Form representation in monkey inferotemporal cortex is virtually unaltered by free viewing

James J. DiCarlo and John H. R. Maunsell

Howard Hughes Medical Institute and Division of Neuroscience, Baylor College of Medicine, One Baylor Plaza, S603, Houston, Texas 77030, USA
Correspondence should be addressed to J.J.D. (dicarlo@focus.neusc.bcm.tmc.edu)

How are objects represented in the brain during natural behavior? Visual object recognition in primates is thought to depend on the inferotemporal cortex (IT). In most neurophysiological studies of IT, monkeys hold their direction of gaze fixed while isolated visual stimuli are presented (controlled viewing). However, during natural behavior, primates visually explore cluttered environments by changing gaze direction several times each second (free viewing). We examined the effect of free viewing on IT neuronal responses in monkeys engaged in a form-recognition task. By making small, real-time stimulus adjustments, we produced nearly identically retinal stimulation during controlled and free viewing. Nearly 90% of neuronal responses were unaffected by free viewing, and average stimulus selectivity was unchanged. Thus, neuronal representations that likely underlie form recognition are virtually unaltered by free viewing.

Object recognition in primates is thought to depend on neuronal activity in the inferotemporal cerebral cortex (IT)[1,2], but most studies of IT neuronal responses have been done under restricted viewing and task conditions. Typically, non-human primates stare at a fixation point while isolated stimuli are flashed on the retina, often in a task that does not require stimulus identification (for example, refs. 3–7). It is not clear if neuronal activity would differ in conditions that more closely approximate the animal's natural behavior. During natural visual exploration (that is, 'active vision'[8]), animals are motivated to find and identify useful objects, objects are seldom seen in isolation, and gaze shifts occur several times per second to bring new objects onto the fovea[9–11]. This study focuses on the effect of these unconstrained changes in gaze ('free viewing').

Free viewing potentially complicates recognition because each gaze shift alters the retinal input associated with an object. On the other hand, theoretical and behavioral studies have exposed potential advantages of gaze shifts during object-recognition tasks[12,13]. Several studies in early visual areas suggest that free viewing appreciably alters neuronal responses[14–17]. However, in all these studies, the source of this response alteration was confounded among several variables, including changes in the retinal image.

Here we asked how neuronal responses change when animals freely view both cluttered and uncluttered visual displays while doing an form-recognition task. We trained monkeys to identify shapes in both controlled and free viewing conditions. We designed the conditions so that retinal stimulation was nominally identical, but we did not attempt to eliminate any other variables that might alter neuronal responses during free viewing, including motor 'intention', para-saccadic suppression or enhancement, stimulus anticipation, stimulus history and stimulus-onset abruptness[14,18].

For almost 90% of neurons, the response to each stimulus was statistically indistinguishable during the epoch of identical retinal stimulation in controlled and free viewing conditions. Among stimulus-selective neurons, average stimulus selectivity was unaffected by free viewing. Thus, we conclude that free viewing has virtually no effect on the IT neuronal representations that likely support object recognition.

Results

Two monkeys performed a form-recognition task. They were required to make a different response (saccade) for each of four target stimuli (Fig. 1). The targets were designed so that they could be discriminated only using the central retina (that is, within 2–3 degrees eccentricity). The recognition task was done in four different conditions (Fig. 2): controlled viewing without clutter, free viewing without clutter, controlled viewing with clutter, and free viewing with clutter (Methods).

During controlled viewing without clutter (Fig. 2a), the animal was required to hold its gaze within a small window around a central fixation point for 300 ms before a target was presented at the center of gaze. During free viewing without clutter (Fig. 2b), a target was presented at the center of gaze (same retinal stimulus as controlled viewing), but just after the animal had completed a saccade toward a stimulus it had not yet identified (saccade amplitude, mean ± s.d., 3.6 ± 1.3 deg). During controlled viewing with clutter (Fig. 2c), the animal was required to hold its gaze within a fixation window before a target embedded in a horizontal row of twenty, equally spaced distractors was presented at the center of gaze. During free viewing with clutter (Fig. 2d), a horizontal row of twenty-one, equally spaced distractors appeared that did not contain a target (although this was unknown to the animal). To locate a target, the animal made searching saccades along the distractor row. During some of these saccades, the entire row was briefly extinguished, and its position was slightly adjusted before it re-appeared just as the saccade ended (Fig. 2d, saccades i, ii, iv). During some of these adjustments, the distractor at the center of gaze was replaced by a target (Fig. 2d, saccade iv). Because the target was inserted just after a moderately large saccade (saccade amplitude, mean ± s.d., 6.6 ± 2.3 deg), the animal was probably unaware that the target location had previously contained a distractor.

Fig. 1. Visual stimuli and recognition task. (**a**) The set of five stimuli used by both animals. For each animal, four were designated as targets, and the other was used as a distractor. (**b**) A scaled depiction of a target stimulus embedded in a full row of (identical) distractor stimuli (clutter). The rectangle represents the display screen (±17 x ±13 deg), and the filled squares represent the response corners. To correctly perform the recognition task, the animal had to identify the target by making an eye movement to one of the four response locations. For monkey 1, the stimulus–response mapping was S3–R1, S1–R2, S4–R3, S5–R4, and S2 was the distractor. For monkey 2, the mapping was S1–R1, S2–R2, S3–R3, S4–R4, and S5 was the distractor. (**c**) A portion of the stimulus row seen during both controlled and free viewing in the presence of visual clutter (approximately half the row shown in **b**). A target stimulus is at the center of the image. The stimuli are scaled and spaced so that if the image is held at approximately arm's length (60 cm), it is comparable to that seen by the animal. The gray scale is an approximate reproduction of that used in the experiment, where the stimuli were precisely calibrated to have the same average luminance as the background. Note that the target cannot be discriminated when fixating more than one or two stimulus elements away, and that stimuli more than 4–5 spacings (7.5–9 deg) from fixation fade into the background. Scale bars, in (**a**), stimulus width (0.52 deg for monkey 1 and 0.68 deg for monkey 2); in (**c**), 0.5 deg.

In controlled viewing conditions, the center of gaze was within 0.5 deg of the (eventual) target center for at least 300 ms before target onset. In the free viewing conditions, because the target was always presented immediately following a moderately large saccade, the center of gaze was typically 3.6 deg (without clutter) or 6.6 deg (with clutter) away from the eventual target center (Fig. 3). Despite these large differences in viewing behavior, nominally identical retinal stimulation was achieved in controlled and free viewing conditions during the short epoch beginning with the presentation of the target and ending with the animal's response. During this epoch, the average absolute deviation of the center of gaze from the target center was less than 0.25 deg for all task conditions (Fig. 3). Thus, the accuracy and precision of retinal stimulation during this epoch were greater across all conditions than typically achieved within conditions in experiments with fixating animals (for example, refs. 4, 5, 7, 19, 20). In all conditions, the animal had to perform the recognition task based solely on visual information available during this epoch because it was the only time the target was present (see Fig. 2).

Both animals performed the task accurately (monkey 1 correctly identified the target in 90% of the trials and monkey 2 in 88%). Accuracy was similar in all conditions (91–92%) except free viewing with clutter (83%). This slight reduction in accuracy was due to a failure to respond to some targets, as if the animal had not noticed them (~7% of trials). The remaining incorrect trials were due to either incorrect identifications (4–5% of trials in all task conditions) or response saccades that did not reach one of the response locations (4–5% of trials in all task condi-

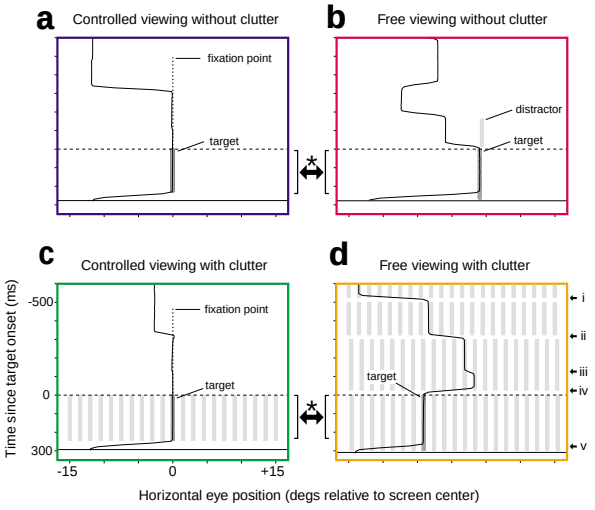

Fig. 2. The four main task conditions. Stimulus presentations and eye movements during a typical trial in each task condition. The color of the border around each example represents that task condition in all figures. The thick, dark gray, vertical bar in the lower third of each panel represents the target, and the thick, light gray bars represent distractors (in **b–d**). The width and spacing of the stimulus bars are scaled to the actual stimulus width and spacing. The top of each stimulus bar indicates the time that the stimulus was first illuminated by the monitor beam, and the bottom of each bar indicates the time that it was last illuminated. Note that the distractors in (**d**) are sometimes extinguished during a saccade (i, ii, iv) and illuminated at a new position (ii, iv) just as the saccade ends. The black horizontal bar near the bottom of each panel indicates the time that the animal's response was completed. The animal typically held its gaze on a target for ~300 ms and then initiated the appropriate response saccade. The starred brackets between the panels indicate that, for each clutter condition, the retinal stimulation is nominally identical during controlled and free viewing from 0 to ~250 ms after target onset. The arrows on the right of (**d**) mark when saccades occurred. See Methods for details.

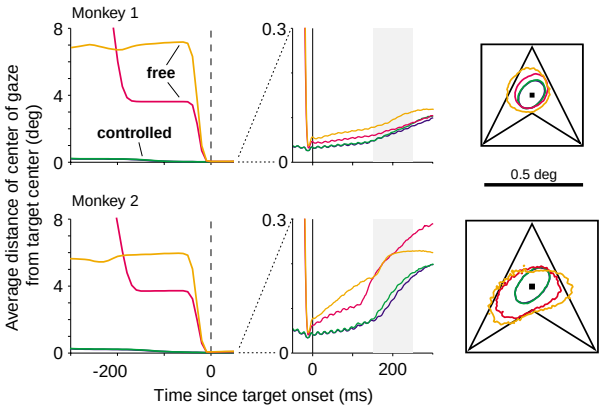

Fig. 3. Task differences in gaze behavior and control of retinal stimulation. Top, monkey 1; bottom, monkey 2. Task conditions are color coded as indicated in Fig. 2. Left, time 0 is the time that the target was first illuminated by the monitor beam. For each task condition, at each time point, the distance values for all of the trials used in the neuronal analyses have been averaged (~4300 trials in each task condition, monkey 1; ~2900, monkey 2). Before target onset, the gaze behavior is different in the controlled and free viewing conditions, but is similar in the two animals because of the task design (Methods). Center, same as left except that the ordinate scale has been greatly expanded, and each trial contributes to the average only up to the time that the animal began its response on that trial. The slight ripple between 0 and 100 ms is due to noise in the eye signal (~0.015 deg RMS) resulting from the monitor refresh. The gray background indicates the 100-ms time window in which the neuronal responses were analyzed (Methods). Right, schematic outlines of target stimuli, scaled to the size used for each animal. The colored lines are contours that contain 95% of the eye-sample values from all trials in the 50-ms period following target onset.

tions; Methods). Target stimulus identity and response type had greater effects on accuracy than did task condition (S1–S5, 86%, 81%, 95%, 93%, 84%; R1–R4, 86%, 81%, 96%, 90%; Fig. 1). All analyses are based on data from correctly completed trials.

Although the animals could respond as slowly as they liked, reaction times were typically short (mean ± s.d., monkey 1, 298 ± 68; monkey 2, 349 ± 79 ms). Both clutter and free viewing tended to slow responses slightly (controlled without clutter, 301; free view without clutter, 306; controlled with clutter, 331; free view with clutter, 356 ms). Target identity and response type had greater effects on reaction time (S1–S5, 342, 377, 291, 304, 336 ms; R1–R4, 307, 357, 283, 347 ms; Fig. 1).

We recorded data from all well-isolated cells in IT that exhibited even weak evidence of a response to any of the five stimuli in any task condition, as assessed using an audio monitor and online histograms (n = 204; monkey 1, 119; monkey 2, 85). Most neurons were recorded on the ventral surface of the brain (82%); the remainder were on the ventral bank of the superior temporal sulcus (STS). Because we could not identify any response differences among recording locations or between animals, all neurons are combined in the analyses.

Consistent with previous reports (reviewed in refs. 1, 2, 21), many IT neuronal responses were target selective. However, their response patterns were typically very similar during controlled and free viewing (for example, Fig. 4). We first looked at the entire recorded population for any overall excitatory or inhibitory effects of free viewing (versus controlled viewing; Fig. 5a and c). Across the wide range of firing rates observed in the population, free viewing had only a slight negative effect. The median response ratio (free/controlled viewing) was 0.97 without clutter and 0.93 with clutter. This corresponds to an average response change of only –1.5 spikes per second (without clutter) and –2.3 spikes per second (with clutter). If all four target stimuli were considered, the sizes of the effects were similarly small (–1.7 and –1.8 spikes/s), but significant (p = 0.0011, p = 0.0007; two-tailed, paired t-test). To eliminate bias resulting from the inclusion of unresponsive or marginally responsive neurons, we repeated the analysis on a subset of neurons that had significant responses (Methods) and obtained nearly identical results (median response ratio, 0.95 without clutter, n = 121; 0.90 with clutter, n = 117; p = 0.009, p = 0.001, two-tailed, paired t-test as above). In summary, free viewing reduced IT neuronal responses to visual stimuli by 3–10% or ~2 spikes per second.

Although free viewing had a small overall effect, we asked whether some IT neurons were substantially affected by free viewing. We were particularly interested in effects on neurons with

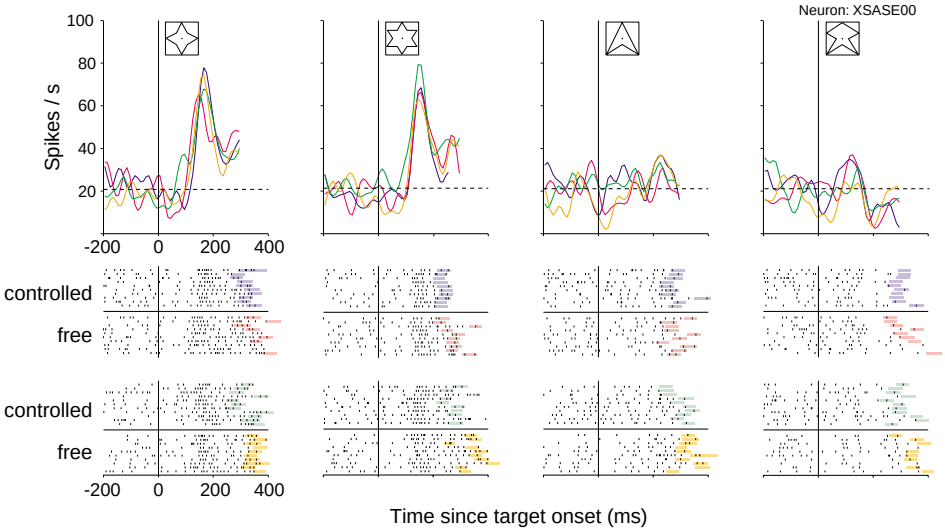

Fig. 4. Response of a target-selective IT neuron to each target stimulus in each task condition. Top, average responses to the target illustrated above; task conditions are color coded as indicated in Fig. 2. Each response curve is the average of ten trials (bottom), smoothed with a Gaussian kernel (15 ms s.d.). The horizontal dashed line indicates the average background firing rate. Bottom, each row is data from a separate trial (all trials were interleaved during collection, see Methods). Each tick mark indicates a single action potential. The colored bands indicate the animal's response saccades. Top pair of raster plots, without clutter; bottom pair, with clutter. During controlled viewing without clutter, the mean firing rate during the analysis time window was 57 spikes/s for the neuron's most-preferred ('best') stimulus and 11 spikes/s for its least-preferred ('worst') stimulus. Although the effect of visual clutter was not the focus of this study, this neuron was typical in that clutter had little effect on its response.

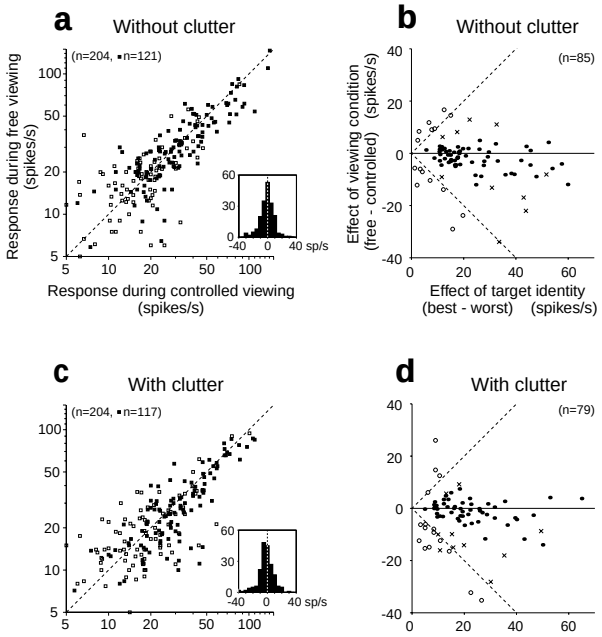

Fig. 5. Effect of viewing condition on single neuronal responses. (**a**) Each square represents a single IT neuron. Abscissa, neuron's response to its most-preferred target stimulus during controlled viewing without clutter (log scale). Ordinate, neuron's response to the same target stimulus during free viewing without clutter (log scale). The dashed line represents equal responses in both viewing conditions. Filled squares indicate neurons that showed a significant response to any target in either viewing condition (Methods). The inset histogram shows the distribution of response differences (free viewing minus controlled viewing). (**b**) Each point represents a single IT neuron that showed a significant effect of target identity or viewing condition (without clutter, Results). Abscissa, magnitude of the neuron's target selectivity (the difference in the neuron's response to its most-preferred target and its least-preferred target, averaged over viewing conditions). Ordinate, neuron's response to target stimuli during free viewing minus its response to the same stimuli during controlled viewing (averaged over target identity). ● neurons that had a significant main effect of target identity (target selective); ○ neurons that had a significant main effect of viewing condition; × neurons where both main effects were significant (Results). The dashed lines represent equally strong effects of target identity and viewing condition. (**c**, **d**) Same as (**a**) and (**b**) except effects were measured in the presence of clutter. In both (**a**) and (**c**), eleven neurons do not appear because their response rates were below the lower limits of the plot.

responses that were sensitive to target identity, because they are likely to underlie recognition. We applied a two-way analysis of variance (ANOVA) to the response rates of each of the 204 neurons, using target identity (4 levels) and viewing condition (2 levels) as factors. Sixty-nine neurons (34% in each monkey) showed a main effect of target identity (were 'target selective'). Twenty-six neurons (13%; 11% in monkey 1, 15% in monkey 2) showed a main effect of viewing condition. Only 1 neuron (< 1%) showed a significant interaction of target identity and viewing condition (at $p = 0.01$, the number expected by chance is 2), indicating that target preferences did not change between viewing conditions (Fig. 4). Similar results were obtained using the data from the two task conditions with visual clutter (31% showed an effect of target identity, 13% showed an effect of viewing condition, < 1% showed an interaction). Overall, 47 neurons (23%) were target selective by both ANOVAs (with and without clutter), but only 2 of these showed a significant effect of viewing condition in both clutter conditions.

Although the ANOVA reveals the frequency of free viewing effects (~13% of neurons), it does not speak to the size and direction of those effects. We plotted the magnitude of the viewing condition effect against the magnitude of target selectivity for all neurons for which either effect was significant by the ANOVA (Fig. 5b and d). Significant effects of target identity and viewing condition were unrelated across the population (Fisher exact test, $p > 0.05$). Without clutter, only 14% of target-selective cells showed a significant effect of viewing condition (15% with clutter). Furthermore, the effects of viewing condition tended to be small relative to the effects of target identity (Fig. 5b and d), and were both positive (free > controlled viewing) and negative (free < controlled). Because four targets but only two viewing conditions were considered, one concern is that the target identity effects are positively biased, relative to the viewing condition effects. However, controlled and free viewing conditions are two of the most extreme viewing conditions in which we could place the animal, but we have sampled the responses to just four, similar stimuli from an essentially infinite set. As a result, it is more likely that target identity effects are negatively biased.

Because only neurons with significant effects are shown (Fig. 5b and d), close inspection reveals the statistical power of the ANOVA used to assess the frequency of effects. In particular, target-identity effects as small as ~7 spikes per second (best versus worst target) and viewing condition effects of ~4 spikes per second were detected. Even though the test is sensitive, only about one third of neurons were significantly target selective because neurons were not selected for recording based on target selectivity. The critical point is that, regardless of target selec-

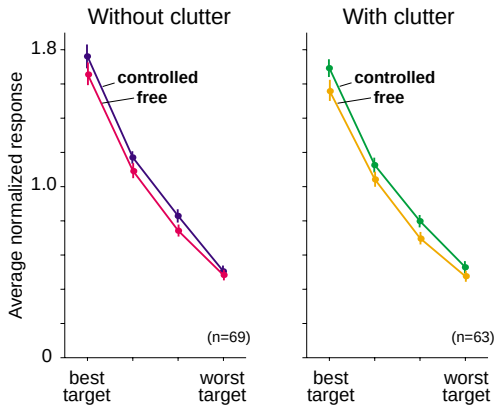

Fig. 6. Average target-sensitivity functions during controlled and free viewing. Task conditions are color coded as indicated in Fig. 2. Abscissa, the four target stimuli, ordered from most-preferred (best) to least-preferred (worst). Ordering was done separately for each neuron before averaging (Results). Ordinate, average response of target-selective neurons where each neuron's response was normalized by its overall average response. The population average response rate was 25 spikes/s without clutter (22 spikes/s with clutter). For reference, a neuron that barely reached statistical significance for target selectivity ($p = 0.009$) had ranked response rates of 15, 15, 8 and 7 spikes/s, whereas a highly significant neuron ($p < 10^{-11}$) had ranked response rates of 81, 38, 21 and 16 spikes/s (averaged over viewing condition). Bars indicate standard error.

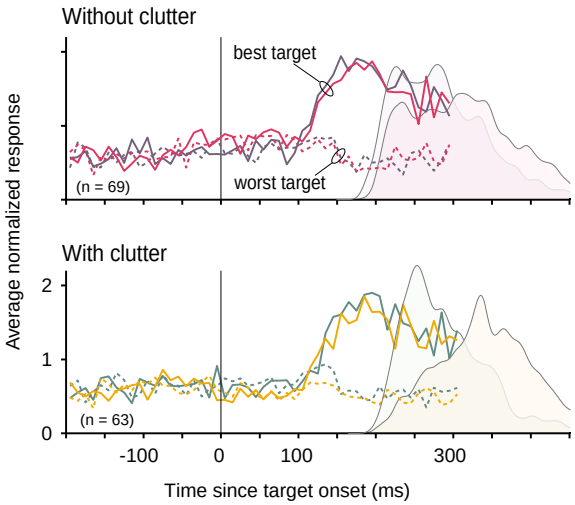

Fig. 7. Time course of the average response during controlled and free viewing. Task conditions are color coded as indicated in Fig. 2. Average response of all target-selective neurons (10-ms bin size; no filtering). Solid lines indicate the average of the each neuron's response to its most-preferred target; dashed lines indicate response to least-preferred target. Before averaging, each neuron's response curve was normalized by its overall average response in the 150–250 ms analysis window (as in **Fig. 6**). The filled distributions are reaction times (response saccade start times) for all trials contributing to each response curve. Top histograms each contain 1180 trials; bottom, 1108 trials. Each reaction time histogram was smoothed with a Gaussian kernel (7 ms s.d.). The plot truncates 2% and 5% of the reaction times in the controlled and free viewing conditions.

tivity, nearly 90% of IT neurons had responses that were statistically indistinguishable during controlled and free viewing, even though the statistical test was sensitive to effects as small as ~4 spikes per second.

Although only a small fraction of IT neuronal responses were modified by free viewing, we sought to estimate the effect of these modifications on the IT neuronal representations that support the recognition behavior. We adopted a common hypothesis, which assumes that target identity is represented by the mean firing rates (during a particular, fixed time window) of a neuronal population (for example, refs. 22, 23). We asked, among neurons with mean responses rates that were sensitive to target identity (target-selective neurons by the ANOVA), does free viewing modify this sensitivity? The responses of each target-selective neuron were rank-ordered by target identity and normalized by the neuron's average response (over all four targets and both viewing conditions), so that all neurons were given equal weight; similar results were obtained without this normalization. For each task condition, an average target-sensitivity function was computed by averaging the normalized responses in that task condition (Fig. 6). The slope of the target-sensitivity function was steep in all task conditions and was unaffected by viewing condition ($p > 0.1$, *t*-test, both with and without clutter). Thus, IT neurons, on average, are equally sensitive to target identity during controlled and free viewing. In terms of absolute spike rates, a change in target identity produced a response change of 25 spikes per second in the average IT neuron (best versus worst target), whereas free viewing produced only a slight, but significant decrease in the response to each target (–1.9 spikes/s; $p < 0.01$, both with and without clutter; two-tailed paired *t*-test).

We next examined whether viewing condition affected the time course of target-selective responses (for example, neuronal activity might develop more slowly or rapidly during free viewing). For both controlled and free viewing, the average response to the 'best' target separates from the response to the 'worst' target at the first change from background firing (Fig. 7). This occurs 75–100 ms before the animal's earliest correct responses. If the animal used IT neuronal responses to identify targets, its behavior was based on only the first 50–100 ms of stimulus-evoked activity. This period is short relative to epochs that have typically been used to characterize IT responses (for example, refs. 3, 5, 7). Nevertheless, in both viewing conditions, even a small subset of IT neurons can support stimulus identification during this brief period (Fig. 7).

We considered the possibility that the stimulus adjustments used during free viewing (that is, stimulus offsets during saccades, and onsets ~10 ms after saccade end; Methods) had altered the neuronal responses relative to those that would have occurred during free viewing of a static image. We tested this hypothesis by recording the responses of 38 IT neurons (22 in monkey 1, 16 in monkey 2) while the animal performed a free viewing task without stimulus adjustments. We compared these responses to those during free viewing with stimulus adjustments programmed to reproduce the stimulus positions obtained during free viewing of the static image. Behavioral accuracy and speed were virtually identical during free viewing of static images and stimulus-adjusted images. Without visual clutter, no neuron showed an effect of the stimulus adjustments (with visual clutter, one neuron did), and no neuron showed an interaction with target identity (two-way ANOVA, $p < 0.01$). Because identical retinal positions were used in both conditions, this does not imply that IT neurons are insensitive to retinal position, but that they are insensitive to brief (20–30 ms) stimulus drop outs and position changes during and just after saccades.

Discussion

These results show that nearly 90% of IT neuronal responses are statistically indistinguishable during controlled and free viewing, and average stimulus selectivity is unaltered. To our knowledge, this is the first direct study of the effect of free viewing on IT neuronal responses. In contrast to this report, previous studies in early visual areas have suggested that free viewing can substantially alter neuronal responses in various ways, including strong response reductions in V4, V2 and V1 (ref. 14), increases in response 'sparseness' in V1 (ref. 16), and more reliance on response 'bursts' in V1 (ref. 15). There are several possible explanations for these contrasting results.

The first explanation is that prior studies were done in brain areas that are earlier than IT in the visual cortical hierarchy[24]. Some psychophysical studies suggest that the magnocellular channel, but not the parvocellular channel, may be suppressed during saccades[25,26] (but see ref. 27). If so, this might explain some of the effects of eye movements reported in the lateral geniculate nucleus[17] and V1 (refs. 15, 16). Furthermore, it would not be surprising to find robust effects of free viewing in areas in the dorsal visual pathway[28], where responses depend on magnocellular input[29], and responses are associated with eye movements[30]. However, these psychophysical studies do not argue strongly for or against free viewing effects in IT and other ventral stream areas where both parvocellular and magnocellular channels contribute[31]. In IT, slight effects of small eye movements on background firing rates[32], and mild excitatory and inhibitory effects of

larger eye movements in complete darkness[33] have been reported. These might explain the small effects of free viewing we did observe.

A second explanation is that some studies may have failed to provide equivalent retinal stimulation during controlled and free viewing. For instance, responses of V4, V2 and V1 neurons during free viewing of large images were reported to be ~50% less than those observed during controlled viewing of small image patches[14]. In the current study, we also found that free viewing tended to reduce responses, but by only 3–10%. One hypothesis is that most of the previously reported effects were due to differences in retinal stimulation and not free viewing *per se*. Conclusive demonstrations of free viewing effects require identical retinal stimulation stimulation across conditons.

A third explanation is that previous studies used animals that were not engaged in a visual task designed to recruit the brain region under study. In particular, although clear effects of task demands have been demonstrated throughout visual cortex[34,35], previous studies of free viewing have involved simple fixation tasks or no task at all[14–17].

A few IT neurons showed a clear effect of free viewing. For some of these neurons, this might be due to differences in retinal stimulation before target onset. For example, during free viewing without clutter, some neurons were strongly activated by the initial distractor stimulus (presented ~150–200 ms before the target), and this activation continued into the analysis epoch. Other neurons had a viewing-condition effect that might be related to anticipation. That is, in controlled viewing conditions, cues reliably predicted the onset of the target stimulus (fixation point onset and tone when fixation was acquired). We did not remove these cues from the task design so that we might maximize any differences between controlled and free viewing. During controlled viewing, but not free viewing, the firing rate of some neurons began to increase before target onset (even though retinal stimulation was constant) and continued to increase into the analysis epoch. Regardless of the cause, these effects were rare (Fig. 5) and are thus probably irrelevant to the function of IT in this task.

It is important to distinguish among variables that are often confounded in the study of natural ('active'[8]) vision, but that may have different, possibly interacting, effects on neuronal responses. In particular, natural vision differs from most controlled viewing studies in at least three general ways: first, stimuli in the real world tend to be complex; second, natural vision is often goal oriented; third, stimulus-directed eye movements occur often during natural viewing. In this study, we examined the effect of the third variable while the other two were held at (nominally) constant levels that approximate those encountered during natural vision (goal-directed object recognition in a cluttered environment). Thus, we report no strong effect of free viewing itself, but we expect that neuronal activity in IT would change if the stimuli were altered to be more complex (for example, natural scenes), or if the animal performed a different task (or no task). Nevertheless, for nearly 90% of IT neurons, our data argue strongly against appreciable effects of motor 'intention', para-saccadic suppression or enhancement, stimulus anticipation, stimulus history and stimulus-onset abruptness.

The results presented here show that activity that is sensitive to object identity is rapidly evoked in IT by the retinal image (Fig. 7), regardless of how that image was brought to the retina. This implies that subjects rely on the same neuronal representation to rapidly recognize an object whose image is suddenly flashed on the retina as when they shift their gaze to foveate the same (previously unknown) object. In the somatosensory system, a similar argument of representation equivalence has been presented, based on psychophysical studies showing that pattern recognition accuracy and confusion errors are indistinguishable if stimuli are presented to a subject's passive hand, or actively scanned by the subject[36]. The present study directly supports the existence of such 'motor-invariant' object representations in the visual system. The observation that such invariant representations exist in IT substantially simplifies the problem of determining how the brain accomplishes visual recognition, but the fundamental question of how each pattern of retinal stimulation is transformed to a useful neuronal representation remains a mystery.

Methods

Animals and surgery. Experiments were done on two male rhesus monkeys (*Macaca mulatta*, 4.5 and 4.7 kg). Before behavioral training, aseptic surgery was performed to attach a head post and place a scleral search coil in the right eye. After 2–3 months of behavioral training (below), a second surgery was performed to place a recording chamber. All procedures were done in compliance with standards of the Baylor College of Medicine Animal Research Committee.

Eye-position monitoring. Horizontal and vertical eye positions were monitored using the scleral search coil[37]. Each channel was low-pass filtered (corner of 400 Hz) and sampled at 1 kHz, with a resolution of ~0.003 deg. Instrumentation time lag was less than 1.5 ms, and RMS noise in each channel was 0.025 deg.

Saccades were detected online. Eye samples were filtered with a ±2 ms boxcar; eye speed at each ms was computed and filtered with a ±2 ms boxcar. If speed exceeded 24 deg/s, the animal was defined to be making a saccade until the speed dropped below 16 deg/s. Saccades greater than 0.2 deg were reliably detected, and the saccades metrics formed a 'main sequence' (peak velocity and duration versus amplitude) that was nearly identical to that reported in humans[38] and monkeys[39].

At the start of each training or recording session, eye position was calibrated over the display region using a small, red fixation point (0.2 × 0.2 deg) displayed in 26 positions. Eight positions were used to achieve 'rough' calibration (± 1 deg) at the edges of the monitor, and 18 were used to achieve 'fine' calibration (± 0.2 deg) over the region where stimuli would be presented (6 horizontal × 3 vertical grid with separations of 4.9 horizontal and 0.8 vertical deg). The animal was rewarded for holding its eyes within ± 2 deg (rough locations) or ± 0.75 deg (fine locations) of the point for 300–600 ms. Average eye sample values were measured during the last fixation in this period, and these values were mapped to monitor positions (n = 26) with a second-order polynomial (6 free parameters for both x and y). The parameters were estimated by least squares[40], with the fine points weighted ten times the rough points. Calibration accuracy was checked using fixation points presented at ten additional monitor locations within the fine region. The errors in the monitor locations predicted by the polynomial fit were always < 0.2 deg, were distributed in all directions, and were independent of the absolute monitor location, suggesting that they were due to limitations in the animals' fixation accuracy and not to instrument calibration. The calibration procedure was occasionally rerun when the mean deviation from the central fixation point was greater than 0.1 deg.

Visual stimuli. Stimuli were presented on a video monitor (37.5 × 28.1 cm, 75 Hz refresh, non-interlaced, 1600 × 1200 pixels) positioned 62 cm from the monkey so that the display subtended ±17 horizontal and ±13 vertical deg of visual angle. The background luminance of the monitor was 22 cd/m^2; it was the only light source in the room. Both animals worked with the same fixed set of five achromatic visual stimuli. Each stimulus was composed of bright, connecting line segments that created a small, square spatial form (Fig. 1). The line segments were filtered with a difference-of-Gaussians spatial filter (0.007 deg s.d. positive, 0.01

deg s.d. negative) so that the average luminance over each stimulus was the same as the monitor background, and the peak luminance was near the monitor maximal white (46 cd/m^2). The spatial form, size and spatial frequency content of the stimuli were tuned to approximately meet three criteria. First, the stimuli were easily discriminable at eccentricities less than ~3 deg. Second, the stimuli were not discriminable beyond ~4 deg eccentricity. This encouraged the animal to search the array of distractors during free viewing, and minimized the chance that the animal could notice the change from distractor to target that occurred during free viewing. Third, the stimuli were not detectable beyond ~8 deg eccentricity. This caused stimulus configurations during controlled viewing to appear nearly identical to configurations during free viewing (up to ±9 deg from the monitor center). The final stimulus size was based on each monkey's recognition performance with stimuli placed at a range of eccentricities (monkey 1, 0.52 deg width; monkey 2, 0.68 deg width; Figs. 1 and 3).

Behavioral task and training. Each animal performed a spatial form-recognition task. Four of the five stimuli were designated as targets; the remaining stimulus was the distractor (Fig. 1). Four response locations near the corners of the monitor (16.8 deg from the display center; Fig. 1b) were indicated by identical white squares (0.6 × 0.6 deg, 46 cd/m^2). For each animal, each response location was assigned a different target, and this mapping never changed. When a target was presented, the animal signaled its identity by making a saccade to the appropriate response location (Fig. 1b). Saccades that ended within a window (±11.9 deg horizontal and ±4 deg vertical) around any response location were scored as a response. All correct trials were rewarded with juice. Reaction times were defined relative to the start of the response saccade.

Animals were first trained to perform the task during controlled viewing. Both animals performed well above chance during both free viewing conditions on the first session in which they were attempted.

Task conditions. For controlled viewing task conditions (Fig. 2a and c), each trial began with the simultaneous presentation of a small, white fixation point (0.1 × 0.1 deg) and a 30-ms tone. The animal was required to bring its gaze to the fixation point and hold its gaze within ±0.4 or ±0.5 deg of the point. The fixation point was extinguished 300 ms after acquisition, and one of the four target stimuli was presented. Because we desired identical retinal stimulation in all conditions, and stimulus position variability can produce neuronal response variability[41], the target was always presented at the current center of gaze. For controlled viewing with clutter, the target was flanked by a horizontal row of 20 identical distractors with a 1.5 deg center-to-center separation (Figs. 1 and 2).

For free viewing without clutter (Fig. 2b), the animal was typically looking around the screen when the trial began. If one of the animal's fixations happened to be near the stimulus presentation region (±12 horizontal × ±0.3 vertical deg region in the center of the display), a single distractor stimulus was presented at a random eccentricity (1.5–6.0 deg, uniform probability), centered vertically on the display. Because the distractor stimulus was within 8 deg eccentricity, it was detected (but presumably not recognized), and the animal typically initiated a saccade to fixate it. During the saccade, the distractor was extinguished. Just as the saccade ended near the (previous) distractor location, a target stimulus was presented on the (just) stabilized center of gaze (mean displacement from the previous distractor, 0.55 ± 0.21 deg). The time between the detected saccade end and target illumination by the monitor beam was 10.8 ± 3.8 (s.d.) ms. For some trials, the target was not presented because the animal's saccade toward the distractor did not end within ±0.75 deg of the just-extinguished distractor.

For free viewing with clutter (Fig. 2d), the trial began with the presentation of a horizontal row of 21 distractors located at the monitor center. The animal typically brought its gaze into the distractor row and began to make saccades in search of a target (mean intersaccadic interval, 199 ms; Fig. 2d). For saccades smaller than 4.5 deg, the display remained unchanged. However, halfway through each saccade estimated (online) to achieve an amplitude greater than 4.5 deg ('large' saccades), the entire distractor row was extinguished (note the gaps in the distractor bars in Fig. 2d after saccades i, ii and iv). Just as each of these large saccades ended, the entire row was redisplayed. The mean delay between detected saccade end and stimulus illumination was 10.1 ms (±4.5 ms s.d.). After ~50% percent of (randomly chosen) large saccades that ended in the stimulus presentation region, a target was presented. This was done before redisplaying the row by replacing the (previous) distractor stimulus location nearest the current center of gaze with a target and shifting the entire row slightly from its previous location so that the target was centered at the center of gaze (Fig. 2d, saccade iv). The stimulus offset distance was never more than 0.88 deg (mean ± s.d., 0.42 ± 0.20 deg). For the remaining large saccades that ended in the stimulus presentation region, the entire redisplayed row was adjusted so that one of the distractors was at the center of gaze (for example, Fig. 2d, after saccade ii) or adjusted 0.2 deg in a random direction. For large saccades that ended outside the stimulus-presentation region, the distractor row was redisplayed at its previous location (for example, Fig. 2d, after saccade i). On average, the animal made 5.7 (± 5.2, s.d.) searching saccades before a target was presented.

During free viewing with clutter, the animal typically responded to any target with an appropriate response saccade (such as Fig. 2d). However, if gaze moved more than 3.0 horizontal or 0.8 vertical deg from the target (for example, along the distractor row), the distractor row remained illuminated, the target was replaced with a distractor, the trial was scored as a failure to respond, and a (new) trial continued without interruption.

Recording and data collection. A guide tube (23 G) was used to reach IT using a dorsal to ventral approach. Recordings were made using glass-coated Pt/Ir electrodes (0.5–1.5 MΩ at 1 kHz), and neuronal spikes were amplified, filtered and discriminated using standard equipment. IT was identified by comparing gray and white matter transitions and the depth of the skull base with standard atlas frontal sections. Recordings were made over a ~10 x 10 area of the ventral STS and ventral surface (Horsely–Clark AP, 10–20 mm; ML, 14–24 mm).

The trials for each task condition were run as a block. Each block contained two correctly completed trials of each target, and each trial lasted ~2–10 s. Both animals typically completed each block in 1–3 minutes before switching to a new block (randomly chosen). The animal cycled through blocks as the electrode was advanced into IT. Data were collected from even marginally responsive cells under the assumption that longer periods of observation might reveal statistically detectable effects. Data were included in the analyses if isolation was maintained for at least six (mean, 8.5; max, 10) presentations of each target in each task condition (~20–35 minutes of recording). A computer controlled all stimulus generation, behavioral monitoring and data recording (temporal resolution, 1 ms).

Analysis. Trials were excluded if eye movements greater than 0.3 deg occurred during the first 50 ms after target onset (< 1% of all correct trials), or if the animal began its response less than 100 ms after target onset (<< 1% of all correct trials). To insure that only conditions with identical retinal stimulation were compared, analyses of the effect of free viewing were performed separately for each clutter condition.

Analyses of the effects of target identity and viewing condition on neuronal responses (Figs. 5 b and d and 6), including two-way ANOVAs, were based on the firing rate in a 100-ms time window beginning 150 ms after target onset. This analysis window was chosen based on visual inspection of the data and previous reports showing than ~90% of neurons in anterior IT have started to respond by 150 ms[19,42]. For trials in which the animal began its response saccade before the end of the analysis window (21% of all trials, see Fig. 7), the average rate from 150 ms after target onset until the beginning of the response was used. The use of other time windows, including windows aligned on response start, produced similar results. ANOVA analysis of square-root-transformed responses (that is, variance stabilizing, assuming Poisson spiking[40]) produced a nearly identical outcome to that described in Results.

For some analyses (for example, Fig. 5a and c), each neuron was tested for any significant responses (not necessarily target selective). The rate in the 100-ms analysis window was compared with the rate in a 200-ms window ending 50 ms before target onset (background rate) using a *t*-test. A neuron was considered to have a significant response if any *t*-test was significant (8 tests = 4 targets × 2 viewing conditions) at a level of 0.0064. (The overall false-positive level of the test was 0.05.)

For analyses where neuronal responses were rank-ordered by target identity (for example, Figs. 5–7), the rank ordering was done once for each neuron and was based on the response rate in the analysis window averaged over controlled and free viewing conditions.

Acknowledgements

We thank C. Boudreau, E. Cook, G. Ghose, and T. Yang for discussions on design, analysis and presentation, and D. Murray for animal husbandry. We also thank K. Johnson and D. Sparks for comments on a previous version of this manuscript. This work was supported by NIH EY05911. J.H.R.M. is an Investigator with the Howard Hughes Medical Institute.

Received 6 April; Accepted 14 June 2000

1. Tanaka, K. Inferotemporal cortex and object vision. *Annu. Rev. Neurosci.* 19, 109–139 (1996).
2. Miyashita, Y. Inferior temporal cortex: where visual perception meets memory. *Annu. Rev. Neurosci.* 16, 245–263 (1993).
3. Booth, M. C. A. & Rolls, E. T. View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. *Cereb. Cortex* 8, 510–523 (1998).
4. Sugase, Y., Yamane, S., Ueno, S. & Kawano, K. Global and fine information coded by single neurons in the temporal visual cortex. *Nature* 400, 869–873 (1999).
5. Missal, M., Vogels, R., Li, C. & Orban, G. A. Shape interactions in macaque inferior temporal neurons. *J. Neurophysiol.* 82, 131–142 (1999).
6. Rollenhagen, J. E. & Olson, C. R. Mirror-image confusion in single neurons of the macaque inferotemporal cortex. *Science* 287, 1506–1508 (2000).
7. Gibson, J. R. & Maunsell, J. H. R. The sensory modality specificity of neural activity related to memory in visual cortex. *J. Neurophysiol.* 78, 1263–1275 (1997).
8. Findlay, J. Active vision: Visual activity in everyday life. *Curr. Biol.* 8, R640–R642 (1998).
9. Motter, B. C. & Belky, E. J. The zone of focal attention during active visual search. *Vision Res.* 38, 1007–1022 (1998).
10. Sommer, M. A. Express saccades elicited during visual scan in the monkey. *Vision Res.* 34, 2023–2038 (1994).
11. Burman, D. D. & Segraves, M. A. Primate frontal eye field activity during natural scanning eye movements. *J. Neurophysiol.* 71, 1266–1271 (1994).
12. Ballard, D. Animate vision. *Artificial Intelligence* 48, 57–86 (1991).
13. Dawkins, M. S. & Woodington, A. Pattern recognition and active vision in chickens. *Nature* 403, 652–655 (2000).
14. Gallant, J. L., Connor, C. E. & Van Essen, D. C. Neural activity in areas V1, V2 and V4 during free viewing of natural scenes compared to controlled viewing. *Neuroreport* 9, 2153–2158 (1998).
15. Livingstone, M. S., Freeman, D. C. & Hubel, D. H. Visual responses in V1 of freely viewing monkeys. *Cold Spring Harb. Symp. Quant. Biol.* 61, 27–37 (1996).
16. Vinje, W. E. & Gallant, J. L. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science* 287, 1273–1276 (2000).
17. Guido, W. & Weyand, T. Burst responses in thalamic relay cells of the awake behaving cat. *J. Neurophysiol.* 74, 1782–1786 (1995).
18. Judge, S. J., Wurtz, R. H. & Richmond, B. J. Vision during saccadic eye movements. I. Visual interactions in striate cortex. *J. Neurophysiol.* 43, 1133–1155 (1980).
19. Vogels, R. & Orban, G. A. Activity of inferior temporal neurons during orientation discrimination with successively presented gratings. *J. Neurophysiol.* 71, 1428–1451 (1994).
20. Liu, Z. & Richmond, B. J. Response differences in monkey TE and perirhinal cortex: stimulus association related to reward schedules. *J. Neurophysiol.* 83, 1677–1692 (2000).
21. Logothetis, N. K. & Sheinberg, D. L. Visual object recognition. *Annu. Rev. Neurosci.* 19, 577–621 (1996).
22. Gochin, P. M., Colombo, M., Dorfman, G. A., Gerstein, G. L. & Gross, C. G. Neural ensembly coding in inferior temporal cortex. *J. Neurophysiol.* 71, 2325–2337 (1994).
23. Britten, K. H., Shadlen, M. N., Newsome, W. T. & Movshon, J. A. The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J. Neurosci.* 12, 4745–4765 (1992).
24. Felleman, D. J. & Van Essen, D. C. Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* 1, 1–47 (1991).
25. Ross, J., Burr, D. & Morrone, C. Suppression of the magnocellular pathway during saccades. *Behav. Brain Res.* 80, 1–8 (1996).
26. Diamond, M. R., Ross, J. & Morrone, M. C. Extraretinal control of saccadic suppression. *J. Neurosci.* 20, 3449–3455 (2000).
27. Castet, E. & Masson, G. S. Motion perception during saccadic eye movements. *Nat. Neurosci.* 3, 177–183 (2000).
28. Ungerleider, L. G. & Mishkin, M. in *Analysis of Visual Behavior* (eds. Ingle, D. J., Goodale, M. A. & Mansfield, R. J. W.) 549–585 (MIT Press, Cambridge, Massachusetts, 1982).
29. Maunsell, J. H. R., Nealey, T. A. & DePriest, D. D. Magnocellular and parvocellular contributions to responses in the middle temporal visual area (MT) of the macaque monkey. *J. Neurosci.* 10, 3323–3334 (1990).
30. Colby, C. L. & Goldberg, M. E. Space and attention in parietal cortex. *Annu. Rev. Neurosci.* 22, 319–349 (1999).
31. Ferrera, V. P., Nealey, T. A. & Maunsell, J. H. R. Responses in macaque visual area V4 following inactivation of the parvocellular and magnocellular LGN pathways. *J. Neurosci.* 14, 2080–2088 (1994).
32. Leopold, D. A. & Logothetis, N. K. Microsaccades differentially modulate neural activity in the striate and extrastriate visual cortex. *Exp. Brain Res.* 123, 341–345 (1998).
33. Ringo, J. L., Sobotka, S., Diltz, M. D. & Bunce, C. M. Eye movements modulate activity in hippocampal, parahippocampal, and inferotemporal neurons. *J. Neurophysiol.* 71, 1285–1288 (1994).
34. Maunsell, J. H. R. The brain's visual world: representation of visual targets in cerebral cortex. *Science* 270, 764–769 (1995).
35. Desimone, R. & Duncan, J. Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* 18, 193–222 (1995).
36. Vega-Bermudez, F., Johnson, K. O. & Hsiao, S. S. Human tactile pattern recognition: active versus passive touch, velocity effects, patterns of confusion. *J. Neurophysiol.* 65, 531–546 (1991).
37. Robinson, D. A. A method of measuring eye movements using a scleral search coil in a magnetic field. *IEEE Trans. Biomed. Eng.* 101, 131–145 (1963).
38. Bahill, A., Clark, M. & Stark, L. Glissades—eye movements generated by mismatched components of the sacadic motorneuronal control signal. *Math. Biosciences* 26, 303–318 (1975).
39. Martinez-Conde, S., Macknik, S. L. & Hubel, D. H. Microsaccadic eye movements and firing of single cells in the striate cortex of macaque monkeys. *Nat. Neurosci.* 3, 251–258 (2000).
40. Snedecor, G. W. & Cochran, W. D. *Statistical Methods* (Iowa Univ. Press, Ames, Iowa, 1967).
41. Gur, M. & Snodderly, D. M. Studying striate cortex neurons in behaving monkeys: benefits of image stabilization. *Vision Res.* 27, 2081–2087 (1987).
42. Baylis, G. C., Rolls, E. T. & Leonard, C. M. Functional subdivisions of the temporal lobe neocortex. *J. Neurosci.* 7, 330–342 (1987).