

Anterior Inferotemporal Neurons of Monkeys Engaged in Object Recognition Can be Highly Sensitive to Object Retinal Position

James J. DiCarlo and John H. R. Maunsell

Howard Hughes Medical Institute and Division of Neuroscience, Baylor College of Medicine, Houston, Texas 77030

Submitted 10 May 2002; accepted in final form 6 February 2003

DiCarlo, James J. and John H. R. Maunsell. Anterior inferotemporal neurons of monkeys engaged in object recognition can be highly sensitive to object retinal position. *J Neurophysiol* 89: 3264–3278, 2003; 10.1152/jn.00358.2002. Visual object recognition is computationally difficult because changes in an object's position, distance, pose, or setting may cause it to produce a different retinal image on each encounter. To robustly recognize objects, the primate brain must have mechanisms to compensate for these variations. Although these mechanisms are poorly understood, it is thought that they elaborate neuronal representations in the inferotemporal cortex that are sensitive to object form but substantially invariant to other image variations. This study examines this hypothesis for image variation resulting from changes in object position. We studied the effect of small differences ($\pm 1.5^\circ$) in the retinal position of small (0.6° wide) visual forms on both the behavior of monkeys trained to identify those forms and the responses of 146 anterior IT (AIT) neurons collected during that behavior. Behavioral accuracy and speed were largely unaffected by these small changes in position. Consistent with previous studies, many AIT responses were highly selective for the forms. However, AIT responses showed far greater sensitivity to retinal position than predicted from their reported receptive field (RF) sizes. The median AIT neuron showed a $\sim 60\%$ response decrease between positions within $\pm 1.5^\circ$ of the center of gaze, and 52% of neurons were unresponsive to one or more of these positions. Consistent with previous studies, each neuron's rank order of target preferences was largely unaffected across position changes. Although we have not yet determined the conditions necessary to observe this marked position sensitivity in AIT responses, we rule out effects of spatial-frequency content, eye movements, and failures to include the RF center. To reconcile this observation with previous studies, we hypothesize that either AIT position sensitivity strongly depends on object size or that position sensitivity is sharpened by extensive visual experience at fixed retinal positions or by the presence of flanking distractors.

INTRODUCTION

Although we effortlessly perform object recognition thousands of times per day, it is a remarkably difficult computational task (Edelman 1999; Ullman 1996). The key computational problem the brain must solve is that the same object can produce a wide variety of sensory images (Edelman 1999; Riesenhuber and Poggio 2000; Ullman 1996). In the visual domain, retinal image variations arise from changes in object position, scale (e.g., viewing distance), orientation, pose, and illumination as well as the presence of other objects in the visual scene. How does the brain tolerate this tremendous variability to identify the object? In this report, we present data

aimed at understanding how behaving animals tolerate one type of image variability—that due to changes in object position relative to the center of gaze.

Object position changes are a common source of image variation because they occur frequently when environments are explored with eye, head, or body movements. Yet even in the face of such position variation, we easily carry out behaviors that depend on recognition. Indeed, some studies suggest that recognition can tolerate changes of $\geq 5^\circ$ (Biederman and Cooper 1991; Ellis et al. 1989). However, others indicate that the position tolerance of recognition depends on visual experience and the similarity of the objects to be distinguished (Dill and Edelman 2001; Dill and Fahle 1997, 1998; Foster and Kahn 1985; Nazir and O'Regan 1990).

Any theory that can explain some range of position tolerance in recognition behavior must include mechanisms that transform retinal images to neuronal signals that are sensitive to object form but are largely insensitive to object position over that range. That is, neuronal signals that are at least as position tolerant as the behavior must exist somewhere in the brain because the behavior dictates their presence at the level of motor neurons. Such neurons could be described as having large receptive fields (RFs) in that they respond selectively to objects over all retinal positions at which recognition occurs. However, because it would be inappropriate to describe motor neurons as having large RFs, we use the term position sensitivity because it can be applied without confusion to the neuronal responses along the entire stimulus-motor chain of processing.

Although many mechanisms have been proposed to create object-selective, position-tolerant signals in the brain (e.g., Biederman 1987; Mel 1997; Olshausen et al. 1993; Riesenhuber and Poggio 1999; Salinas and Abbott 1997; Ullman 1996), the actual mechanisms are unknown, and the brain regions thought to contain these signals are poorly understood. The dominant hypothesis is that these mechanisms operate in the ventral visual processing stream of the cerebral cortex and produce position-tolerant patterns of neuronal activity at the highest level of that stream—the anterior inferotemporal cortex (AIT) (Gross 1973; Logothetis and Sheinberg 1996; Tanaka 1996; Ungerleider and Mishkin 1982). Indeed, inferotemporal cortex (IT) likely plays a central role in object recognition because IT lesions (Dean 1982; Weiskrantz and Saunders 1984) or inactivation (Horel 1996) impair recognition, and IT

Address for reprint requests: Corresponding author: James J. DiCarlo, Massachusetts Institute of Technology, E25–242, 45 Carleton St., Cambridge, MA 02139. (617) 452-2045, FAX: (617) 253-2964, e-mail: dicarlo@mit.edu

The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

neuronal responses are selective for complex stimulus forms (Logothetis and Sheinberg 1996; Miyashita 1993; Tanaka 1996), such as faces (Desimone et al. 1984; Perrett et al. 1982).

The strongest statement of the IT position-tolerance hypothesis predicts that IT responses should be highly sensitive to stimulus form (i.e., identity) and completely insensitive to stimulus position (within the visual field). It is already well known that this strict interpretation is not true because previous studies show that IT neurons have finite RFs and that IT responses often decrease with changes in stimulus position away from the RF center (Boussaoud et al. 1991; Desimone et al. 1984; Gross et al. 1969, 1972; Ito et al. 1995; Kobatake and Tanaka 1994; Leuschow et al. 1994; Logothetis et al. 1995; Missal et al. 1999; Op de Beeck and Vogels 2000; Richmond et al. 1983; Sary et al. 1993; Schwartz et al. 1983; Tovée et al. 1994). Furthermore, IT neurons are often described as having only a relative form of position tolerance in which the neuron's overall responsiveness decreases with changes in position but its rank order of target preferences remains the same (e.g., Logothetis and Sheinberg 1996). We do not yet know if or how this relative position tolerance supports nonrelative behavioral position tolerance. Nevertheless, IT neurons have been shown to maintain this relative position tolerance over visual regions $\geq 10^\circ$ in diameter (Ito et al. 1995; but see Logothetis et al. 1995 and discussion; Sary et al. 1993; Schwartz et al. 1983; Tovée et al. 1994). Thus all of these studies suggest that IT neurons maintain responsiveness over large regions of visual space—that is, they have large RFs. Indeed, standard RF mapping methods indicate that AIT neurons have very large RFs ($10\text{--}30^\circ$ in diameter) (Boussaoud et al. 1991; Desimone et al. 1984; Gross et al. 1969, 1972; Kobatake and Tanaka 1994; Op de Beeck and Vogels 2000; Richmond et al. 1983).

Although previous studies indicate that AIT neurons maintain relative form selectivity over large RFs, it is not known if or how these neuronal responses compare with the position tolerance of the recognition behavior they are thought to support. We therefore sought to understand the neuronal responses to one or more recognition targets placed within the large RFs of form-selective AIT neurons while animals performed form-recognition tasks. To this end, we trained animals to recognize and report the identity of familiar objects and developed a technique that allowed presentation of visual stimuli to arbitrary retinal positions with an accuracy of $\sim 0.1^\circ$, even in free-viewing animals (DiCarlo and Maunsell 2000). We first sought to confirm the large RF property of AIT neurons by presenting stimuli at three closely spaced retinal positions (-1.5 , 0 , and $+1.5^\circ$). Based on the studies described in the preceding text, these positions should have all been well within the RFs of essentially all AIT neurons. Unexpectedly, most AIT neurons were highly sensitive to these small changes in stimulus position.

METHODS

Animals and surgery

Experiments were performed on two male rhesus monkeys (*Macaca mulatta*) weighing 4.5 and 4.7 kg. Before behavioral training, aseptic surgery was performed to attach a head post to the skull and to implant a scleral search coil in the right eye. After 2–3 mo of behavioral training (following text), a second surgery was performed to place a recording chamber (18 mm diam) to reach the anterior half

of the left temporal lobe (chamber Horsley-Clark center = 15 mm A). All animal procedures were performed in compliance with the standards of the Baylor College of Medicine Animal Research Committee and the American Physiological Society.

Eye-position monitoring

Horizontal and vertical eye positions were monitored using the scleral search coil (Robinson 1963). Each channel was low-pass filtered at a corner frequency of 400 Hz and was digitally sampled at 1 kHz with a resolution of $\sim 0.003^\circ$. The instrumentation time lag was < 1.5 ms, the RMS noise in each channel was 0.025° , and accuracy was $\sim 0.1^\circ$. Saccades greater than $\sim 0.2^\circ$ were reliably detected in real time using speed criteria (saccade start: speed $> 24^\circ/\text{s}$; saccade end: speed $< 16^\circ/\text{s}$). The methods for detecting saccades and calibrating retinal locations with monitor locations are described in detail elsewhere (DiCarlo and Maunsell 2000).

Visual stimuli

Stimuli were presented on a video monitor (37.5×28.1 cm, 75 Hz frame rate, $1,600 \times 1,200$ pixels) positioned 62 cm from the monkey so that the display subtended ± 17 (h) and ± 13 (v) $^\circ$ 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 forms (Fig. 1A). Each form was constructed by connecting line segments (0.02° width) to form the stimulus outline. This outline shape was then convolved with a difference-of-Gaussians spatial filter (0.01° SD positive, 0.02° SD negative) so that the average luminance over each form was the same as the monitor background (Fig. 10A). The peak luminance was set to the monitor maximal white (46 cd/m^2). The size and spatial frequency content of the forms were tuned to allowed us to study both the effects of free viewing (DiCarlo and Maunsell 2000) and of stimulus position

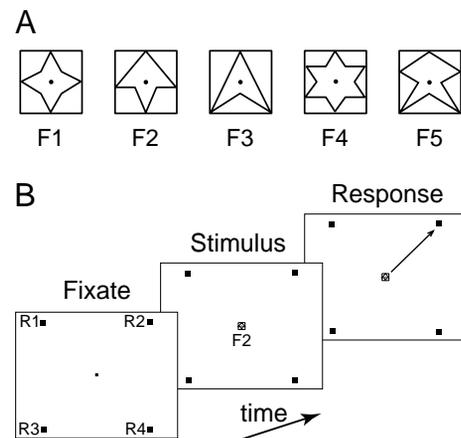


FIG. 1. Stimuli and behavioral task. A: schematic illustrations of the 5 visual forms used by both animals (F1–F5). For approximate gray-scale reproductions of each form, see Fig. 10A and Fig. 1 of (DiCarlo and Maunsell 2000). For each animal, 4 forms were designated as targets, and the other was used as a distractor (visual clutter, see Fig. 13). The form width (edge to edge) was 0.52° for the *monkey 1* and 0.68° for *monkey 2*. B: temporal sequence illustrating 1 trial of the primary behavioral task. Each panel represents the display screen ($34 \times 26^\circ$); ■, the response corners (R1–R4). Trials began with the animal fixating a small point in the center of the display. After 300 ms of fixation, 1 of the 5 forms was presented in 1 of 3 retinal positions along the horizontal meridian (1.5° left of the center of gaze, at the center of gaze, or 1.5° right of the center of gaze; the central condition is illustrated here). To correctly perform the recognition task, the animal had to identify the form by making an eye movement (saccade) to the appropriate response location. For *monkey 1*, the stimulus form to response mapping was: (F3–R1), (F1–R2), (F4–R3), and (F5–R4) and F2 was the distractor. For *monkey 2*, the mapping was: (F1–R1), (F2–R2), (F3–R3), and (F4–R4) and F5 was the distractor.

(current study). Specifically, based on the animal's performance with stimuli placed at a range of eccentricities, we chose the stimulus size so that recognition accuracy was good for stimuli at 1.5° eccentricity (Fig. 2) but was approaching chance levels for stimuli at $\sim 5^\circ$ eccentricity (*monkey 1* = 0.52° width, *monkey 2* = 0.68° width). Although acuity limits depend on the forms to be distinguished, at 1.5° eccentricity acuity is reduced to 40–60% of that observed at the center of gaze (Ludvigh 1941; Merigan and Katz 1990), and retinal cone density is $\sim 40\%$ of maximal (Curcio et al. 1987; Perry and Cowey 1985).

Some neurons in *monkey 1* were also studied with a second set of target objects that had the same shapes as the original stimuli but substantially different elemental spatial frequency content (Fig. 10, see RESULTS). These were constructed with the same outline shapes, except that the outlines were 0.04° wide and were not filtered with the difference-of-Gaussians spatial filter. Instead, to keep the average luminance over the stimulus near the background luminance, each of these stimulus shapes was added to a negative, (i.e., below the average luminance), circularly symmetric Gaussian (0.3° SD). The amplitude of this Gaussian was set so that the average luminance over a 2° square window centered on the stimulus was the same as the background luminance.

Basic form recognition task

Both animals performed a form recognition task. Four of the five stimulus forms were designated as targets; the remaining form was the distractor (Fig. 1A). Four response locations near the corners of the monitor (16.8° from the display center) were at all times indicated by identical white squares ($0.6 \times 0.6^\circ$, 46 cd/m^2 ; Fig. 1B). For each animal, each target form was assigned a different response location, and this mapping never changed. When a target was presented, the animal was required to signal the target form by making a saccade to the appropriate response location. Saccades that ended within a window [$\pm 11.9^\circ$ (h) and $\pm 4^\circ$ (v)] around any response location were scored as a response. The horizontal width of these windows was chosen to ensure that the animal would register a response if it produced the same saccade vector from a broad range of absolute horizontal eye positions where targets could be encountered during free viewing studies, described elsewhere (DiCarlo and Maunsell 2000). Correct responses produced a juice reward and a brief tone. Reaction time was defined as the duration between target onset and the start of the response saccade.

Each trial began with the presentation of a small, white fixation point ($0.1 \times 0.1^\circ$) near the display center (Fig. 1B). The animal was required to bring and hold its gaze within $\pm 0.5^\circ$ of the point. The fixation point was extinguished 300 ms after acquisition, and one of the five forms was immediately presented in one of three positions: at the center of gaze, 1.5° to the left of the center of gaze (ipsilateral to the recorded hemisphere), or 1.5° to the right of the center of gaze (contralateral to the recorded hemisphere). Because we desired identical retinal stimulation for all trials within a condition and because position variability on the retina can produce neuronal response variability (Gur and Snodderly 1987), the three positions were always specified relative to the animal's center of gaze at the end of the fixation period. That is, the three positions were specified in retinal coordinates rather than monitor coordinates. Over all recording sessions, the mean center of gaze at the end of the fixation period was 0.01° (h) and 0.13° (v) (*monkey 1*) and -0.02° (h) and 0.14° (v) (*monkey 2*) from the fixation point center (h and v SD $\sim 0.14^\circ$ in both monkeys). On each trial, the stimulus form (4 target forms) and the position of the form (3 possible positions) were each randomly chosen with equal likelihood and were presented only briefly (mean: ~ 290 ms, see following text). Thus the animal could not bias spatial or featural attention differently on each trial because it could not predict the position or form of the target. These 12 trial types were presented

in blocks such that a correctly completed trial type was not presented again until all trial types were correctly completed.

After a target form was presented, the animal was allowed to respond as rapidly as it liked. If the animal made a saccade that ended $>3^\circ$ (h) or 1° (v) from the fixation point but did not reach one of the response widows, the trial was scored as a failed trial ($\sim 4\%$ of trials). Any eye movement that brought the center of gaze out of the fixation window ($\pm 0.5^\circ$ around the initial fixation point) caused the stimulus to be immediately extinguished. Indeed, on $\sim 97\%$ of trials in which a form was presented to the left or right of the center of gaze, the animal made a small "adjustment" saccade (mean amplitude = 1.1° ; mean duration = 23 ms; latency mean and SD = 140 ± 33 ms) toward the form, and the form was extinguished during this saccade. The animals generated these adjustment saccades without training, and we did not attempt to modify this behavior. Extinguishing the target during the adjustment saccade ensured that the animal could not acquire information about target form from a retinal position other than the initially stimulated position (see Fig. 11). In these trials, the monitor phosphors that comprised the form were last excited 22.5 ms (mean; 95% range = 10–36 ms) before the saccade out of the fixation window was completed. Because the phosphors decayed exponentially with a time constant of <1 ms, the extinguished form could not have been visible at the end of the adjustment saccade (Michelson contrast $<10^{-9}$ on average; 95% upper bound = 5×10^{-5}). After the adjustment saccade, the animal's gaze typically remained at the new, now empty, position (i.e., near the original target position) for ~ 150 ms before the animal began its response saccade (i.e., to 1 of the 4 response locations). This pattern of eye movements was observed in essentially all correct trials in both animals (*monkey 1*: 93% of central position trials, 95% of eccentric position trials; *monkey 2*: 88%, 99%; see Fig. 3, top). In the remaining central position trials, the animals made a small saccade (typically $<0.5^\circ$) before the response saccade. In the remaining eccentric position trials, no adjustment saccade was detected.

Additional task conditions

We also recorded data while the animal performed the basic recognition task in the presence of visual clutter. For these trials, the single target form was embedded in a horizontal row of 20 identical distractor forms with a 1.5° center-to-center separation (see Fig. 13) (see also Fig. 1 of DiCarlo and Maunsell 2000). Trials run with clutter were run in separate blocks, and these blocks were interleaved with the primary behavioral task blocks.

Monkey 1 was also studied in a version of the basic recognition task in which target shapes were presented not just at the central three positions but also at more eccentric positions along the horizontal meridian ($\pm 4.5^\circ$ in 1.5° increments). Initially, the animal's performance was better than chance for targets presented in these more eccentric positions ($\sim 52\%$ accuracy; chance is 25%), indicating that the animal had generalized the task (i.e., shape identification regardless of retinal position). After ~ 2 wk of training, performance gradually improved but was still not as good as the central three positions (see RESULTS) and was very poor for some target shapes. Because of this, we did not force the animal to complete an equal number of correct trials for each target in each position but instead included neuronal response data from all trials in which the target was presented, regardless of the behavioral outcome (i.e., correct, wrong, or failed).

Recording and data collection

A guide tube (23 G) was used to reach AIT using a dorsal to ventral approach. Recordings were made using glass-coated Pt/Ir electrodes ($0.5\text{--}1.5 \text{ M}\Omega$ at 1 kHz), and spikes from individual neurons were amplified, filtered, and isolated using conventional equipment. The superior temporal sulcus (STS) and the ventral surface were identified

by comparing gray and white matter transitions and the depth of the skull base with atlas sections. Penetrations were made over a $\sim 10 \times 10$ mm area of the ventral STS and ventral surface (Horsley-Clark AP: 10–20 mm, ML: 14–24 mm) of the left hemisphere of each animal. In both animals, the penetrations were concentrated near the center of this region, where form selective neurons were more reliably found. Using electrolytic lesions and fluorescent dye (DiI, Molecular Probes) to coat the electrode (DiCarlo et al. 1996), we confirmed that the bulk of the recordings from the first animal were on the ventral surface, centered ~ 10.5 mm posterior of the temporal pole, lateral of the anterior middle temporal sulcus (AMTS). Based on the anterior-posterior coordinates, and the sulci, this region is approximately the anterior third of IT and is contained in area TE (Felleman and Van Essen 1991; Logothetis and Pauls 1995; Logothetis and Sheinberg 1996). We refer to this region as AIT (Felleman and Van Essen 1991).

The animal cycled through behavioral blocks as the electrode was advanced into AIT. Responses from every isolated neuron were assessed with an audio monitor and on-line histograms, and data were collected from even marginally responsive cells under the assumption that longer periods of observation might reveal statistically detectable effects. Data from each recorded neuron were considered for further analysis if isolation was maintained for at least six presentations (mean = 8.5, maximum = 10) of each target form in each position during all task conditions (~ 20 – 35 min of recording). The responses of 220 AIT neurons (*monkey 1* = 119, *monkey 2* = 101) were recorded. Among these, 74 (33%) were not considered for further analysis because they failed to produce a statistically significant response to any of the three tested retinal positions (described in the following text). The presence of these 74 unresponsive neurons in the recorded data set is consistent with our low threshold for selecting neurons during the recording sessions. Most of the neurons were located on the ventral surface (127 of 146; 87%); the rest were in the ventral bank of the STS. For brevity, the data from both animals were combined in some plots, and summary values for each animal are indicated in the text and figure legends.

Analysis

Only neuronal responses collected during correctly completed behavioral trials were included in the analyses (88% of trials; except Fig. 8, see METHODS). We also excluded trials in which eye movements $> 0.3^\circ$ occurred during the first 50 ms after target onset ($< 1\%$ of all correct trials) or those in which the animal began its response saccade < 100 ms after target onset ($\ll 1\%$ of all correct trials). We estimated the background firing rate of each neuron as the mean rate of firing over all trials in a 100-ms-duration window that directly preceded target onset. For the majority of the data (where only 3 positions were tested), we quantified the response of each neuron to each of the 12 stimulus conditions (4 forms \times 3 positions) as the mean response in a 150-ms window that began 100 ms after target onset. One advantage of the behavioral task is that the choice of the temporal analysis window was constrained by both the start of the AIT responses (~ 100 ms after stimulus onset, see Fig. 13) (see also Baylis et al. 1987; Vogels and Orban 1994) and by the animal's reaction times (~ 300 ms after stimulus onset, see Fig. 2B). The results were largely unaffected by the details of the analysis time window (see RESULTS).

The mean response above background for each of the 12 stimulus conditions (4 target forms \times 3 positions) was used to determine the form and position preferences of each neuron. Eight neurons that showed decreases in firing rates in all 12 conditions were excluded from further analyses. We defined the neuron's best and worst target forms as those that produced the largest and smallest mean response over all three positions. Likewise, we defined the neuron's best and worst positions as those that produced the largest and smallest mean response over all four target forms. Responsive neurons ($n = 146$ of 220) were defined as those that showed a statistically significant increase in firing rate (relative to background rate) to their best target

form presented in any of the three positions (3 *t*-test, each run at $P = 0.017$). Because we selected the neuron's best target before running these tests, Monte Carlo simulation shows this gives an overall false positive level of 0.075. The main result (Fig. 6) was unaffected when false positive levels of 0.05 ($n = 140$), 0.01 ($n = 128$), and 0.001 ($n = 101$) were applied.

In Fig. 6, we used the RF data of Op de Beeck et al. (Op de Beeck and Vogels 2000) to predict the expected neuronal sensitivity to our tested positions. That report is the most quantitative study of IT RFs currently available. It showed that Gaussian sensitivity profiles fit most of the measured IT RFs, and it provided the distribution of RF sizes and RF centers. Based on those data, we simulated the position sensitivity of 10,000 randomly selected (normal), circularly symmetric Gaussian RFs using the following parameters: mean RF size (square root of RF area) = 10.3° ; RF size SD = 5° ; min RF size = 2° ; mean RF center azimuth = 1.5° (contralateral), mean RF center elevation = 0.0° ; RF center SD = 1.5° (azimuth and elevation).

RESULTS

Two monkeys were trained to identify four target forms by making a saccade to one of four fixed locations (Fig. 1). Each target form was presented to the fixating animal at one of three retinal positions on the horizontal meridian (center of gaze, 1.5° left of center, and 1.5° right of center). Both animals were highly accurate at this task (Fig. 2A). Accuracy was best at the central position (*monkey 1* = 94% correct, *monkey 2* = 88% correct) and only slightly reduced at the eccentric positions

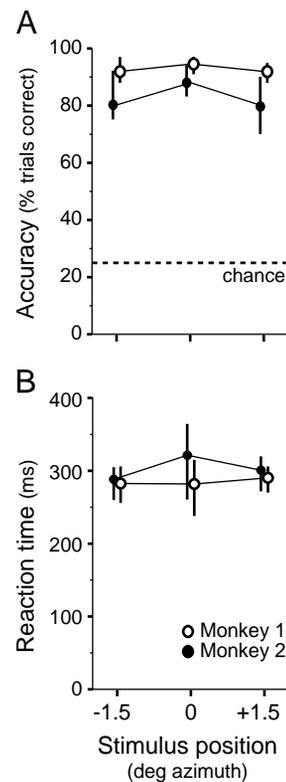


FIG. 2. Behavioral performance over changes in target position. A: mean accuracy for each animal at each retinal position. \circ , data from *monkey 1*; \bullet , data from *monkey 2*. Bars indicate the upper and lower quartiles of accuracy across neuronal recording runs ($n = 119$ for *monkey 1*; $n = 101$ for *monkey 2*). Because 4 target forms were used, the accuracy that would occur by guessing is 25% (dashed line). B: mean reaction time for each animal at each retinal position. Bars indicate the upper and lower quartiles of reaction time across trials.

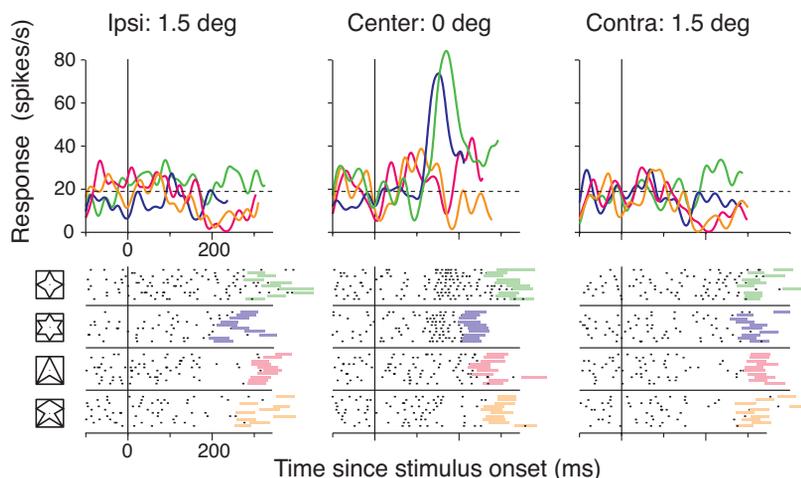


FIG. 3. Response of an anterior inferotemporal cortex (AIT) neuron to each target form in each retinal position. Columns show data from each position. The abscissa represents time since stimulus onset. *Top*: the ordinate represents the firing rate of the neuron. The response to each of the 4 target forms is indicated by a different color (for target form mapping, see the colored bands in the rasters below). Each response curve is the average of ten trials (*bottom*), smoothed with a Gaussian filter (10 ms SD). The horizontal dashed line indicates the background firing rate of the neuron. Lower panels: Each row is data from a separate trial. Each tick mark indicates a single action potential. Target forms are indicated by the icons at the left, and are ordered from “best” to “worst” for this neuron. The colored bands indicate the duration of the animal’s response saccades. This neuron is typical in that it showed comparable sensitivity to target form and target position.

(*monkey 1* = 3% decrease in accuracy; *monkey 2* = 8%). Mean reaction times were short in both animals (*monkey 1* = 285 ms; *monkey 2* = 303 ms) and were little affected by position (Fig. 2B). Although these behavioral effects of position were small, most were statistically significant because of the large number of behavioral trials examined (~4,500 trials for each animal in each position; accuracy: *monkey 1*: $\chi^2 = 3.0$, $P > 0.05$; *monkey 2*: $\chi^2 = 22.2$, $P < 0.01$, $df = 2$; reaction time: *monkey 1*: $F = 53$, $P < 0.01$; *monkey 2*: $F = 287$, $P < 0.01$). In sum, the behavior showed excellent position tolerance—both animals could rapidly and accurately identify each target form, regardless of its position, and without foreknowledge of precisely where it would appear.

If individual AIT neurons were underlying the animal’s recognition, the behavioral observations suggested that these neuronal responses should be largely unaffected by these small position changes. Likewise, previous studies showing AIT RFs to be 10° or more in diameter (see INTRODUCTION) also predicted that the neuronal responses should be largely unaffected by our small position changes. To examine these predictions, we analyzed data from all 146 recorded neurons that were responsive in at least one position (72 from *monkey 1*, 74 from *monkey 2*; see METHODS). Consistent with previous studies (Logothetis and Sheinberg 1996; Miyashita 1993; Tanaka 1996), many of the recorded neurons were selective for stimulus form ($n = 54$ of 146, see later). However, the AIT neuronal responses in our animals were largely inconsistent

with the large RFs previously reported in AIT (see INTRODUCTION). In particular, almost all neurons showed a stronger than expected sensitivity to small (1.5°) position changes, and some were exquisitely sensitive to these position changes. Responses from one such neuron are shown in Fig. 3. *Middle* shows that when targets were presented at the center of gaze, the neuron responded strongly to two of the target forms but gave little response to the other two. That is, this neuron was highly form selective at the center of gaze (ANOVA, $P < 10^{-7}$). However, the neuron produced almost no response when the same target forms appeared either 1.5° ipsilateral or 1.5° contralateral to the center of gaze. Thus this neuron was selective for stimulus form but responded only over a very limited range of stimulus positions (assuming that positions more eccentric than the tested three would yield little or no response, see following text). It should be emphasized that all three tested retinal positions were within the fovea ($\pm 2^\circ$). One interpretation of these observations is that the neuron had a very small RF near the center of gaze (i.e., $< 2^\circ$ in diameter). However, because we did not perform full RF mapping for most neurons and because some neurons showed more than one hot spot in their RF (e.g., Fig. 4), we use the term position sensitivity to describe the effect of our tested position changes on the neuronal responses.

The neuron in Fig. 3 could contribute to form discrimination at the central fovea, but it is poorly suited for the eccentric positions just 1.5° away. However, the animals were highly accurate at identifying target forms at all three retinal positions.

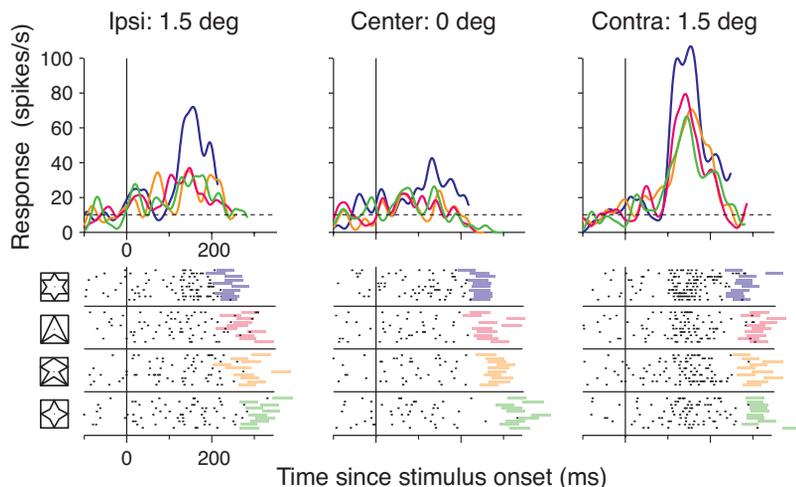


FIG. 4. Response of an AIT neuron that preferred target forms at eccentric positions. Format is described in Fig. 3.

If AIT supported recognition at all three positions, one would expect to find neurons that showed form selectivity at eccentric positions. Indeed, we also encountered many neurons that preferred stimuli at one or both of the eccentric locations. For example, the response pattern of the neuron shown in Fig. 4 was complementary to that of the previous neuron in that it was most responsive to stimuli presented in the contralateral position, with some response in the ipsilateral position, and almost no response in the central position.

In light of previous studies, the observation that AIT neuronal responses change with stimulus position is not surprising. Indeed, any neuron must show some position sensitivity—at least at the edges of its RF. However, the neuronal position sensitivity was typically much larger than that previously reported or expected based on reported RF sizes in AIT. Indeed, many neuronal responses were so strongly affected by retinal position that they failed to respond at one or two of the three tested locations (all were within the fovea). Among the neurons that were responsive in at least one location, 77 (52%) gave no statistically significant response for one or both of the remaining positions (*t*-test), and 18 of these gave no statistically significant response to the central fovea (using the best target form for all tests). This was not due to the neurons being poorly responsive overall because the mean driven response rate at preferred positions was 24.3 spikes/s ($n = 146$)—comparable to rates previously reported in AIT (20–40 spikes/s) (Leuschow et al. 1994; Missal et al. 1999; Op de Beeck and Vogels 2000). The examples in Fig. 5 illustrate the range of position and form sensitivities seen in the recorded population.

To summarize the position sensitivity of each neuron, we plotted its reduction in response when its best target form was presented in its worst position (relative to the response in its best position; Fig. 6). The median relative response was 0.41. In other words, the response of the typical AIT neuron in our sample could be reduced by $\sim 60\%$ when the neuron's preferred stimulus form was moved within a region of only $\pm 1.5^\circ$

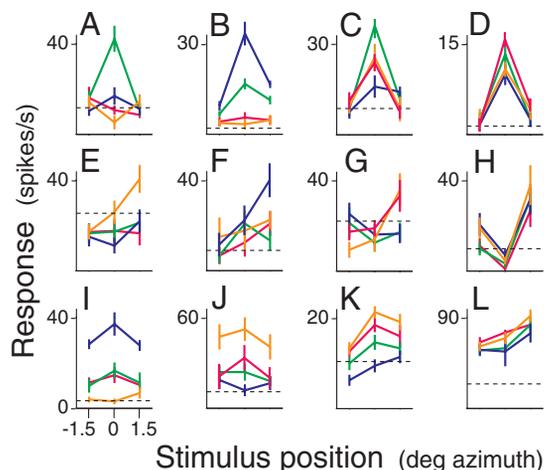


FIG. 5. Responses to each stimulus condition (4 target forms \times 3 positions) from 12 representative AIT neurons. For each panel, abscissa represents retinal position; ordinate represents mean firing rate in the analysis time window (100–250 ms after stimulus onset). Colors indicate responses to different target forms. ---, each neuron's background firing rate. We did not observe obvious grouping of neurons with particular patterns of position and form selectivities but instead a continuum of properties. *Top*: preferred target forms at the central position; *middle*: preferred target forms at the contralateral position; *bottom*: forms in all 3 positions. *Left*: strong form sensitivity; *right*: weak form sensitivity. Error bars indicate SEs.

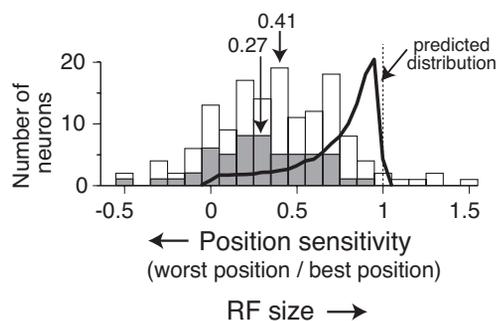


FIG. 6. Distribution of position sensitivity. The abscissa indicates the ratio of the mean response to the worst position and the mean response to the best position (using the best target form). A value of one indicates no position sensitivity; a value of 0 (or <0) indicates that the neuron's response at the worst position is at (or below) background. The ordinate is the number of neurons showing the position sensitivity specified on the abscissa. ▣, form selective neurons ($n = 54$, see RESULTS). The median position sensitivity index was 0.41 (*monkey 1* = 0.34; *monkey 2* = 0.46) for the 146 responsive neurons and 0.27 for form selective neurons. For 108 (of 146) the best and the worst positions were 1.5° apart; for the remaining neurons they were 3° apart. Of the 146 neurons, 86 preferred the center position, 55 preferred the contralateral position and 5 preferred the ipsilateral position. The thick, solid line shows the distribution of position sensitivities predicted from AIT RF data (median predicted position sensitivity = 0.82; see METHODS).

around the center of gaze. If we only consider neurons that prefer the center of gaze (i.e., where we clearly included the RF center), assume 2D Gaussian shaped RFs, and define RF cut-off at 50% (as in previous studies, see Op de Beeck and Vogels 2000), then this median decrease over a position change of 1.5° corresponds to a median RF diameter of 2.6° . This is not an artifact of noisy responses—the result was nearly identical when the data were split in half and one group was used to compute the best and worst targets and positions and the other group used to compute the position sensitivity.

Because form-selective neurons are most likely to underlie the recognition behavior, it is possible that they have less position sensitivity (because the behavior showed virtually no position sensitivity). However, examination of the 54 neurons (37%) that were selective for stimulus form (ANOVA, $P < 0.05$) revealed even greater position sensitivity (median = 0.27) than that seen in the entire responsive population (Fig. 6). Under the RF assumptions described above, this corresponds to a median RF diameter of 2.2° .

To compare the distribution of position sensitivities of the recorded population (Fig. 6) with that predicted from previous studies, we estimated the expected AIT position sensitivity using the RF data from a recent, thorough study of AIT RFs (Op de Beeck and Vogels 2000) (see METHODS). That data predict that the median AIT neuron should have shown only an 18% maximal response change across our three tested positions, nearly fourfold less than we observed.

The stronger than expected position sensitivity could be due to changes in overall responsiveness at some retinal positions (e.g., due to small RFs), changes in form preference at each retinal position, or both. The example neuronal data (Figs. 3–5) suggest the former hypothesis. This hypothesis also seemed most likely because previous studies have reported that the rank order of form preference is largely unaltered by changes in position (e.g., Desimone et al. 1984; Ito et al. 1995; Sary et al. 1993; Schwartz et al. 1983). However, because we found much greater position sensitivity than previous studies, we

sought to confirm that it acted across all stimulus forms. Because the position sensitivity of the neuronal responses was so strong, we could not test this hypothesis for about half the neurons because the 1.5° position shifts eliminated the response (e.g., Fig. 3). Even when responses remained at non-preferred positions, they were so weak that most neurons were no longer significantly form selective at those positions. Specifically, 54 of the 146 responsive neurons (37%) were significantly form selective at their best position but less than half of these (25 of 54) were still significantly form selective at their second best position. Nevertheless, 24 of these 25 neurons maintained the rank order of their best and worst forms at their second best position.

To summarize the average effect of position changes on form selectivity, we split the 54 form-selective neurons into three groups, where each group preferred one of the three tested positions ($n = 2$, $n = 35$, $n = 17$ for the ipsi, central, and contra positions). We then rank-ordered the target forms for each neuron and averaged the normalized (to best response) responses of all neurons in the group for each rank-ordered form in each position (Fig. 7). This analysis showed that, on average, neurons that preferred the central position (Fig. 7, left) maintained their rank order of form preferences at the eccentric positions and showed a strong response reduction in each side position that operated largely as a decrease in response gain over all four target forms (gain of ~ 0.4 across the 1.5° position changes). Results were similar for neurons that preferred the contralateral position, but the decrease in response gain was slightly weaker (Fig. 7, right). In summary, although we found much greater position sensitivity than most previous studies, the results were consistent with other studies in that, when it could be measured, the rank order of target form preference was largely unaffected by position. Thus the strong position sensitivity observed in this study is most consistent with the hypothesis that the neurons have small RFs ($\sim 2.5^\circ$ diam), or that those RFs contain unresponsive locations (e.g., Fig. 4).

We could not fully characterize the spatial RFs of the neurons because we tested only three positions. Because the animal's task was to identify forms at these positions, our logic was that the position sensitivity of AIT neurons responding to

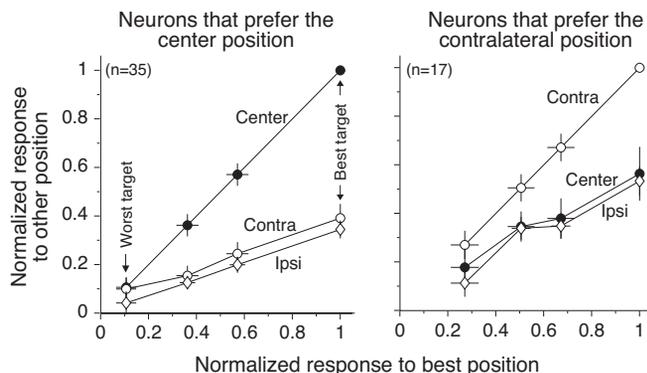


FIG. 7. Average effect of stimulus position on form selectivity. Left: the abscissa represents the normalized response to forms presented in the best position. The ordinate represents the normalized response to forms presented in each of the 3 tested positions (●, central position; ○, contralateral position; ◇, ipsilateral position). Each data point is the mean response of the population of form selective neurons that preferred the central (left) and contralateral (right) positions. Before averaging, each neuron's target form preferences were rank-ordered from best to worst and its response to each of those target forms was normalized by its response to its best target form in its best position.

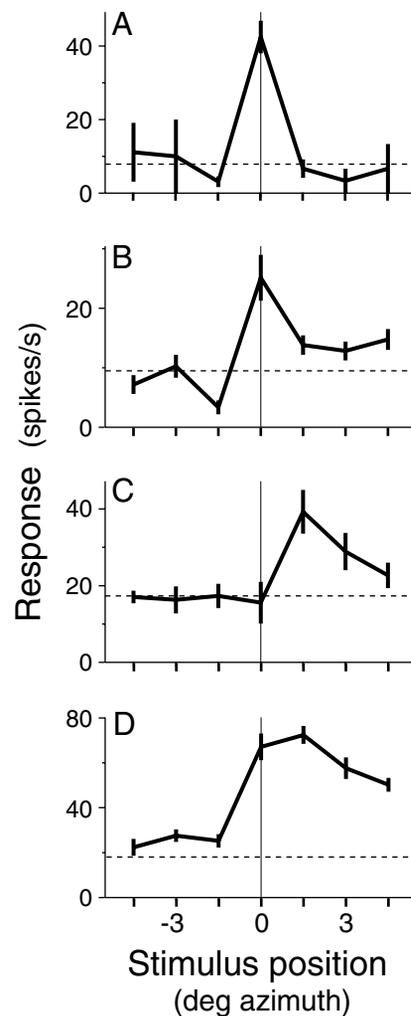


FIG. 8. Extended mapping of position sensitivity. Each panel shows data from 1 neuron. For each panel, the abscissa indicates the horizontal retinal eccentricity of the neuron's best target form (deg azimuth along the horizontal meridian); the ordinate indicates the mean response rate 100–250 ms after stimulus onset. Error bars show the SE. The dashed line is the background rate (see METHODS).

any of these positions would provide the most appropriate measurement of the position sensitivity of AIT neurons that might support the behavior. Exploration of additional retinal positions could only show that we had underestimated the neuronal position sensitivity. However, we wondered if our measurements were on the edge of some RFs or if they always included the RF center (i.e., maximal response position). Although a thorough exploration of these RF issues is the focus of future studies, we have collected preliminary data from 17 responsive neurons in one animal (*monkey 1*). For these neurons we extended our measure of position sensitivity along the horizontal meridian by placing stimuli at four additional positions eccentric to those tested for the larger neuronal population. In particular, we tested horizontal eccentricities of -4.5 to $+4.5^\circ$ in 1.5° increments (Fig. 8). Although the animal performed well above chance the first day it saw these new positions, the animal received additional training to better acclimate it to the occurrence of targets at these new positions (see METHODS). After training, the animal's performance at these positions was reduced relative to the more central posi-

tions, but was well above chance (70 and 62% correct at 3.0 and 4.5° eccentricity, respectively). Each neuron's preferred target form was determined from the central three positions as before, and the response to that target plotted as a function of position. Of the 17 neurons tested, no neuron gave a significantly larger mean response to any of the more eccentric positions than it did to the best of the original, central three positions (t -test, $P = 0.05$). Data from four representative neurons are shown in Fig. 8. Thus although the RF shape varied from neuron to neuron, the extended field mapping suggests that the RF centers of the tested neurons were within the original three positions.

Time course of position sensitivity

We next sought to determine if the position sensitivity was present in the earliest part of the responses or if it developed over time. For example, perhaps the AIT neurons had different response latencies for different positions. Inspection of the data revealed little evidence of large differences in latency across stimulus position (e.g., Fig. 4), but we examined the time course for subtle effects. As a first step, we re-analyzed the entire data set using two other analysis windows (100–200 and 150–250 ms after stimulus onset) with little effect on any of the results. The median position sensitivity ratios using these time windows were similar (0.36 and 0.38, respectively; cf. Fig. 6). An ideal analysis would estimate each neuron's response latency for each position, but this is problematic because of the limited number of trials and because many neurons did not respond to nonpreferred positions. Instead we estimated the population time course of the position sensitivity by computing the population average response to each neuron's best target form presented in the neuron's best and worst positions (Fig. 9). For the best position, AIT neurons began to respond ~100 ms after stimulus onset. For the worst position, the average response began slightly later, rose more slowly, and

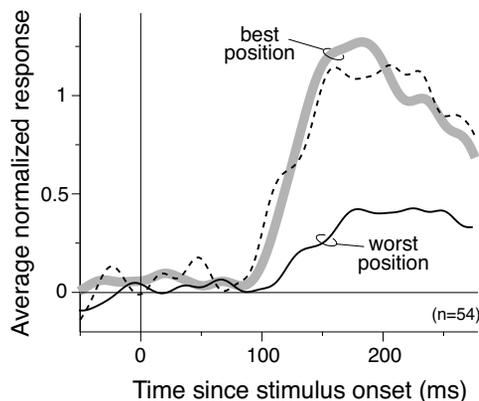


FIG. 9. Time course of position sensitivity. The abscissa represents time since stimulus onset. The ordinate represents the average driven response for the population of form selective AIT neurons. Before averaging, each neuron's response was normalized by its response to its best target form in its best position (as in Fig. 7). The thick gray line indicates the population average response to the best target form in the best position; the thin line indicates the response to the best target form in the worst position. The data were binned at 1 ms and Gaussian filtered (10 ms SD). The dashed line is a scaled (2.7 times) and temporally shifted (-19 ms) version of the response to the worst position that best fits the response to the best position (see RESULTS). Fifty-four form-selective neurons were included in the average. The plots were nearly identical when all 146 neurons were included.

reached a lower peak. The plot suggests that latency differences across stimulus position account for only a small amount of the position sensitivity reported above. To quantify this, we found the temporal shift and scale factor that could be applied to the average response in the worst position to best match the average response in the best position (RMS error function). The fit was good (correlation coef = 0.976, 0–300 ms after stimulus onset; dashed line in Fig. 9), and it required a temporal shift of 19 ms and a vertical scale factor of 2.7. The scale factor is an estimate of the amount of position sensitivity not due to latency differences, and it shows that mean position sensitivity (worst/best position) was 0.37 (i.e., $1/2.7$), which is comparable to the median effect of 0.41 already described. In summary, changes in response gain with position underlie almost all of the position sensitivity reported in this study.

Possible artifacts

Because we found much greater position sensitivity than almost all previous studies of AIT (but see DISCUSSION), we considered factors that might explain this finding. The most intriguing possibilities require further systematic study (see DISCUSSION). However, here we report our examination of three possible artifacts that might have contributed to our findings: stimulus spatial frequency content, differences in eye movements across position, and differences in stimulus duration across position.

The first factor we considered was the spatial frequency composition of the target forms. The target forms were made of line segments with a high spatial frequency content (~ 25 cycles/°, see METHODS). Because stimulus form (identity) depended on the spatial arrangement of these line segments, the spatial frequencies that supported the animal's differentiation of the forms were much lower (~ 5 cycles/°)—near the maximal contrast sensitivity for primates (Merigan and Maunsell 1993). Indeed, the stimuli had spatial frequency content similar to that of individual letters during normal reading. Nevertheless, we considered the possibility that the spatial frequency content of the stimulus elements was responsible for the strong position sensitivity. We created a set of four new targets that had the same size and spatial layout as the original four targets, but whose line segments contained lower spatial frequencies (Fig. 10). One of the animals (*monkey 1*) was retrained to respond to these four modified targets using the same form-response mapping as the four original targets even when both target types were randomly interleaved across trials (~ 1 wk of training). We recorded the responses of an additional 15 AIT neurons to each of the eight targets in each of the three original positions. We measured position sensitivity for each spatial-frequency condition exactly as before with the exception that each neuron's best target and best and worst positions were chosen after averaging the data from the two spatial-frequency conditions (results were nearly identical when each condition was considered separately). The analysis showed that some neurons were less sensitive to the position of the modified stimuli (Fig. 10C) but that other neurons were equally (Fig. 10D) or more position sensitive (Fig. 10E). Over the population ($n = 15$), the median position sensitivity for the original stimuli was nearly identical to that measured in the larger group of neurons (0.37) and was not significantly different from the population position sensitivity measured with the

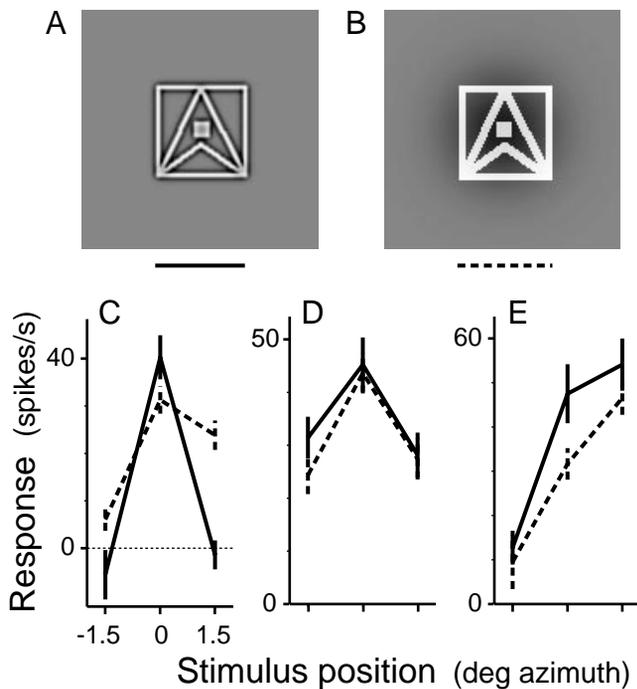


FIG. 10. Effect of stimulus spatial frequency content on position sensitivity. *Top*: one of the target forms in the 2 spatial-frequency conditions: original condition (A), used throughout the study, and modified condition (B). These images are only approximate reproductions of the stimuli used in the task (see METHODS). *Bottom*: the position sensitivity data (driven response to the best target in each of the 3 positions) from 3 representative neurons in the 2 spatial frequency conditions (—, original condition; - - -, modified condition).

modified stimuli (median = 0.33; t -test, $P = 0.60$). Thus these data suggest that the strong position sensitivity cannot be simply explained by the spatial-frequency content of the stimulus elements per se (but see DISCUSSION).

The second and third potential artifacts we considered were differences in eye movements and differences in stimulus duration across target position. As described in METHODS, we did not place strong constraints on the animal's eye movements but ensured that the target was only presented at the intended retinal position. Because of this, the animal's pattern of eye movement and the stimulus duration were both confounded with the primary variable of retinal position. These confounds are illustrated in Fig. 11, A–C. We admitted these confounds in our design because we wanted the task to remain as natural as possible while still varying the retinal position of the target forms. As a result, it is possible that the shorter stimulus exposure durations used for eccentric stimuli (~150 ms) relative to the central stimuli (~300 ms) could affect response rate and cause apparent strong position sensitivity. This seemed unlikely because rapid presentation of stimuli indicates little peak response reduction for stimulus exposure durations greater than ~50 ms (Keyser et al. 2001) and because the latency of AIT neurons to stimulus onset is ~100 ms (Fig. 9) (Baylis et al. 1987; DiCarlo and Maunsell 2000; Vogels and Orban 1994). If stimulus offset requires the same latency as stimulus onset to alter AIT firing rates, then the offset of the target form would not alter the response until the end of the analysis window (i.e., 100 ms after the form offset is ~250 ms). A second possibility is that the neuronal processes that

produce eye movements toward the target ("adjustment saccades" in Fig. 11; see METHODS) could cause a change in ongoing AIT neuronal activity (e.g., a "reset" signal or saccadic suppression). The fact that the monkeys' reaction times were nearly identical for central and eccentric stimulus positions argues against this possibility (Fig. 2) but does not exclude it. Because the two confounding factors (stimulus exposure duration and time of adjustment saccade) were perfectly correlated in our design, we cannot distinguish their effects, so we considered them to be a single confound and performed analyses to isolate the effect of this confound from that of stimulus position.

One analysis is summarized in Fig. 11 (D and E). Each point in each panel is the response rate of one neuron on one trial relative to the average response rate of this neuron over all trials with the neuron's best form in one position. These normalized trial-by-trial responses are plotted relative to the time that the adjustment saccade (i.e., the confound) occurred for that trial. Thus these plots show the average effect of the confound on response rate (isolated from the effect of stimulus position). If the confound had a consistent effect across the population of AIT neurons (e.g., decrease in ongoing neuronal responses), the running averages in the plots should show a trend. Instead, no trends were apparent and the correlation coefficients were not significantly different from zero (–0.012, –0.030, $P > 0.1$). The two symbol types in the plots indicate data from the two monkeys, illustrating that *monkey 2* tended to make adjustment saccades at shorter latencies than *monkey 1*. This difference in behavior does not obscure a relationship between the time of the adjustment saccade and response rate because the within-animal correlations are also not significantly different from zero (*monkey 1*: –0.051, –0.021; *monkey 2*: 0.013, –0.043; $P > 0.1$ all cases). In addition, the mean of the normalized responses on trials where no adjustment saccade occurred was not significantly different from that expected based on trials where an adjustment saccade was made (t -test against a value of 1, $P > 0.1$ for the ipsilateral and contralateral conditions). If the confound causes some neurons to increase their firing rates and others to decrease, the analysis in Fig. 11 might fail to detect these effects. However, a neuron-by-neuron analysis revealed that only ~5% of neurons (8% for ipsilateral stimuli, 3% for contralateral stimuli) showed any significant correlation of response rate with adjustment saccade latency (Spearman ranked correlation, $P < 0.05$), which is approximately the number expected by chance. Furthermore, a mixture of positive and negative effects should increase the variability of relative response rates (i.e., the SD of the ordinate values in Fig. 11) relative to that which would have been observed without the effects. Instead, the observed SDs (ipsi: 0.50, contra: 0.47) were slightly below those obtained from simulated trial-by-trial responses using the average rates observed in the actual population and Poisson firing statistics (ipsi: 0.53, contra: 0.50) (see Shadlen and Newsome 1994 for Poisson assumption; Softky and Koch 1993). In summary, because these analyses failed to find a significant effect of the time of the adjustment saccade (and stimulus offset) on the response rate, we conclude that these factors did not significantly modify the AIT responses and thus they cannot explain the position sensitivity of those responses.

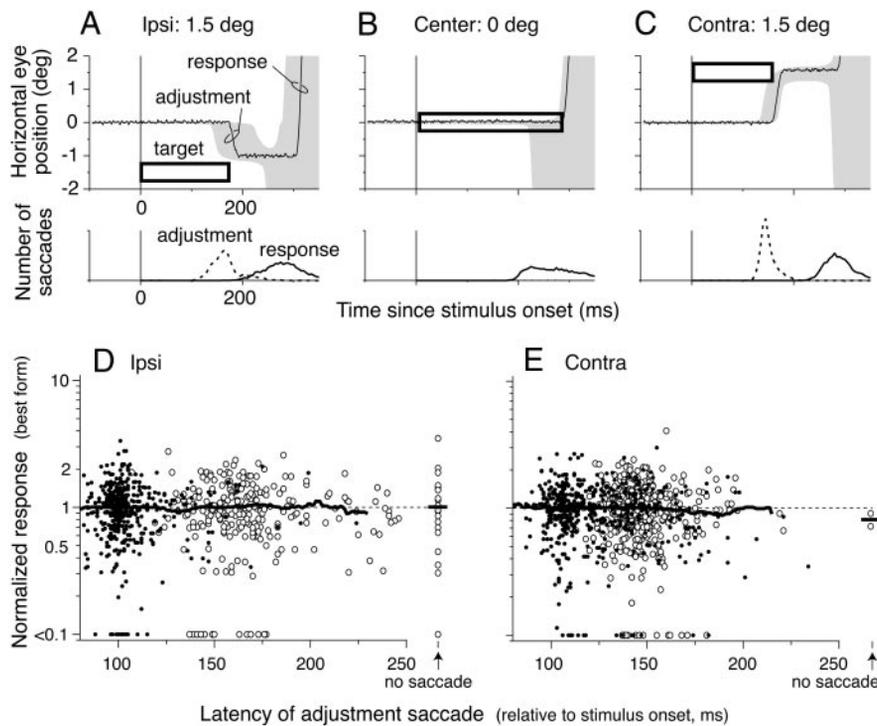


FIG. 11. Effect of eye movements and stimulus exposure duration. *A–C*: gaze behavior for presentation at each of the 3 retinal positions. *Top*: a typical eye position trace obtained from a single trial (—) and the range of eye traces across all trials (□). At each time point, □ contains 75% of the eye position values. For the 2 eccentric retinal positions, the animal typically made a saccade toward the target (labeled “adjustment”), and the target was removed from the display during this saccade (see METHODS). □, the spatial and temporal extent of the target form during the 3 example trials. Response saccades began ~300 ms after stimulus onset and ended near one of the response locations (see Fig. 1). *Middle*: the temporal distributions of adjustment saccade onsets (---) and response saccade onsets (i.e., reaction times; —). Approximately 4,300 trials contributed to each of the plots. The ends of the plots cut off 4, 10, and 3% of the response saccade distribution for the ipsi, central, and contra positions, respectively. Data are from *monkey 1*. *D* and *E*: effect of adjustment saccade and stimulus offset on responses. In both panels, the abscissa represents the latency of the start of the adjustment saccade, relative to stimulus onset (e.g., see *A*). The ordinate represents the normalized response rate to the best target form presented in the specified position: ipsilateral position (*D*) and contralateral position (*E*). Each data point indicates the response rate obtained from a single trial, and the corresponding adjustment saccade latency for that trial (○, data from *monkey 1*; ●, data from *monkey 2*). Each single-trial response rate was normalized by the mean response rate to the best target form presented in the specified position (thus the 6–10 data points contributed from each neuron always have a mean value of 1 on the plot). The data from a neuron were included in each plot if the neuron showed a significant response to the best target form in the specified position [number of neurons = 82(*D*), 111(*E*); *t*-test against background rate, $P < 0.05$]. Points at the right side of each plot are data from trials in which no adjustment saccade was detected (i.e., the animal made a saccade directly from the central position to the correct response location). —, a running mean computed from points within ± 20 ms.

Behavioral significance of neuronal position sensitivity

Unlike almost all previous studies of AIT RFs or AIT position tolerance, the current data were collected while the subjects performed recognition across changes in object position. Thus we were also able to examine position sensitivity in the context of that behavior. Here we present three such analyses.

In the first analysis, we adopt a standard view of AIT in which the purported role of AIT neurons is to extract object identity and to support the “perceptual equivalence” of the same object over changes in, for example, object position (e.g., Desimone et al. 1984; Gross and Mishkin 1977). This hypothesis predicts that individual AIT neurons should be capable of signaling object identity across changes in object position that are “perceptually equivalent.” Testing this prediction depends on defining both perceptual equivalence and the manner in which AIT neurons signal or code object identity. The spirit of perceptual equivalence is that the subject’s interpretation of the identity of the object remains the same over changes in, for example, object position. The animal’s accurate identification

of each object across changes in position (even for less trained positions, see METHODS) suggests that it treats each object as equivalent across position. Thus we assume that AIT neurons should signal object identity across these same position changes. We defined an AIT neuron’s ability to signal object identity as its response to its best target form relative to a distractor response (d'). The distractor response was taken to be the maximal response to the neuron’s worst target form over all three positions. We then asked, how well does each neuron continue to signal its preferred object across the tested position changes?

The results from the 54 form-selective neurons are shown in Fig. 12. Almost all of these neurons provided a strong signal of target identity at their preferred position. In particular, 41 of the 54 neurons (76%) had d' values > 1.35 (discrimination performance of 75% correct) at their preferred position. However, only three of the neurons (6%) could continue to provide this target identity signal ($d' > 1.35$) at all three of the tested positions. Put another way, the typical form-selective neuron could correctly discriminate its best target from the distractor on 83% of the trials (median $d' = 1.89$), but a position change

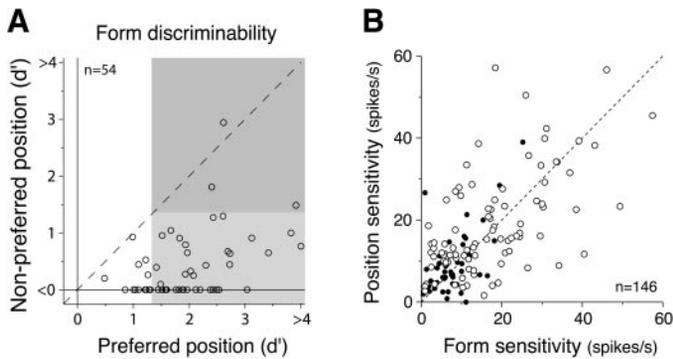


FIG. 12. *A*: the effect of a position change on form discriminability. Discriminability (d') is the difference between the response to the neuron's best target form and a distractor form (see RESULTS), normalized by the root mean square of the response SDs in each condition (Green and Swets 1966). The abscissa shows discriminability when the best target form is at the neuron's preferred (best) position. The ordinate shows discriminability at the neuron's least preferred (worst) position (of the 3 tested positions). Data from the 54 form-selective neurons are shown (ANOVA, see RESULTS). Neurons in the light gray region can only reliably signal their preferred target at their preferred position (>75% correct performance; $d' > 1.35$). Neurons in the dark gray region can reliably signal their preferred target at all 3 tested positions. *B*: comparison of position sensitivity and form sensitivity. Form sensitivity (abscissa) is the difference between the response to the best form and the worst form in the best position. Position sensitivity (ordinate) is the difference between the response to the best form in the best and the worst position. Each point represents the data from a single neuron ($n = 146$). Open circles ($n = 94$), neurons with a significant main effect of stimulus form, stimulus position, or an interaction of those effects (2-way ANOVA, $P < 0.05$).

within 1.5° of the fovea caused that same neuron's performance to fall to near chance (median $d' = 0.15$; 53% correct discrimination; 50% is chance). In sum, these data show that only a few AIT neurons are individually capable of mediating perceptual (behavioral) equivalence.

In the second analysis, we ask: were the AIT neurons better at signaling object identity or object position? Tovée et al. (1994) asked this question in passively fixating animals and showed that the median AIT neuron carried four times as much information about object identity as it did about object form. However, comparison of position sensitivity and form sensitivity is problematic because it depends on the tested range of objects and positions. The comparison is only meaningful in the context of a behavioral task. In particular, if the putative role of AIT neuronal responses is to inform the animal about object identity regardless of small changes in object position, then AIT responses must be more sensitive to an identity change that is critical to the animal's task than to a position change that is irrelevant in that task. Our behavioral task was specifically designed to test this hypothesis, because it required the animal to signal object identity (stimulus form) regardless of position.

We compared the position and form sensitivity of the population of AIT neurons. The median position sensitivity was 11.5 spikes/s ($n = 146$; best-worst position; *monkeys 1* and *2* = 13.1 and 10.9) and the median form sensitivity was 10.4 spikes/s (best-worst form; *monkeys 1* and *2* = 10.5 and 10.2). If we consider only the 94 neurons that showed a statistically significant effect of either identity or position or an interaction (2-way ANOVA, $P < 0.05$), the median sensitivity differences were 14.3 spikes/s (position) and 13.8 spikes/s (form) and the median sensitivity ratios were 3:1 (position) and 2.4:1 (form). In summary, the AIT neurons were slightly more sensitive to

differences in position within the fovea that were irrelevant to the task than they were to differences in target form that were critical to the task. These data cannot rule out the possibility that the object position information conveyed in the AIT responses is completely ignored by downstream brain areas. However, these data suggest that the role of AIT neurons is to provide the animal with a representation of *both* object identity and object position and that the representation of object position can be of much higher spatial resolution than previously appreciated.

So far we have focused on the idea that to perform position-tolerant recognition, the brain should seek large RFs and thus less neuronal position sensitivity. However, there may be competing behavioral demands for small RFs and thus more position sensitivity (i.e., as seen in this study). In this third analysis, we consider one of those behavioral demands—recognition in visual clutter. Before any recording began, both animals were successfully trained to recognize each target in each position even when the target form was flanked on both sides by a row of distractors (see Fig. 13*A* and METHODS; mean behavioral accuracy was 87% with clutter vs. 88% without clutter). We considered the hypothesis that small RFs (i.e., high position sensitivity) might have developed to protect each neuron's

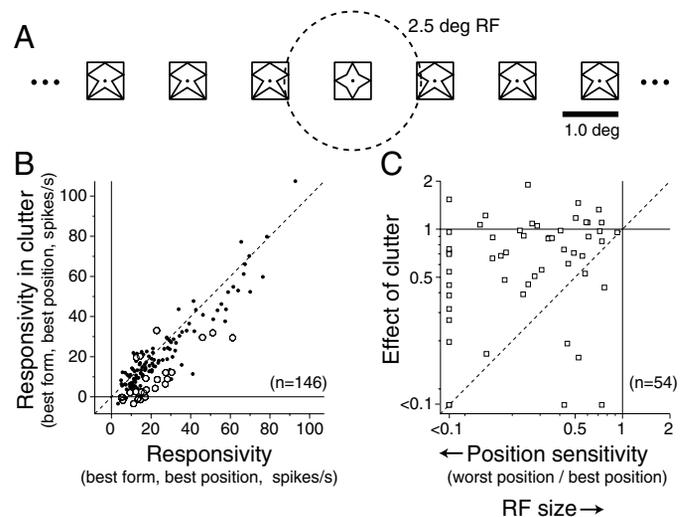


FIG. 13. Relationship of position sensitivity to visual clutter interference. *A*: schematic illustration of a target form embedded in the visual clutter. The clutter consisted of 20 identical distractor forms with 1.5° center-to-center spacing. ---, the RF size that would have produced the median observed position sensitivity (Fig. 6; see RESULTS). *B*: effect of visual clutter on responsivity. The abscissa represents the driven response when the neuron's best target form was presented in its best position. The ordinate represents the driven response when the same target form was presented in the same position but was embedded in a horizontal row of distractor forms (illustrated in *A*). \circ , neurons with responses that were significantly affected by the clutter ($n = 22$ of 146, paired *t*-test, $P < 0.05$). The median effect of clutter ($n = 146$) was a 23% response reduction (*monkey 1* = 25%; *monkey 2* = 22%). *C*: relationship between position sensitivity and the effect of clutter on form sensitivity. The abscissa represents an index of position sensitivity (response to the worst position/response to the best position); a value of 1 indicates no position sensitivity; values < 1 indicate increasing sensitivity to position. The ordinate represents an index of the effect of clutter on form sensitivity (form sensitivity in clutter/form sensitivity without clutter). Form sensitivity in each condition was defined as the response to best target form minus the response to worst target form, in the best position. An ordinate value of 1 indicates no effect of clutter on form sensitivity; values < 1 indicate increasing interference of clutter on form sensitivity. Data from the 54 form-selective neurons are shown (see RESULTS).

response, and thus the animal's behavior, from the influence of flanking visual clutter by limiting its intrusion into the RF.

We compared each neuron's responses at its best position with and without the flanking distractors. Consistent with the small RF hypothesis, addition of this flanking visual clutter only slightly decreased each neuron's response to its best target at its best position (median 23% decrease; Fig. 13*B*). Similarly, clutter had only modest effects on form selectivity. For 52 of the 54 (96%) form selective neurons, the response to the best target form remained above the response to the worst target form when clutter was added to the display, and clutter reduced median form selectivity by 24% (20.0–15.2 spikes/s; best-worst form). Because these effects of clutter are relatively mild, they are consistent with the small RF hypothesis. A more convincing test would ask if neuronal immunity to clutter is negatively correlated with RF size. However, when we took position sensitivity as an inverse measure of RF size, and form sensitivity in clutter as a measure of clutter immunity, we found no such relationship (Fig. 13*C*). We also observed little relationship between position sensitivity and responsivity in clutter (data not shown). In summary, this study suggests a relationship between position sensitivity and clutter immunity because it reports a much stronger effect of position than most previous studies and a weaker effect of clutter than most previous studies (Chelazzi et al. 1998; Miller et al. 1993; Missal et al. 1999; Rolls and Tovee 1995; Sato 1988). However, this relationship may not be simply explained by the hypothesis that these neurons have small RFs because the stronger position sensitivity was not associated on a neuron-by-neuron basis with improved immunity to clutter.

DISCUSSION

It is thought that the position tolerance of object recognition is supported by the large RFs of individual AIT neurons and their ability to maintain target preferences within those large RFs. Here we provide data relevant to that hypothesis by examining the effect of small differences in object position on recognition behavior and AIT neuronal responses. Behavioral accuracy and reaction times were largely unaffected by the differences in position. However, individual AIT responses were remarkably sensitive to position. The median AIT neuron showed ~60% decrease in response when stimuli were shifted within $\pm 1.5^\circ$ from the center of gaze, and 52% of neurons were unresponsive to one or two positions within this range. Although we did not systematically characterize the size of the AIT RFs, the position sensitivity would be explained by a median RF diameter of $\sim 2.5^\circ$. For comparison, a recent, systematic study (Op de Beeck and Vogels 2000) of AIT RFs estimated a mean diameter of $\sim 10^\circ$. Most studies have reported even larger RFs (e.g., 30° in diameter or more) (Bous-saoud et al. 1991; Desimone et al. 1984; Gross et al. 1969, 1972; Kobatake and Tanaka 1994; Richmond et al. 1983). Although we report much greater position sensitivity than previous studies, we do not refute or discount the results of those studies. Instead we believe that our observations point to several hypotheses whose exploration might unify previous observations and, in the process, provide a much deeper understanding of the tolerance properties of AIT neurons.

Consistent with previous studies, we found that many AIT neurons were highly sensitive to the form of visual stimuli

(reviewed by Logothetis and Sheinberg 1996; Miyashita 1993; Tanaka 1996). We considered the possibility that the strong position sensitivity of our recorded neurons was due to changes in form preferences across position. However, our findings were consistent with other reports (Desimone et al. 1984; Leuschow et al. 1994; Logothetis and Sheinberg 1996) in that the rank order of target preferences was largely maintained across responsive locations. In sum, the primary novel finding of this study is that AIT neurons can be highly sensitive to retinal position and thus appear to have much smaller RFs than previously reported. Control experiments and analyses revealed that this observation was largely unaffected by substantial changes in the spatial frequency content of the stimuli, was not an artifact of missing the RF center, and was not due to differences in eye movements or stimulus exposure duration at each position. Earlier studies did not use such small, precisely positioned stimuli and therefore would not have been able to measure position sensitivity at this spatial scale.

Like this study, several other studies have probed IT position tolerance by testing a few positions for changes in responsivity and selectivity. Ito et al. (1995) selected stimuli to optimize IT neuronal responses in anesthetized monkeys and then reported that a position change of 5° produced a $\sim 30\%$ response decrease. This is about sixfold less position sensitivity than reported here. Studies in awake, passively fixating animals typically show even less position sensitivity (Desimone et al. 1984; Gross et al. 1969, 1972; Kobatake and Tanaka 1994; Richmond et al. 1983; Tovee et al. 1994). For example, Tovee et al. (1994) showed that neurons responding best to face stimuli did not decrease their response rates $< 50\%$ of maximal until the center of the face was displaced by $> 15^\circ$. In monkeys performing a delayed match-to-sample task, Leuschow et al. (1994) showed that a position change of 5° produced only a $\sim 25\%$ decrease in response rate.

Notably, the data that most closely approach those reported here come from monkeys trained to recognize wire-frame objects (Logothetis et al. 1995). In that study, Logothetis and colleagues tested the position sensitivity of nine AIT neurons that were tuned for specific views of the wire-frame objects, while the animal maintained passive fixation. For three neurons, they reported that responses were largely insensitive to position differences of at least $\pm 2^\circ$ but less than $\pm 7.5^\circ$ (i.e., a RF size of $4\text{--}15^\circ$ in diameter) (Logothetis et al. 1995). When position tolerance was measured as the size of the region where the response to the best form remained above the responses of a large set of distractors, the typical tolerance region was found to be only $\sim 4^\circ$ in diameter (see Riesenhuber and Poggio 1999). This means that the RF size in that study was at least twice as large as that predicted by the current observations. Nevertheless, the results of Logothetis and colleagues may be the most consistent with our results because they also studied animals highly trained to recognize specific stimuli (see following text).

At least four possibilities could explain the unexpectedly strong position sensitivity of the AIT neurons in this study. First, the animal was actively performing a recognition task, whereas most previous studies of position effects in IT have been carried out in anesthetized or passively fixating animals (Desimone et al. 1984; Gross et al. 1969, 1972; Ito et al. 1995; Kobatake and Tanaka 1994; Logothetis et al. 1995; Op de Beeck and Vogels 2000; Richmond et al. 1983; Tovee et al. 1994). Second, by allowing the animal to respond as rapidly as

it liked, we obtained a short (150 ms), physiologically meaningful time window in which to analyze neuronal response rates. Previous studies generally averaged response rates over much longer, arbitrary periods of time. It seems unlikely that either of these possibilities account for the strong position effects reported here. This first possibility is unlikely because task effects in IT are typically weak (Vogels et al. 1995) and greater position tolerance has been observed in behaving animals (Leuschow et al. 1994; Logothetis et al. 1995). The second possibility is unlikely because initial response transients can dominate response rates computed over longer time windows in early visual areas (e.g., Muller et al. 2001) and perhaps IT (e.g., Logothetis et al. 1995) (see also Figs. 3 and 4).

A third possibility is stimulus size—we used stimuli that were much smaller (0.6° width) than those used in previous studies. Although the effect of stimulus size on AIT position tolerance has not been thoroughly studied, some data suggest that AIT neurons are somewhat less tolerant to position changes of small stimuli (Op de Beeck and Vogels 2000), and at least one computational model of recognition predicts that the position tolerance of AIT neurons will decrease with smaller stimuli (Gochin 1994). Furthermore, comparison across studies suggests that position tolerance is roughly proportional to stimulus size. For example, Tovee et al. (1994) employed the largest stimuli ($8.5\text{--}17^\circ$) of any study of IT position tolerance and also reported the most position tolerance (only a 50% response decrease over $>12^\circ$). Most studies used $\sim 5^\circ$ wide stimuli and found that IT response rates fell by $\sim 25\%$ over a 5° position change (Desimone et al. 1984; Ito et al. 1995; Leuschow et al. 1994; Missal et al. 1999). Using these results as a benchmark, a straightforward scaling of position tolerance with stimulus width predicts that our 0.6° wide stimuli should have caused response rates to drop by 62% with our 1.5° position change. Indeed, we found that the median neuron's response rate fell by $\sim 60\%$.

In the limit, AIT position sensitivity must depend on stimulus size. An AIT neuron (and the subject) can only be position tolerant within the sampling limits of the retina—it cannot respond selectively to stimuli that are not sampled by the retina with sufficient resolution to distinguish between them. As the retinal images of the stimuli to be recognized are made smaller (e.g., increasing viewing distance), this loss of discriminability should first occur at eccentric positions where retinal sampling density is lowest (Curcio et al. 1987; Perry and Cowey 1985). Thus even if AIT neurons were always maximally position tolerant, the tolerance region must be smaller for smaller stimuli. This could manifest itself in several ways. For instance, AIT RFs might shrink toward the fovea when measured with smaller and smaller stimuli. Alternatively, AIT neurons with large regions of position tolerance measured with large stimuli might simply not respond to small stimuli, as other AIT neurons begin to respond, albeit with less position tolerance (i.e., smaller RFs). However, this alternative may be inconsistent with data showing IT neurons to be largely size-invariant (Desimone et al. 1984; Ito et al. 1995; Logothetis et al. 1995; Sary et al. 1993; Schwartz et al. 1983). One of the goals of future studies is to address these issues by comparing both the size and position tolerance properties of AIT neurons with the tolerance properties of the recognition behavior they are thought to support.

Although retinal sampling sets an upper limit on behavioral

and AIT neuronal position tolerance, the actual limit may be imposed by further neuronal processing. If this neuronal processing is modified with experience, AIT position tolerance will not correspond to a single measurement, or to a measurement that scales simply with stimulus size, but instead it will be highly dependent on the experience of the observer with the objects to be discriminated. Thus a fourth possibility that might explain the strong position sensitivity found in this study is visual experience. In particular, our animals had extensive experience with four fixed target objects at the three highly controlled retinal positions. This visual experience might have caused some AIT neurons to become tuned to target forms at those specific positions. Indeed, this hypothesis is suggested by the profile of some of the AIT RFs (Fig. 8). Besides the position-specific experience, the experience with visual clutter may have also enhanced position sensitivity (see Fig. 13). Because the animals had approximately equal visual experience with each position during several months of recording, and we found that most neurons preferred the central position (Fig. 6), visual experience alone may not suffice to explain our results. Nevertheless, if the strong position sensitivity of the AIT neurons reported here depends at all on visual experience, these effects must be understood, because they bear directly on the mechanisms that underlie AIT position tolerance.

To our knowledge, no study has asked if AIT *position* sensitivity can be modified by visual experience, but some studies have touched on closely related issues. Several previous studies have shown that experience results in IT neurons that are tuned to the *form* of familiar stimuli (Kobatake et al. 1998; Logothetis et al. 1995; Miyashita 1988). For example, some AIT neurons are tuned to specific, trained views of familiar objects (Booth and Rolls 1998; Logothetis et al. 1995; Logothetis and Pauls 1995). Indeed, some studies suggest that AIT neurons become tuned to discriminate stimulus forms that are often encountered by the animal (Kobatake et al. 1998; Sigala et al. 2002; Young and Yamane 1992). At the level of behavior, a large literature has emerged describing perceptual learning tasks in which performance improvements on various types of visual discriminations (e.g., orientation discrimination) are specific to the trained retinal location but typically show inter-ocular transfer (e.g., Crist et al. 1997; Goldstone 1998; Schoups et al. 1995). These studies illustrate that recognition (i.e., stimulus form discrimination) is not always fully position-tolerant, even at equal eccentricities. Because IT RFs are generally thought to be too large to account for the position specificity of perceptual learning, it has been argued that the locus of plasticity must be in visual areas where RFs are small but typically binocular, such as V1 or V2 (e.g., Crist et al. 2001; Fahle 1994; Schoups et al. 1995). However, several recent monkey studies in which extensive training resulted in perceptual learning, found significant, but subtle changes in the response properties of V1 and V2 neurons (Crist et al. 2001; Ghose et al. 2002; Schoups et al. 2001). Because we found that AIT neurons can be sensitive to stimulus features (i.e., form) over small visual field regions, this raises the possibility that AIT plasticity could contribute to performance improvements that are specific to trained retinal positions. In other words, the position sensitivity reported here cautions that IT cortex should not be ruled out as a locus of plasticity underlying perceptual learning.

In the present study, we found that individual AIT re-

sponses were much more sensitive to target position than was the animal's behavior and that neurons preferred different locations within the trained region of the visual field. This suggests that the tiling of visual space by the RFs of form-selective AIT neurons produced the position-tolerant recognition behavior. If each of these tiles resulted from experience with a particular form at a particular retinal position, then position-tolerant recognition might depend on such experience over a range of positions (Dill and Fahle 1997; Hebb 1949; Nazir and O'Regan 1990). Although computational considerations argue that position tolerance could be achieved with built-in, general-purpose mechanisms that require experience with only a single retinal image of an object (e.g., Olshausen et al. 1993; Salinas and Abbott 1997; Ullman 1996; Vetter et al. 1995), this does not rule out the possibility that the brain has adopted an experience-dependent, "brute force" solution. Indeed, the idea of learned tolerance is not new, and it has been proposed to explain several types of tolerance at various cortical levels (e.g., Poggio 1990; Wallis and Rolls 1997), including the limited position tolerance of complex cells in V1 (Foldiak 1991).

The idea that position-tolerant recognition depends on visual experience at those positions seems at odds with the widely held belief that if we learn to recognize an object at one retinal position, recognition automatically transfers to other retinal positions. However, this expectation appears to rest not on a body of psychophysical data, but from introspection in everyday situations where recognition is assisted by both eye movements and extensive retinal experience with the objects we recognize (Nazir and O'Regan 1990). In fact, few psychophysical studies have examined the position tolerance of recognition. Although some suggest automatic position tolerance, (Biederman and Cooper 1991; Ellis et al. 1989), others, including the perceptual learning studies already mentioned, indicate that recognition is best at positions where the subject has the most experience (Dill and Edelman 2001; Dill and Fahle 1997, 1998; Foster and Kahn 1985; Nazir and O'Regan 1990). The likely resolution of these different results is that the role of experience in position tolerance depends on the stimuli to be discriminated (Dill and Edelman 2001).

Present address of J. J. DiCarlo, McGovern Institute for Brain Research, Dept. of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139

We thank C. Boudreau, E. Cook, G. Ghose, C. Hocker, and T. Yang for discussions on design, analysis and presentation, D. Murray and T. Williford for technical assistance, and T. Poggio, M. Riesenhuber, and S. Treue for helpful comments.

This work was supported by National Eye Institute Grant EY-05911. J.H.R. Maunsell is an investigator with the Howard Hughes Medical Institute.

REFERENCES

- Baylis GC, Rolls ET, and Leonard CM.** Functional subdivisions of the temporal lobe neocortex. *J Neurosci* 7: 330–342, 1987.
- Biederman I.** Recognition-by-components: a theory of human image understanding. *Psychol Rev* 94: 115–147, 1987.
- Biederman I and Cooper EE.** Evidence for complete translational and reflectional invariance in visual object priming. *Perception* 20: 585–593, 1991.
- Booth MCA and Rolls ET.** View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. *Cereb Cortex* 8: 510–523, 1998.
- Boussaoud D, Desimone R, and Ungerleider L.** Visual topography of area TEO in the macaque. *J Comp Neurol* 306: 554–575, 1991.
- Chelazzi L, Duncan J, Miller EK, and Desimone R.** Responses of neurons in inferior temporal cortex during memory-guided visual search. *J Neurophysiol* 80: 2918–2940, 1998.
- Crist RE, Kapadia MK, Westheimer G, and Gilbert CD.** Perceptual learning of spatial localization: specificity for orientation, position, and context. *J Neurophysiol* 78: 2889–2894, 1997.
- Crist RE, Li W, and Gilbert CD.** Learning to see: experience and attention in primary visual cortex. *Nat Neurosci* 4: 519–525, 2001.
- Curcio CA, Sloan KR Jr, Packer O, Hendrickson AE, and Kalina RE.** Distribution of cones in human and monkey retina: individual variability and radial asymmetry. *Science* 236: 579–582, 1987.
- Dean P.** Visual behavior in monkeys with inferotemporal lesions. In: *Analysis of Visual Behavior*, edited by Ingle D, Goodale M, and Mansfield J. Cambridge, MA: MIT Press, 1982, p. 587–627.
- Desimone R, Albright TD, Gross CG, and Bruce C.** Stimulus-selective properties of inferior temporal neurons in the macaque. *J Neurosci* 4: 2051–2062, 1984.
- DiCarlo JJ, Lane JW, Hsiao SS, and Johnson KO.** Marking microelectrode penetrations with fluorescent dyes. *J Neurosci Methods* 64: 75–81, 1996.
- DiCarlo JJ and Maunsell JHR.** Form representation in monkey inferotemporal cortex is virtually unaltered by free viewing. *Nat Neurosci* 3: 814–821, 2000.
- Dill M and Edelman S.** Imperfect invariance to object translation in the discrimination of complex shapes. *Perception* 30: 707–724, 2001.
- Dill M and Fahle M.** The role of visual field position in pattern-discrimination learning. *Proc R Soc Lond B Biol Sci* 264: 1031–1036, 1997.
- Dill M and Fahle M.** Limited translation invariance of human pattern recognition. *Percept Psychophys* 60: 65–81, 1998.
- Edelman S.** *Representation and Recognition in Vision*. Cambridge, MA: MIT Press, 1999.
- Ellis R, Allport DA, Humphreys GW, and Collis J.** Varieties of object constancy. *Q J Exp Psychol A* 41: 775–796, 1989.
- Fahle M.** Human pattern recognition: parallel processing and perceptual learning. *Perception* 23: 411–427, 1994.
- Felleman DJ and Van Essen DC.** Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex* 1: 1–47, 1991.
- Foldiak P.** Learning invariance from transformation sequences. *Neural Comp* 3: 194–200, 1991.
- Foster DH and Kahn JI.** Internal representations and operations in the visual comparison of transformed patterns: effects of pattern point-inversion, position symmetry, and separation. *Biol Cybern* 51: 305–312, 1985.
- Ghose GM, Yang T, and Maunsell JHR.** Physiological correlates of perceptual learning in monkey V1 and V2. *J Neurophysiol* 87: 1867–1888, 2002.
- Gochin PM.** Properties of simulated neurons from a model of primate inferior temporal cortex. *Cereb Cortex* 4: 532–543, 1994.
- Goldstone RL.** Perceptual learning. *Annu Rev Psychol* 49: 585–612, 1998.
- Green DM and Swets JA.** *Signal Detection Theory and Psychophysics*. New York: Wiley, 1966.
- Gross CG.** Visual functions of inferotemporal cortex. In: *Handbook of Sensory Physiology*, Berlin: Springer-Verlag, 1973, p. 461–482.
- Gross CG, Bender DB, and Rocha-Miranda CE.** Visual receptive fields of neurons in inferotemporal cortex of the monkey. *Science* 166: 1303–1307, 1969.
- Gross CG and Mishkin M.** The Neural Basis of Stimulus Equivalence Across Retinal Translation. In: *Lateralization in the Nervous System*, edited by Harnad S. New York: Academic, 1977, p. 109–122.
- Gross CG, Rocha-Miranda CE, and Bender DB.** Visual properties of neurons in inferotemporal cortex of the macaque. *J Neurophysiol* 35: 96–111, 1972.
- Gur M and Snodderly DM.** Studying striate cortex neurons in behaving monkeys: benefits of image stabilization. *Vision Res* 27: 2081–2087, 1987.
- Hebb DO.** *The Organization of Behavior: A Neuropsychological Theory*. New York: Wiley, 1949.
- Horel JA.** Perception, learning and identification studied with reversible suppression of cortical visual areas in monkeys. *Behav Brain Res* 76: 199–214, 1996.
- Ito M, Tamura H, Fujita I, and Tanaka K.** Size and position invariance of neuronal responses in monkey inferotemporal cortex. *J Neurophysiol* 73: 218–226, 1995.
- Keyser C, Xiao DK, Foldiak P, and Perrett DI.** The speed of sight. *J Cogn Neurosci* 13: 90–101, 2001.
- Kobatake E and Tanaka K.** Neuronal selectivities to complex object-features in the ventral visual pathway of the macaque cerebral cortex. *J Neurophysiol* 71: 856–867, 1994.

- Kobatake E, Wang G, and Tanaka K.** Effects of shape-discrimination training on the selectivity of inferotemporal cells in adult monkeys. *J Neurophysiol* 80: 324–330, 1998.
- Leuschow A, Miller EK, and Desimone R.** Inferior temporal mechanisms for invariant object recognition. *Cereb Cortex* 5: 523–531, 1994.
- Logothetis NK, Pauls J, and Poggio T.** Shape representation in the inferior temporal cortex of monkeys. *Curr Biol* 5: 552–563, 1995.
- Logothetis NK and Pauls JP.** Psychophysical and physiological evidence for viewer-centered object representation in the primate. *Cereb Cortex* 5: 270–288, 1995.
- Logothetis NK and Sheinberg DL.** Visual object recognition. *Annu Rev Neurosci* 19: 577–621, 1996.
- Ludvigh E.** Extrafoveal visual acuity as measured with Snellen test-letters. *Am J Ophthalmol* 46: 102–113, 1941.
- Mel BW.** SEEMORE: combining color, shape, and texture histogramming in a neurally inspired approach to visual object recognition. *Neural Comput* 9: 777–804, 1997.
- Merigan WH and Katz LM.** Spatial resolution across the macaque retina. *Vision Res* 30: 985–991, 1990.
- Merigan WH and Maunsell JHR.** How parallel are the primate visual pathways? *Annu Rev Neurosci* 16: 369–402, 1993.
- Miller EK, Gochin PM, and Gross CG.** Suppression of visual responses of neurons in inferior temporal cortex of the awake macaque by addition of a second stimulus. *Brain Res* 616: 25–29, 1993.
- Missal M, Vogels R, Li C, and Orban GA.** Shape interactions in macaque inferior temporal neurons. *J Neurophysiol* 82: 131–142, 1999.
- Miyashita Y.** Neuronal correlate of visual associative long-term memory in the primate visual cortex. *Nature* 335: 817–820, 1988.
- Miyashita Y.** Inferior temporal cortex: where visual perception meets memory. *Annu Rev Neurosci* 16: 245–263, 1993.
- Muller JR, Metha AB, Krauskopf J, and Lennie P.** Information conveyed by onset transients in responses of striate cortical neurons. *J Neurosci* 21: 6978–6990, 2001.
- Nazir TA and O'Regan JK.** Some results on translation invariance in the human visual system. *Spat Vis* 5: 81–100, 1990.
- Olshausen BA, Anderson CH, and Van Essen DC.** A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *J Neurosci* 13: 4700–4719, 1993.
- Op de Beeck H and Vogels R.** Spatial sensitivity of macaque inferior temporal neurons. *J Comp Neurol* 426: 505–518, 2000.
- Perrett DI, Rolls ET, and Caan W.** Visual neurons responsive to faces in the monkey temporal cortex. *Exp Brain Res* 47: 329–342, 1982.
- Perry VH and Cowey A.** The ganglion cell and cone distributions in the monkey's retina: implications for central magnification factors. *Vision Res* 25: 1795–1810, 1985.
- Poggio T.** A theory of how the brain might work. *Cold Spring Harb Symp Quant Biol* 55: 899–910, 1990.
- Richmond BJ, Wurtz RH, and Sato T.** Visual responses of inferior temporal neurons in awake rhesus monkey. *J Neurophysiol* 50: 1415–1432, 1983.
- Riesenhuber M and Poggio T.** Hierarchical models of object recognition in cortex. *Nat Neurosci* 2: 1019–1025, 1999.
- Riesenhuber M and Poggio T.** Models of object recognition. *Nat Neurosci* 3 Suppl: 1199–1204, 2000.
- Robinson DA.** A method of measuring eye movements using a scleral search coil in a magnetic field. *IEEE Trans Biomed Eng* 101: 131–145, 1963.
- Rolls ET and Tovee MJ.** The responses of single neurons in the temporal visual cortical areas of the macaque when more than one stimulus is present in the receptive field. *Exp Brain Res* 103: 409–420, 1995.
- Salinas E and Abbott LF.** Invariant visual responses from attentional gain fields. *J Neurophysiol* 77: 3267–3272, 1997.
- Sary G, Vogels R, and Orban GA.** Cue-invariant shape selectivity of macaque inferior temporal neurons. *Science* 260: 995–997, 1993.
- Sato T.** Effects of attention and stimulus interaction on visual responses of inferior temporal neurons in macaque. *J Neurophysiol* 60: 344–364, 1988.
- Schoups AA, Vogels R, and Orban GA.** Human perceptual learning in identifying the oblique orientation: retinotopy, orientation specificity and monocularity. *J Physiol* 483: 797–810, 1995.
- Schoups AA, Vogels R, Qian N, and Orban GA.** Practising orientation identification improves orientation coding in V1 neurons. *Nature* 412: 549–553, 2001.
- Schwartz EL, Desimone R, Albright TD, and Gross CG.** Shape recognition and inferior temporal neurons. *Proc Nat Acad Sci USA* 80: 5776–5778, 1983.
- Shadlen MN and Newsome WT.** Noise, neural codes and cortical organization. *Curr Opin Neurobiol* 4: 569–579, 1994.
- Sigala N, Gabbiani F, and Logothetis NK.** Visual categorization and object representation in monkeys and humans. *J Cogn Neurosci* 14: 187–198, 2002.
- Softky WR, and Koch C.** The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *J Neurosci* 13: 334–350, 1993.
- Tanaka K.** Inferotemporal cortex and object vision. *Annu Rev Neurosci* 19: 109–139, 1996.
- Tovee MJ, Rolls ET, and Azzopardi P.** Translation invariance in the responses to faces of single neurons in the temporal visual cortical areas of the alert monkey. *J Neurophysiol* 72: 1049–1060, 1994.
- Ullman S.** *High Level Vision*. Cambridge, MA: MIT Press, 1996.
- Ungerleider LG and Mishkin M.** Two cortical visual systems. In: *Analysis of Visual Behavior*, edited by Ingle DJ, Goodale MA, and Mansfield RJW. Cambridge, MA: MIT Press, 1982, p. 549–585.
- Vetter T, Hurlbert A, and Poggio T.** View-based models of 3D object recognition: invariance to imaging transformations. *Cereb Cortex* 3: 261–269, 1995.
- Vogels R and Orban GA.** Activity of inferior temporal neurons during orientation discrimination with successively presented gratings. *J Neurophysiol* 71: 1428–1451, 1994.
- Vogels R, Sary G, and Orban GA.** How task-related are the responses of inferior temporal neurons? *Vis Neurosci* 12: 207–214, 1995.
- Wallis G and Rolls ET.** Invariant face and object recognition in the visual system. *Prog Neurobiol* 51: 167–194, 1997.
- Weiskrantz L and Saunders RC.** Impairments of visual object transforms in monkeys. *Brain* 107: 1033–1072, 1984.
- Young MP and Yamane S.** Sparse population coding of faces in the inferotemporal cortex. *Science* 256: 1327–1331, 1992.