference in
the IT neural responses at different stimulus discrimination levels (Suzuki et al. 2006).

These studies have explored the task related modulation of the amplitude of the IT neural
response to the visual stimuli. However, the effect of the task demands on the object category
selectivity of the IT neurons has not been explored yet. Selectivity of the IT neurons to complex
visual objects is a specific property of the IT neurons in the visual hierarchy and underlies our
visual categorization ability. By a micro-stimulation study we have previously shown that
activity of the category selective neural clusters in the IT is causally linked to the perceptual
categorization performance (Afraz et al. 2006). The effect of micro-stimulation is correlated with
the level of category selectivity of the neurons. These results suggest that category selective
neurons, such as face and body neurons are linked to the visual categorization performance.
Given the importance of our understanding about the neural mechanisms of visual object
categorization we designed the current experiment to study how behavioral needs modulate the
category representation in the IT cortex. The monkeys in the current study performed two tasks
with the same visual stimuli but different behavioral demands: passive fixation and
categorization. By analyzing the spiking activity of the IT single neurons recorded while
monkeys were performing these two tasks, we explored how category selectivity of IT neurons
was modulated by the behavioral demands.

METHODS

Subjects
Two male adult macaque monkeys were used in this experiment. Before training, the monkeys
were prepared with head restraints and recording chambers implanted stereotaxically on the
dorsal surface of their skull under aseptic conditions while monkeys were anesthetized with
sodium pentobarbital. All experimental procedures were in accordance with the National
Institutes of Health guide for the care and use of laboratory animals. They were also approved by
the animal care and use committee of Institute for Research in Fundamental Sciences (IPM).

**Stimuli**

The stimuli were $7^\circ \times 7^\circ$ in size grayscale photographs of body (human, monkey and four-
leg subcategories) and object categories (aircraft, car and chair subcategories). There were 90
images in each category (30 images per subcategory) (Emadi and Esteky 2013). Body images
had no facial features. Each image was presented in four different noise levels. Each noise level
was generated by assigning a uniformly distributed grayscale value to X% of image pixels,
where X was the absolute noise level and had one of the values of 10, 30, 45 or 60. These 720
noisy images [(2 categories) x (90 images in each category) x (4 noise levels)] and 90 full noise
images (100% noise) were randomly presented to the monkeys without repetition. The pattern of
the noise was fixed for one image across days, but different among stimuli. The stimuli were
presented on a 19 inch CRT monitor (resolution: 800 x 600 pixels), placed 57 cm in front of the
monkey seated in a primate chair.

**Tasks**

Passive fixation task: monkeys were trained to perform a passive viewing task (Fig. 1A). Following 400 ms of fixation on a white fixation point at the center of the screen, a randomly
selected sequence of images was presented to the monkey. Each image was presented for 70 ms
with a variable inter-stimulus interval (850 to 950 ms). The monkey was rewarded with a drop of
apple juice every 1.5–2 seconds as long as its gaze was fixated within a $2.4^\circ \times 2.4^\circ$ invisible
fixation window at the center of the screen. The sequence of images stopped when the monkey
broke the gaze fixation and the fixation point reappeared after 1500 ms of blank interval. Stimuli
in the broken trials were later randomly selected for the re-presentation.

Two-alternative forced choice body/object categorization task: we trained two monkeys
for ‘body/object’ categorization task (Fig. 1B). The monkeys started a trial by fixating on a
fixation point within a $2.4^\circ \times 2.4^\circ$ window at the center of the screen. The fixation point duration
was randomly chosen from three variable durations (350, 400 or 450 ms). This pre-stimulus
phase was followed by a brief (70 ms) presentation of an image chosen pseudo-randomly from
the set of noisy body and object images. After 500-ms delay a green and red response target were
presented $10^\circ$ to the left and right of the screen center, respectively. The monkey was required to
make a saccadic eye movement to one of the targets to indicate whether the image was a body or an object. Monkeys were trained to make the saccade no later than 300 ms after the target onset and keep their gaze fixed within 2.4° x 2.4° window on the saccade point for 150 ms. The assignment of body and object choices to the left and right targets was reversed between the two monkeys. Each correct response (i.e. when the image was correctly categorized by monkey) was rewarded by a drop of apple juice. For ‘full-noise’ stimuli (0% visual signal), the monkey was rewarded randomly in 50% of the trials. Correct and wrong trials were followed by 750 and 1500 ms inter-trial interval, respectively.

To examine the unlikely event of whether the monkeys simply memorized all the images in the image set, a behavioral experiment was conducted after the completion of the training for each monkey. In these experiments, 40 novel images (20 bodies and 20 objects) were intermixed with 40 familiar body and object images (randomly chosen from the learned images). The stimuli were presented to the monkeys without any visual noise. The body/object categorization performance of the monkeys was measured in several behavioral sessions. In each session, we used a new set of novel images. The monkeys’ performance for novel stimuli was not different from familiar stimuli from the very beginning of the behavioral sessions (Monkey 1: familiar stimuli: 90.00% ± 2.75, novel stimuli: 83.33% ± 3.42, t-test, p = 0.13; Monkey 2: familiar stimuli: 92.50% ± 2.41, novel stimuli: 86.67% ± 3.12, t-test, p = 0.14).

In each recording session monkeys performed both passive categorization tasks. At the beginning of each recording session 810 noisy stimuli in the image set were randomly divided into 9 blocks of 90 images separately for each task. Monkeys were presented with an interleaved order of passive and categorization blocks starting randomly with either of them in each recording session (Fig. 1C). Only recording sessions in which the monkeys successfully completed at least half of the trials in both tasks were included in the analysis. Median number of trials per category in both tasks was 360 (mean±sem; passive: bodies = 342 ± 10, objects = 340 ± 9.5; categorization: bodies = 338 ± 10.3, objects = 336 ± 9.9). The eye position was measured by an infra-red eye-tracking system.

**Recording**

In a preparatory surgery, head restraints and recording chambers were stereotaxically implanted under aseptic conditions on the dorsal surface of the skull of the monkeys while the animals were
anesthetized with sodium pentobarbital. Craniotomy was performed to record from the inferior temporal cortex of the monkeys. After recovery, single units were recorded with tungsten electrodes (FHC). The position of the recording chamber and the recording sites were defined by the stereotactic measurements and magnetic resonance images (MRIs) acquired prior to surgery. Subdivisions of the IT cortex were defined using the location of cortical sulci as illustrated by Tanaka and colleagues (Saleem et al. 2000; Saleem and Tanaka 1996; Tamura and Tanaka 2001). Recordings were made over the lower bank of superior temporal sulcus (STS) and TE cortices (12 to 18 mm and 13 to 20 mm anterior to interauricular line in monkey1 and monkey2, respectively). During each recording session, a single tungsten electrode was inserted into the IT cortex. The electrode was advanced with an Evarts-type manipulator (Narishige, Japan) from the dorsal surface of the brain through a stainless steel guide tube inserted into the brain down to 10-15 mm above the recording sites. After isolating units in real-time by a template matching algorithm (Worgotter et al. 1986) while inserting the electrode into IT cortex, the monkeys were required to perform a short passive fixation task. In this task 10 randomly selected noiseless images from each subcategory were presented. Each image was presented once for 70 ms with 630 ms blank interval in between. The PSTHs of the units were plotted real-time and visually inspected. If the unit was visually responsive, monkeys were presented with the interleaved categorization and passive tasks. Spiking activity of the isolated single neurons was recorded extracellularly while monkeys were performing the tasks. A total of 95 visually responsive single neurons were recorded from two monkeys (74 from monkey1 and 21 from monkey2). Visual responsiveness was defined as significantly larger responses (in any of the sliding 100-ms windows from 100 to 400 ms after the stimulus onset) relative to the baseline activity (-50 to 50 ms relative to the stimulus onset) following the presentation of body or object images in at least one of the noise levels (t-test, alpha = 0.01).

Data Analysis

- Based on the similar trend of monkeys’ behavior and a similar pattern in other results, data from two monkeys were combined in all of the analyses.
- In the categorization task we only analyzed the correctly completed trials.
• Trials with different levels of noise (10, 30, 45 and 60 percent) are used in all of the analyses unless otherwise mentioned. Because of the lack of category information in the fully noisy images (100% noise), this condition was excluded from the analyses.

• In all of the analyses a 300-ms window from 100 to 400 ms after the stimulus onset was used.

• PSTHs were smoothed by convolving with a 30-ms Gaussian kernel.

• Before running any parametric tests, the normality of the distributions was confirmed by Kolmogorov–Smirnov test ($P < 0.01$ in all of the distributions).

• All of the statistical tests were one-tailed one-sample t-test unless otherwise mentioned.

• P-values less that 0.00001 are shown as $p < 10^{-5}$.

Analysis of the response modulation Index (RMI)

RMI was used as a normalized index to examine the rate modulation in categorization task compared to the passive task:

\[
RMI = \frac{\mu(C) - \mu(P)}{\mu(C) + \mu(P)} \times 100
\]

\(\mu(C)\) and \(\mu(P)\) were the averaged evoked response of each neuron in categorization (only correct trials) and passive tasks. Positive and negative values represent response enhancement and suppression in categorization compared to passive, respectively.

Analysis of the response variability (Fano Factor)

Fano Factor (FF) was used as the index of response variability:

\[
FF = \frac{\sigma^2}{\mu}
\]

\(\sigma^2\) and \(\mu\) were the variance and mean of the spike count, respectively. The FF was measured in 100-ms windows with 1-ms steps for the responses evoked by each category. The number of spikes was calculated in each window in each trial. Then the FF within each window was computed as the ratio of variance in spike counts to mean spike count across all trials. For rate matching between two tasks in each neuron, the most common firing rate value, across all
windows in both tasks, was selected as the matching rate. The FFs of all windows with firing rates ± 10% of the selected rate were averaged in each task.

**D-prime**

D-prime (d-prime) was used as an index of the neural category selectivity for bodies versus objects:

\[
Dprime = \frac{\mu(B) - \mu(O)}{\sqrt{\sigma^2(B) + \sigma^2(O)/2}}
\]

\(\mu(B)\) and \(\mu(O)\) were the mean response of each neuron to body and object images, respectively. \(\sigma^2(B)\) and \(\sigma^2(O)\) were the variance of the distributions of neural responses to body and object images, respectively. Positive and negative values represent body and object category selectively, respectively. We measured d-prime in a 300-ms window from 100 to 400 ms after the stimulus onset. For plotting d-prime in time, d-prime was measured in sliding 100-ms windows with 1-ms steps.

**Classification**

A linear ‘support vector machine’ was used to assess the neural performance in body vs. object categorization. The neural performance was measured by using the “Support Vector Machines” functions from the Statistics toolbox of MATLAB software. We used sequential minimal optimization method to find the separating hyperplane. The input for each neuron was the response of the neuron to body and object categories in a 300-ms window from 100 to 400 ms after stimulus onset. In each round of classification, we randomly selected the 75% of trials of every neuron in each task for training the classifier. The classification performance of the neural population was tested on the remaining 25% of the trials. In all of neurons equal numbers of trials were used in each task and image category. Trials were pseudorandomly selected from the pool of trials in each condition. This procedure was repeated 100 times for each step of increasing number of neurons in Figure 2G. Percent of task-dependent categorization improvement in each cell group was measured as:

\[
PerformanceModulation(\%) = \frac{CategorizationPerformance - PassivePerformance}{PassivePerformance} \times 100
\]
RESULTS

We trained two monkeys to perform two different tasks: passive fixation (called “passive”) and two-alternative forced choice body/object categorization (called “categorization”). In these two tasks the same images were presented to the monkeys in two different behavioral contexts. In the passive task monkeys passively viewed the images to get a reward (Fig. 1A), while in the categorization task they were required to correctly categorize the images as bodies or objects to get a reward (Fig. 1B). Images were selected from different body and object subcategories and were presented to the monkeys with different levels of added noise (Fig. 1C-D). Blocks of these two tasks were presented to the monkeys in an interleaved order (Fig. 1E). Fig. 1F shows that monkeys’ performance in the categorization task decreased as noise level increased. As expected, monkeys had a better performance categorizing less noisy stimuli. Performance for fully noisy images was 48.48±0.60, which was different from chance (t-test, p=0.016), indicating a small bias (1.52%) in this noise level toward object choice.

Response Selectivity in Categorization vs. Passive Viewing

To study the behavior-related modulation of the IT category responses, we recorded spiking activity of 95 single IT neurons in monkeys performing the passive and categorization tasks. We measured d-prime which is an indication of the signal to noise ratio of the neural responses (see Methods) and captures the separation between the distributions of responses to body and object images in each neuron. Larger d-prime values mean less overlap between these two distributions, or higher signal to noise ratio for category selectivity. By measuring d-prime we found 51 body and 44 object neurons using the responses during the categorization task (see Methods). Absolute d-prime values were higher in body compared to object neurons (two-sampled two-tailed t-test, p = 0.0168, body neurons: n = 51, object neurons: n = 44).

To study the neural signature of the behavioral demands we plotted the d-prime over time in two different tasks for exemplar body (C52) and object (C83) neurons, measured in 100-ms sliding windows (Figures 2A-B). We observed that category selectivity was improved during the evoked response in the categorization task compared to the passive task, making each group of neurons more selective to their preferred categories (1000 bootstrap replicates, t-test, 100 to 400 ms, p < 10^-5 in both neurons). We studied these changes at the population level and observed
similar results for body and object neurons (Figures C-D, t-test, 100 to 400 ms, p < 10^{-5} in body neurons (n = 51), p = 0.000011 in object neurons (n = 44)).

To better explore this modulation, we measured the difference between the d-prime values in categorization and passive tasks in all of the single neurons. Figure 2E shows a positive correlation between the d-prime modulation and d-prime in single neurons (monkey 1: r = 0.71, p = 0.0003, n = 74; monkey 2: r = 0.72, p < 10^{-5}, n = 21). In neurons with positive d-prime (body selective), d-prime improvement in categorization compared to passive task was 0.0488 ± 0.0102 (mean±sem). We observed significant d-prime increase in 38 out of 51 (74.5%) body neurons (t-test, alpha = 0.05, 1000 bootstraps). In neurons with negative d-prime (object selective), d-prime modulation was -0.0552 ± 0.0105. We found significant d-prime decrease in 33 out of 44 (75%) of object neurons (t-test, alpha = 0.05, 1000 bootstraps). These findings indicate that selectivity of 74.5% and 75% of body and object neurons to their preferred categories were improved in the categorization task. Improvement of category selectivity of the single neurons during the categorization task is consistent with the monkeys’ behavioral demand for a higher categorization performance to obtain more reward.

To explore the effect of task difficulty, we measured the d-prime modulation in each noise level separately. The modulation was 0.0223 ± 0.0171 in 10%, 0.0444 ± 0.0150 in 30%, 0.0476 ± 0.0179 in 45% and 0.0625 ± 0.0153 in 60% noise levels. Figure 2F shows a non-significant trend for larger d-prime modulation in more ambiguous conditions (Pearson correlation, r = 0.0881, p = 0.0872, n = 95).

Category Information at the Population Level in Categorization vs. Passive Viewing

To see how category representation at the population level was modulated between these two tasks, a linear classifier was trained on the population of neurons to discriminate body and object images in each task. Figure 2G shows the performance of the classifier as a function of the number of body and object neurons for categorizing images into bodies and objects in each task. Increasing the number of neurons enhanced the categorization performance from chance level (50%) to around 70% for body neurons in categorization task. This performance was close to the monkeys’ averaged behavioral performance (72.9%). Body neurons showed higher categorization performance compared to object neurons in both tasks (44 neurons of each group, t-test, p < 0.00001 in both tasks, 100 bootstraps). Performance in both groups of neurons was
larger in categorization compared to passive task (significancy line in Figure 2G, t-test, alpha = 0.05, 100 bootstraps). To compare this task-dependent categorization improvement between body and object neurons, we measured the percent of performance modulation for both cell groups (44 neurons of each group, see also Methods). These measurements revealed that performance modulation in body neurons (11.67% ± 2.77) was significantly larger than object neurons (6.24% ± 2.89; t-test, p = 0.021, 100 bootstraps, n= 44).

**Firing Rate in Categorization vs. Passive Viewing**

By definition d-prime takes both mean and variability of the responses into account. D-prime improvement in categorization task can be the result of signal increase (rate enhancement) and/or noise decrease (response variability decrease).

To explore the modulation of the firing rate we plotted the response of the exemplar body neuron to the preferred and non-preferred categories in passive (left panel) and categorization (right panel) tasks (Figures 3A). We observed that the response difference of this neuron to the preferred versus non-preferred category was enhanced during the categorization task (1000 bootstraps, t-test, p < 10^{-5}). Figure 3B shows similar results for the exemplar object neuron (1000 bootstraps, t-test, p < 10^{-5}). We also plotted the averaged response of the population of body (Figure 3C) and object (Figure 3D) neurons to their preferred and non-preferred categories in different tasks and observed similar results at the population level.

To quantify this observation we measured a response modulation index (RMI) in each neuron for different image categories (see Methods). Positive RMI values were indicative of response enhancement in categorization task compared to the passive task. We measured RMI values of the body neurons for body images (4.97% ± 1.09) and object images (1.93% ± 1.11). RMI in the population of body neurons for both preferred and non-preferred categories was significantly larger than zero (t-test, preferred category: p = 0.000016, non-preferred category: p = 0.043, n = 51). RMI of the population of object neurons for both body (3.92% ± 1.18) and object images (7.72% ± 1.40) were also significantly larger than zero (t-test, preferred category: p = 0.000001, non-preferred category: p = 0.00095, n = 44).

We also compared the RMI values for the preferred and non-preferred categories. RMI difference was measured as RMI for preferred category minus RMI for non-preferred category. The results showed that RMI values were larger for preferred than non-preferred categories in...
both cell groups at the population level (Figures 3E; body neurons: ΔRMI = 3.05% ± 0.77; t-test; monkey 1 = 0.0009, n = 74; monkey 2 = 0.0295, n = 41, both monkeys: p = 0.00011, n = 95; object neurons: ΔRMI = 3.80% ± 0.84; monkey 1 = 0.0003, n = 74; monkey 2 = 0.0188, n = 21; both monkeys: p = 0.000024, n = 95). These findings indicate that during the categorization with higher level of task engagement compared to the passive task, responses to both image categories were enhanced. However, this response enhancement was specific in a way that larger response enhancement occurred for the preferred category, resulting in the category selectivity improvement at the level of single neurons. This effect was not significantly different between the two groups of body and object neurons (t-test, p = 0.74, body neurons: n = 51, object neurons: n = 44). At the level of single neurons, 38 body (74.51%) and 36 object (81.82%) neurons showed significantly larger rate modulation for their preferred than non-preferred category (t-test, alpha = 0.05, 1000 bootstraps).

Response Variability in Categorization vs. Passive Viewing

The variability of the neural response to different members of each category is an important factor in measuring the d-prime or how well a neuron could discriminate between two categories. We plotted the Fano Factor (FF) of the exemplar body (C52, Figure 4A) and object (C83, Figure 4B) neurons across time, measured in 100-ms sliding windows (see Methods). We observed a decrease in FF after the stimulus presentation in both neurons which was stronger in the categorization compared to the passive task. We also plotted the FF in time for the population of neurons and found lower FF in the categorization compared to the passive task later during the evoked response (Figure 4C, significance line: t-test, alpha = 0.05). To better explore this effect, we measured the rate matched FF in each single neuron during the evoked responses (see Methods). Figure 4D shows FF of the single neurons in the categorization versus passive task. 35 body (63.6%) and 25 object (56.8%) neurons showed significantly lower FF during the categorization task (t-test, alpha = 0.05, 1000 bootstraps). The modulation of the FF was 0.1044 ± 0.0297 in body and 0.0460 ± 0.0366 in object neurons. The decrease of FF in the categorization compared to the passive task was a small but significant effect in the population (t-test; monkey 1: p = 0.0275, n= 74; monkey 2: p = 0.0086, n = 21; both monkeys: p = 0.0013, n= 95). This effect was not significantly different between body and object neurons (two-tailed two-sampled t-test, p = 0.15, body neurons: n = 51, object neurons: n = 44). We also compared the FF modulation (categorization FF minus passive FF) for body and object images separately.
and found no difference in the population for FF modulation in these two image groups (two-tailed one-sampled t-test, \( p = 0.14 \), \( n = 95 \)). These finding suggests that the variability of the response to different members of the preferred and also non-preferred categories was decreased during the categorization compared to the passive task.

The results of RMI and FF analysis suggest that the larger d-prime values in the categorization task are related to the larger rate enhancement for the preferred category besides smaller response variability for both preferred and non-preferred categories in the categorization task. Figure 4E summarizes our findings and shows how category representation improves in the categorization task. Larger rate enhancement for the preferred category and the smaller response variability for both preferred and non-preferred categories result in a greater separation of the signal and noise distributions. This enhances the category discriminability of the neural response during the categorization task which is consistent with the monkeys’ behavioral demand.

DISCUSSION

By comparing the activity of IT neurons in monkeys performing passive fixation and body/object categorization tasks, we found that behavioral demand modulates category representation in the IT cortex. Category selectivity was improved in the categorization task both at the level of the single neurons and the population. The category representation improvement resulted from two different mechanisms. First, the rate enhancement observed in the categorization task was larger for the preferred category. Second, the response variability to both preferred and non-preferred categories decreased in categorization compared to the passive task.

The effect of the task demand on the response amplitude of the IT neurons has been explored previously. The results show increase, decrease or no modulations in the responses of IT neurons depending on the task in hand (Anderson and Sheinberg 2008; Fuster and Jervey 1981; Koida and Komatsu 2007; Richmond and Sato 1987; Sato 1988; Suzuki et al. 2006; Vogels et al. 1995). We observed that during the categorization task with higher levels of task engagement compared to the passive task, firing rates were enhanced for both preferred and non-preferred categories. More importantly, we found that this rate enhancement was larger for the preferred category compared to the non-preferred category. This specific rate modulation results in category selectivity enhancement at the level of single neurons. It is possible that the general
increase in the responses to the preferred and non-preferred categories is related to the enhanced alertness in the categorization task while the selective larger response enhancement to the preferred category is related to the task specific visual attention.

Few studies have explored the effect of the top-down signals on the reliability of the neural responses in the ventral visual pathway. It was previously shown that response reliability of the V1 and V4 neurons is not affected by top-down signals (McAdams and Maunsell 1999; Wiener et al. 2001). However, more recent studies have reported that the response reliability of the V4 neurons improves by attention related top-down signals (Cohen and Maunsell 2009; Mitchell et al. 2007; Noudoost and Moore 2011). We found that the IT response reliability increased in the categorization compared to the passive task, which is consistent with the more recent reports on the effect of top-down signals on the response reliability of V4 neurons. However, our result extends those earlier findings in three aspects. First, we report how response reliability changes based on the behavioral demands in two different tasks compared to the attention modulation in the same task reported in previous studies (Cohen and Maunsell 2009; Mitchell et al. 2007). Second, we measured response reliability modulation in the IT cortex. Third, we measured the variability of the responses of IT neurons to a richer stimulus set where the response to image exemplars of a category were only presented once. All of the previous studies have explored the variability of the responses to the repeated presentations of one single stimulus. More reliable response to different members of each category decreases the overlap between the response distributions to those categories and enhances the category discriminability.

We found that the amplitude and the variability of the responses were modulated in a direction that led to higher discriminability power. There was a larger rate enhancement for the preferred category with similar reliability enhancement for both categories. Larger distance between the mean of the responses to the two categories (larger between- categories response difference) and more reliable response to the members of each category (smaller within-category response difference) both enhance the discrimination between the two categories and improve the category selectivity of the neural responses. It has been shown that engagement in a spatial attention task decorrelates the ongoing fluctuations of the neural population activity in macaque area V4 (Mitchell et al. 2009). This decorrelation reduces the response variability caused by the
task-irrelevant intrinsic cortical activity and thereby enhances the signal-to-noise ratio. Our results showing a similar decline in the variability of the responses to the preferred and non-preferred categories could be related to a similar mechanism.

It is important to note that this effect is different from the change in the response selectivity of the IT neurons as a result of learning and training (De Baene et al. 2008; Freedman et al. 2006; Kobatake et al. 1998; Sigala 2004; Sigala and Logothetis 2002; Vogels 1999). Here monkeys switched between two tasks several times in a session and this flexible category representation was modulated from one pattern to the other consistent with the changes in the behavioral demands in a session.

The origin of the modulation in the category representation could be a “hierarchical” top-down feedback about the task rule from downstream areas like the prefrontal cortex (Muhammad et al. 2006). This modulation could be the result of changes in the synaptic weight and functional connectivity within IT and also between IT and upstream/downstream areas. In our experiment monkeys were well-trained for discriminating bodies from other objects. It has been shown that there is a specialized neural system for the representation of bodies as one of the most important visual stimuli that we perceive (Downing et al. 2001). Learning a task is a top-down guided process (Ahissar and Hochstein 2004). Based on the “reverse hierarchy theory” in a highly trained subject, the ability to perform a task could progress backwards to the lower levels with a better signal-to-noise ratio (Ahissar and Hochstein 2004). In our well-trained monkeys it is possible that two synaptic/connectivity states in the IT cortex for each task have been established. The top-down signal might be necessary just to switch the IT from one state to the other when the task changes. The IT remains in that state until the next task switch happens. Understanding the origin of the task-related signals is a challenging topic that needs further studies with specific design to address this question.

In conclusion, we found that selective rate enhancement and lower response variability in the categorization compared to the passive task, improves category representation in the IT cortex during a categorization task. Our findings suggest that modulation in the category selectivity of the same population of IT neurons could account for the behavioral demands in different tasks. These results are supported by previous studies indicating that the information for
two different tasks (categorization and identification) exists in the same neural population (Hung et al. 2005; Sugase et al. 1999). The present results provide new insight into the adaptable category representation in the IT cortex depending on the behavioral demands and also the underlying neural mechanisms of the visual object categorization.

ACKNOWLEDGMENTS

This work was supported in part by a grant from Iranian National Science Foundation (INSF). We thank Reza Rajimehr for his comments on the manuscript.

AUTHOR CONTRIBUTIONS

N.E. collected the data and performed the analysis. N.E. and H.E. designed the study and wrote the manuscript.

REFERENCES

See Online Experimental Procedures.


**FIGURE LEGENDS**

**Fig. 1.** Task and performance.

(A) Passive fixation task (ISI: inter-stimulus interval). (B) Body/object categorization task. (C) The noiseless image set. (D) Exemplar body and object images in different noise levels. Numbers below the images show percent of the added noise in each column of images. (E) Order of the tasks in a session. Blocks of passive and categorization tasks were interleaved. There were 90 trials in each block. A session could start with a categorization or passive block, which was chosen randomly at the beginning of the session. Each bar represents one block of 90 trials of
categorization (C, dark gray) or passive (P, light gray) tasks. (F) Monkeys’ performance in the categorization task. Error bars indicate the SEM.

**Fig. 2.** Improvement of the category representation during the categorization task compared to the passive task.

(A-B) D-prime of exemplar body (A) and object (B) neuron in the categorization and passive tasks measured in 100-ms sliding windows. The gray box represents period of evoked activity used for the further analysis. The dashed lines represent the SEM of each condition with 1000 bootstraps. The black bar on the x-axis represents the stimulus epoch. Data points are plotted at the middle of each bin. (C-D) Mean d-prime of population of body neurons (C) and object neurons (D) in categorization and passive tasks. The dashed lines represent the SEM of each condition across neurons. The gray line above the x-axis shows the significant difference between d-prime in categorization versus passive task (t-test, alpha = 0.05). Conventions as in Fig. 2A. (E) D-prime modulation versus categorization d-prime. Each data point represents one neuron. Error bars represent the SEM of each condition with 1000 bootstraps. Gray and black data points represent monkey 1 and 2, respectively. The arrows represent exemplar body and object neurons. (F) D-prime modulation in different noise levels. Error bars represent the SEM in each condition across neurons. (G) Performance of the classifier in categorizing images into bodies and objects as a function of the number of neurons. Error bars show the SEM for 100 random selections of the neurons. Performance of 50% represents the chance level. Solid and dashes lines above the x-axis shows the significant difference between performance in categorization and passive tasks for body and object neurons, respectively (t-test, alpha = 0.05).

**Fig. 3.** Modulation of firing rate in categorization task compared to passive task.

(A-B) The histogram of the response of the exemplar body (A) and object (B) neurons to different image categories in passive (left) and categorization (right) tasks. Shaded areas represent the SEM of each condition across trials. Conventions as described in Fig. 2A. (C-D) The histogram of the normalized response of the population of body (C) and object (D) neurons to different
image categories in passive (left) and categorization (right) tasks. Shaded areas represent the SEM of each condition across neurons. Conventions as described in Fig. 2A. (E) Response Modulation Index (RMI) difference was measured as RMI for preferred category minus RMI for non-preferred category in body (left) and object (right) neurons. The arrow represents the mean value of the distribution. The dashed vertical dashed line represents zero or no difference. Black data points represent the exemplar single neurons in each cell group.

Fig. 4. Modulation of the response variability in the categorization task compared to the passive task.

(A-B) Fano Factor (FF) of the exemplar body (A) and object (B) neurons, measured in 100-ms sliding windows in different tasks. Data points are plotted at the middle of each bin. Conventions as described in Fig. 2A. (C) Mean FF of the population of neurons in different tasks. The dashed lines represent the SEM of each condition across neurons. The gray line above the x-axis shows the significant difference between FF in categorization and passive tasks (t-test, alpha = 0.05). Conventions as described in Fig. 2A. (D) Rate-matched FF in categorization versus passive task. Each data point represents one neuron. The diamond and square data points represent the exemplar single body and object neurons, respectively. The oblique dashed line represents the diagonal line. One data point from monkey 2 is not shown here due to not being within the limits of x- and y-axes (x=5.87, y=5.19). (E) Schematic model of category representation improvement in the categorization compared to the passive task. Black and gray lines represent the categorization and the passive tasks, respectively. Solid and dashed lines represent the distribution of the neural responses to the preferred and non-preferred categories, respectively. Responses to both image categories enhanced in the categorization compared to the passive task, but more rate enhancement happened for the preferred category (both black distributions moved to the right relative to the corresponding gray ones, but the solid one moved more). Response variability decreased for both categories (narrower black distributions compared to the gray ones). Both of these modulations resulted in the larger separation of the response distributions for preferred and non-preferred and a better category discrimination (larger d-prime) in the categorization task compared to the passive task.
Figure 1

A

B

C

D

E

F

10% 30% 45% 60%

244x672

254x594

245x574

10 30 45 60

40 50 60 70

90 90 90 90

90 90 90 90

90 90 90 90

90 90 90 90

Performance (%)

10 30 45 60

Visual noise (%)
Figure 3

A

C52

Response (Hz)

Time relative to stimulus onset (ms)

-200 -100 0 100 200 300 400 500

Preferred category

Non-preferred category

B

C83

Response (Hz)

Time relative to stimulus onset (ms)

-200 -100 0 100 200 300 400 500

C

n = 51

Normalized response

Time relative to stimulus onset (ms)

-200 -100 0 100 200 300 400 500

D

n = 44

Normalized response

Time relative to stimulus onset (ms)

-200 -100 0 100 200 300 400 500

E

n = 51

Number of cells

RMI difference

-20 -10 0 10 20

n = 44

Number of cells

RMI difference

-20 -10 0 10 20
Figure 4

A, B: Graphs showing Time from stimulus onset (ms) vs. FF for C52 and C83. The graphs display data for Passive and Categorization conditions.

C, D: Graphs showing FF vs. Time from stimulus onset (ms) for C52 and C83, with a note that n = 95.

E: Graph showing Firing rate vs. Number of trials with a dashed line indicating d'.

Legend:
- Passive
- Categorization

n = 95