The Receptive Fields of Inferior Temporal Cortex Neurons in Natural Scenes

Edmund T. Rolls, Nicholas C. Aggelopoulos, and Fashan Zheng
University of Oxford, Department of Experimental Psychology, South Parks Road, Oxford OX1 3UD, United Kingdom

Inferior temporal cortex neurons have generally been found to have large visual receptive fields that typically include the fovea and extend throughout much of the visual field. However, a problem of such a large receptive field is that it does not easily support object selection by subsequent processing areas, in that all objects within such a large receptive field might activate inferior temporal cortex cells. To clarify this, we recorded from inferior temporal cortex neurons while macaques searched for objects in complex natural scenes or in plain backgrounds, as normally used. Inferior temporal cortex neuron receptive fields were much smaller in natural scenes (mean radius, 11°) than in plain backgrounds (39°). With two objects in a scene, one of which was a target for action (a touch), the firing rates were equally high during foveation of the effective stimulus when it was the target and when it was the distractor in both the plain and the complex scenes. With a plain background and two objects present, the receptive fields were much larger (24°) for the stimulus when it was the target than when it was the distractor (9°). This effect of object-based attention was much less evident in the complex scene, when the receptive fields were small both when the stimulus was a distractor and when it was a target. The results show that the temporal visual cortex provides an unambiguous representation in natural scenes by responding to the object shown at or close to the fixation point.

Key words: **visual search; translation invariance; attention; object recognition; rhesus monkey; active vision**

Introduction

Inferior temporal cortex (IT) neurons of macaques have responses that provide information about objects or faces (Gross et al., 1972; Perrett et al., 1982; Rolls, 1992, 2000; Booth and Rolls, 1998; Rolls and Deco, 2002). The responses of these neurons are often relatively invariant with respect to the position in the visual field, size, and even view of the object (Gross et al., 1972; Rolls and Baylis, 1986; Tovee et al., 1994; Booth and Rolls, 1998). This is an important property, because when areas that receive from the inferior temporal visual cortex such as the orbitofrontal cortex, amygdala, and hippocampus learn about one view, position, or size of an object, the learning then generalizes to other views, positions, or sizes of the same object (Rolls and Treves, 1998; Rolls and Deco, 2002).

Much visual neurophysiology is conducted with one visual stimulus present in an otherwise blank visual scene (Hubel and Wiesel, 1982; Gross et al., 1985). Even in studies of the neuronal mechanisms of selective attention, there are usually only two small visual stimuli present in the visual field, which is otherwise blank (Chelazzi et al., 1993, 1998; Desimone and Duncan, 1995; Chelazzi, 1998; Chelazzi and Corbetta, 2000). In conditions in which one visual stimulus is present, inferior temporal cortex neurons typically have large receptive fields, $\geq$50° in diameter, under anesthesia (Gross et al., 1972) and when performing a visual fixation task (Tovee et al., 1994). However, a problem of such large inferior temporal cortex neuron receptive fields is that they do not easily support object selection by subsequent processing areas, in that all objects within such a large neuronal receptive field might activate different inferior temporal cortex cells, so that the output of the inferior temporal cortex might appear as a "tower of Babel." For example, if multiple objects were present within the large receptive field of inferior temporal cortex neurons, the orbitofrontal cortex and amygdala would retrieve many different reward–punishment associations simultaneously, and the rest of the brain would not know what to approach or avoid. Nor would there be any way of directing action at the correct goal object in such a scene with multiple objects, or indeed in any cluttered natural scene. Therefore, the issue arises of how the visual system operates in a natural and cluttered visual scene.

The aim was to investigate the sizes of the receptive fields of inferior temporal cortex neurons in natural scenes.

Materials and Methods

We measured the magnitude of the responses of inferior temporal cortex neurons when an effective stimulus was shown in blank scenes, in complex natural scenes, and in scenes with one other image present, as is typical in previous studies of attention. In the visual search task, in one condition the effective image was the object of attention, in the sense that the monkey was required to search for that object on the screen and touch it. In another condition, the effective image for the neuron was not the object of attention, in that the monkey was searching for another object to touch.

Recording techniques. The activity of single neurons was recorded with glass insulated tungsten microelectrodes in two macaque monkeys (*Macaca mulatta*; weight, ~4–6 kg) in a primate chair using techniques described previously (Rolls et al., 1990; Tovee et al., 1993; Booth and Rolls, 1998). All preparative and subsequent procedures were performed in accordance with the National Institutes of Health *Guide for the Care and Use of Laboratory Animals* and were licensed under the UK Animals (Scientific Procedures) Act of 1986. The action potentials of single neurons were amplified (Rolls et al., 1979) and converted to digital pulses using the trigger circuitry of an oscilloscope and analyzed online using an

Received April 24, 2002; revised Oct. 9, 2002; accepted Oct. 18, 2002.

This work was supported by the Wellcome Trust and by the Medical Research Council Interdisciplinary Research Centre for Cognitive Neuroscience.

Correspondence should be addressed to Prof. E. T. Rolls, Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford OX1 3UD, UK. E-mail: edmund.rolls@psy.ox.ac.uk.

 0270-6474/02/220339-10$15.00/0

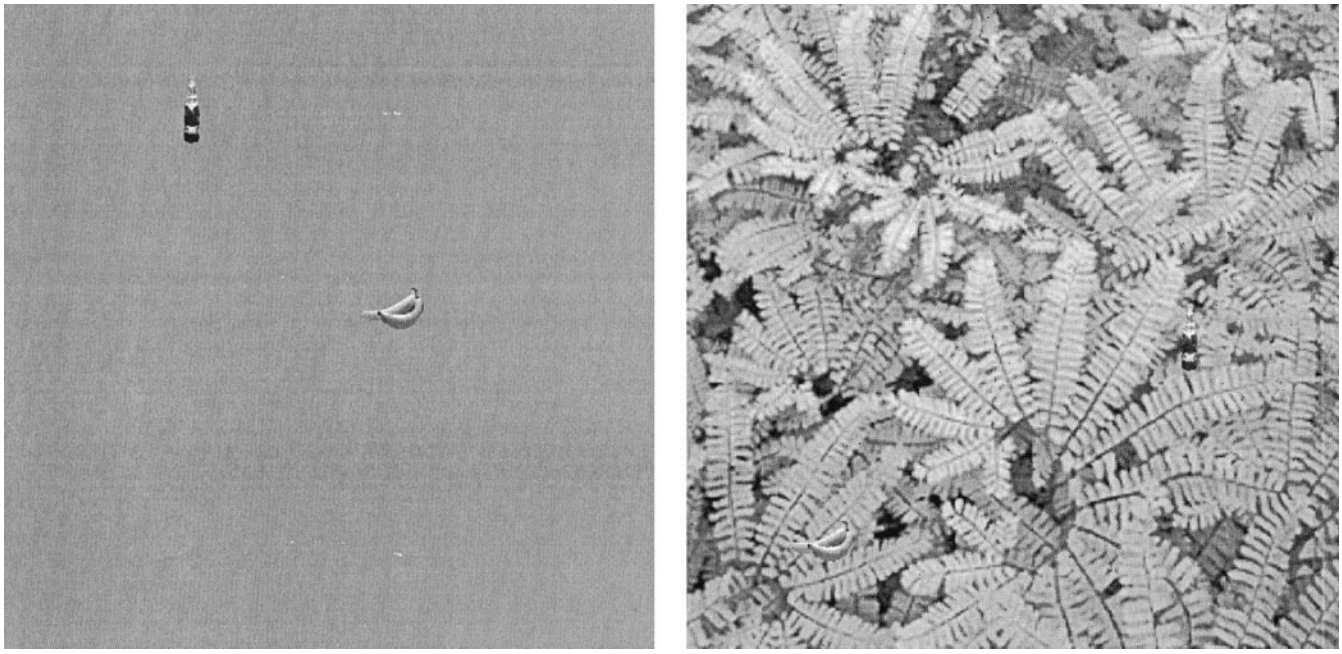

Figure 1. The visual search task. The monkey had to search for and touch an object (in this case a banana) when shown in a complex natural scene or when shown on a plain background. In each case a second object is present (a bottle), which the monkey must not touch.

IBM-compatible personal computer. The isolation of single neurons was ensured using Brainwave enhanced Discovery data acquisition, using cluster cutting for offline data analysis (DataWave Technologies, Longmont, CO), and establishing that no spikes occurred very close together in time (<3 msec) in the interspike interval histogram. Eye position was monitored and measured with the scleral search coil technique (Judge et al., 1980) using 1 kHz digitization and storage of new values every 20 msec.

Stimuli. The monkeys performed a visual search task in which if a particular image shown on a computer monitor was touched, the monkey obtained two to three drops of fruit juice for every touch. The monitor was at a distance of 23 cm from the monkey. The entire screen subtended 70 (horizontal) × 55° at the retina (with 512 × 512 pixels), and an object typically subtended 9 × 7° (~3.6 × 2.8 cm on the screen) in a first series of experiments and 5 × 3.5° in a second series of experiments. An example of a typical visual display is shown in Figure 1. (The refresh rate was 100 Hz.) The object had a resolution of 64 × 64 pixels but was prepared in such a way that each object with its particular outline could be presented on either a complex background or a blank background that had a resolution of 512 × 512 pixels. The target object occurred in random positions on the screen from trial to trial within the boundary of the screen. (We note that if the target was diagonally opposite the fixation position, the eccentricity could be >70°, given that the screen width was 70°.) Each trial was preceded by a 0.5 sec tone cue to enable the monkey to look at the screen. The monkey was allowed to touch up to four times to obtain separate aliquots of a taste reward before the next trial started. Trials in which the target object appeared in a blank screen or a natural scene were run in random order within a block of trials. The first block of trials typically involved measuring the neuronal responses when a single object was shown in a blank screen or in a complex natural scene. Other blocks of trials are described below.

Procedure. Tracks were made into the cortex of the superior temporal sulcus (STS) and the IT; the responses of isolated neurons were measured for a wide variety of small stimuli on the video monitor. These included faces, objects, sine-wave gratings, and boundary curvature descriptors (Rolls and Tovee, 1995). If the neuron responded to some but not other stimuli, the search continued to find an object to which the neuron had a large response (e.g., >50 spikes/sec); it was easy to find another object to which the neuron did not respond. It was also a condition for running the experiment that the neuron did not respond to the background. Most anterior inferior temporal cortex neurons (at coordinates that were typically 3–7 mm posterior to the sphenoid reference; see Fig. 7) did not respond to the background image (which for most experiments was that shown in Fig. 1), but if the neuron did respond, other background images were tried. (In the more posterior inferior temporal visual cortex, at coordinates ~9–11 mm posterior to the sphenoid reference, clusters of neurons were frequently encountered that did respond to the background images, and the experiment could not be performed with such neurons.)

Once a stimulus-selective cell was found without responses to the complex background natural scene image, blocks of trials were run in which stimuli were shown for which the target object appeared in a blank screen or a natural scene in random order. The first block of trials typically involved measuring the neuronal responses when a single object was shown in a blank screen or in a complex natural scene. The second block of trials typically involved measuring the neuronal responses when two objects were shown, one of which was a target that when touched led to the delivery of a taste reward, and the other of which if touched led to the delivery of aversive saline. The two objects were shown in a random sequence in plain backgrounds or complex natural scenes. In this second block of trials, a noneffective stimulus for the cell was normally the target, so that the effects of attention directed away from the effective visual stimulus could be measured as a function of the distance of the fovea from the effective stimulus when it was not the target for action. In all cases in these experiments, the ineffective stimulus produced no difference of the firing rate from the spontaneous value. The third block of trials was typically similar to the second block, except that in this case the effective object of the pair was the target for action. This condition enabled the neuronal responses to be measured to an effective stimulus when it was the target for action and when it was shown with a single distractor in either a plain background or in a complex scene. In the third block of trials, the monkey quickly learned, when touches of the previously rewarded object resulted in the delivery of a drop of aversive saline, to search for and touch the other object to obtain fruit juice. (It was found, as shown in the figures, that there was no effect on the neuronal responses of this stimulus–reward reversal, in that the neuron responded to the stimuli independently of the reward association provided that the monkey looked at the stimuli, as in previous studies (Rolls et al., 1977). Because trial block 3 was run last, the neuronal responses were sometimes slightly smaller on average than in trial blocks 1 and 2.

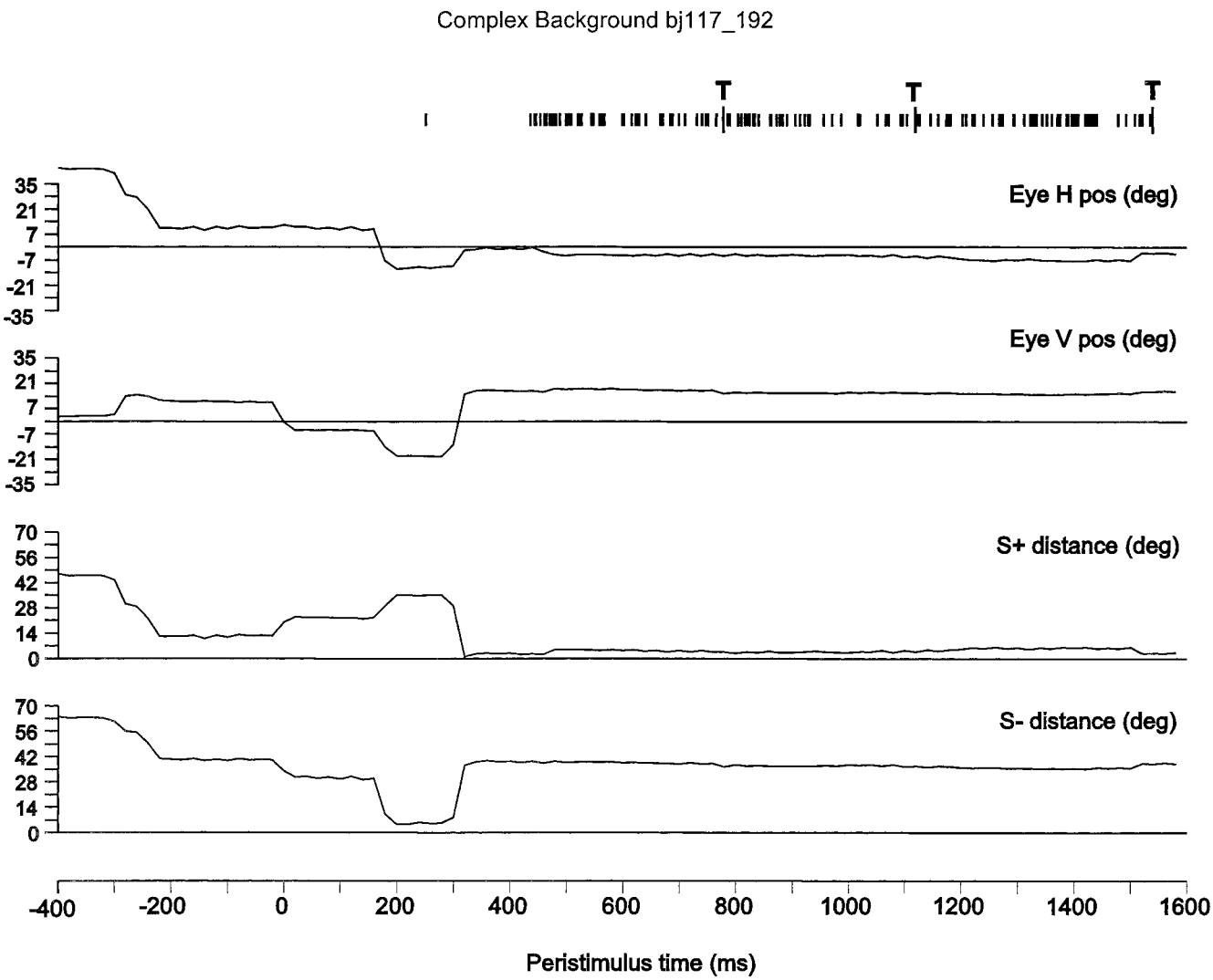

Figure 2. Eye position and neuronal response data collection during the performance of the visual search task for one neuron. The horizontal (*H*) and vertical (*V*) eye position traces are calibrated with respect to the center of the screen in degrees (with −35° horizontal and −35° vertical being the bottom left of the screen). Separate traces show the distance of the fovea from the target search object (S+) and from the distractor object (S−). A rastergram is shown *above*, with each *vertical line* representing an action potential from the neuron. The visual display was switched on at time 0, and after looking at two different positions for which there was only spontaneous neuronal activity, the eyes saccaded to the target object at ~340 msec after the stimulus, the neuron responded ~100 msec later, and the monkey then made multiple touches (*T*) of the object to obtain fruit juice.

Data analysis. The aim of the data analysis was to obtain measures of the firing rate of the cell when fixation was at different distances from the effective object, in both the complex natural scene and in the blank background.

During the experiment, calibration trials for the scleral eye position recording system were run so that the output of the eye position measurement system could be obtained in degrees for every stimulus position on the screen. To obtain calibration data, the output of the eye position monitoring system was measured while the monkey performed a visual fixation task with fixation points in a five-position array (top left, top right, screen center, bottom left, and bottom right) (or, for five neurons in the sample of neurons tested with small stimuli, in a 12-position calibration array). Similarly good eye calibration data were obtained in a similar task in which small objects were shown at the same grid positions, and the monkey touched the objects to obtain a fruit juice reward. In the touch task, the monkey fixated the small stimuli in the period before it touched them (as will be illustrated below). The values obtained during the calibration task enabled other eye position values obtained during the main experiment to be converted to degrees relative to screen center. The conversion program took into account any rotation and shear that was required to transform the eye position values into screen coordinates in degrees relative to screen center. Proof that this procedure worked accurately was that the transform applied to the calibration data placed the stimuli on the screen accurately to within ~1° over the entire width and height of the screen. One way in which this was confirmed was by showing that the final eye position measured by the procedure was within 1–2° of the stimulus that was being touched wherever the stimulus was on the screen.

Typical eye position data collected during the performance of the task are illustrated in Figure 2. The visual stimulus appeared at time 0 in a complex background. The monkey had to make several saccades (three on the trial shown) around the scene before the stimulus was found. Eventually a saccade found the target, and the eyes tended to remain still fixating the object for several hundreds of milliseconds while the monkey repeatedly touched the touchscreen to obtain up to four aliquots of a fruit juice reward. (One aliquot of ~0.15 ml was delivered for each press from a tube placed 2 mm in front of the monkey's mouth. The monkey typically had to touch the object within 3° of the center of the object to obtain the reward. If more than one object was present in the scene, touching the wrong object, the S− in what was essentially a visual discrimination task, resulted in the delivery of an aliquot of aversive saline taste.) After each saccade the eyes remained still for periods that were typically in the range

of 150–250 msec. In the complex natural scene, the monkey sometimes had to make up to eight saccades before its search found the target. There was no clear pattern to these saccades, and it was only when a saccade landed near the object that the monkey reached to touch the object if it was the target of the search. In the blank scene, often one saccade was sufficient, but especially when two stimuli were on the screen, one or two more saccades were sometimes needed, because sometimes the first saccade was to the nontarget object.

The firing rates of a cell as a function of the distance from the effective stimulus were measured during each period in which the eyes were still for >100 msec during the search task. The algorithm implemented in a computer program searched for any such period in which the eyes were still to within <1° for ≥100 msec and measured the number of spikes in one or more periods each 100 msec long in which the eyes remained still. This resulted in a large number of firing rate measurements, each of which was at a given distance from the effective visual stimulus for the cell. [The distance from the target object (S+) and the distractor object (S−) is plotted in Figure 2. It is clear when the monkey found the target, because after that time the eyes remained relatively still on or close to the target, and ~250 msec later the monkey had touched the screen, as indicated by the *T* values.] It is shown in Figure 2 that, typically for the cells analyzed here, the neuron responded to the stimulus ~100 msec after the eyes landed on the target. This is a typical response latency for anterior inferior temporal cortex neurons (Baylis et al., 1987); it reflects retinal delays as well as the cortical processing in each of the stages from the primary visual cortex to the inferior temporal cortex, which is on the order of 15 msec per stage (Panzeri et al., 2001). If the delay was different for a particular cell (as shown by latency measurements both in the touchscreen task and in a visual fixation task), the lag parameter was adjusted accordingly (to compensate for the delay between the arrival of the stimulus at the retina and the neuronal response). The output of the firing rate measurement algorithm consisted of several hundred 100 msec firing rate measurements, each taken when the monkey was fixating at different angular distances from the effective stimulus for the neuron being recorded. The firing rate measurements were then binned into 2° bins, the first including data for 0–2° from the center of the stimulus, the second 2–4°, etc. (For additional analyses, and in some of the graphs for clarity, 5° bins were used.) The means and SEs of the firing rates at different eccentricities were calculated and graphed to give an indication of which effects were significant. Statistical tests such as *t* tests were performed to test whether (within each block of trials), for example, the firing rates of a cell were different when an effective stimulus was being fixated in plain and complex natural scenes, and one-way and two-way ANOVAs were performed to test additional hypotheses, as indicated in Results. The statistical analyses included checks that the data were approximately normally distributed and nonparametric analyses for additional confirmation, using the methods described by Siegel and Castellan (1988) and Meddis (1984).

Recording sites. X-radiographs were taken at the end of each recording session to determine the position of the microelectrode relative to bony landmarks and the permanently implanted reference electrodes. At the end of the final tracks, microlesions were made in the areas of cortex in which recordings were made to mark typical recording sites (Feigenbaum and Rolls, 1991). Reconstructions of the tracks were made in serial 50 μm histological sections using the positions of the microlesions and the reference electrodes in the histology, the corresponding x-ray coordinates, and the x-ray coordinates of all recorded cells, to determine the locations of all of the cells.

Results

We recorded from well isolated inferior temporal cortex neurons in three hemispheres of two monkeys. In the course of these recordings, it was possible to find a reasonable number of neurons that responded well and with selectivity to some of the small test images of objects, faces, etc. that were available, with the proportions of different types of responsive cells similar to those we have reported previously (Baylis et al., 1987) (It was also a condition of the experiment that the neurons did not respond to the complex natural background image.) It was possible to complete all of the extensive testing required in these experiments, which took several hours, for one subset of nine cells tested with $9 \times 7°$ stimuli (supplemented by seven additional neurons tested in the condition with one stimulus present in either a plain or complex background) and a second subset of eight cells tested with $5 \times 3.5°$ stimuli. The same results were found in all 17 neurons on which it was possible to complete sufficient testing in all three conditions. [Of the other neurons recorded that could not be used for the experiments described here, only a proportion (~22%) responded differentially to any of the images of objects and faces in the set of images available, as described by Baylis et al. (1987).] In addition, some of the neurons did not respond with large responses to the small images used in this investigation; ~20% responded to the complex background stimuli of natural scenes and so could not be used in this particular experiment; and some were not held sufficiently long enough for a data set to be obtained.

The data obtained in one of these experiments with 9° stimuli are shown in Figure 3. The firing rate of cell bj168 was ~25 spikes/sec when the monkey fixated the effective stimulus in the plain background; this value was little affected when the monkey fixated the same object when shown in the complex natural scene (Fig. 3, *left*). The firing rate remained high even when the monkey fixated far from (up to 40° away from) the effective stimulus in the plain (blank) background. However, with the complex background, the firing rate fell markedly as a function of the distance of the position being fixated from the effective stimulus. The half-amplitude radius of the receptive-field size (the angle at which the firing rate had dropped to half its value relative to the spontaneous rate when the object was fixated) was ~17°. The means and SEs of the responses give an indication that for example the firing rate as a function of eccentricity was markedly different in the blank and complex natural scenes; indeed, a two-way ANOVA (performed within trial block 1 and with two conditions, background type and eccentricity of fixation) revealed a highly significant interaction ($F_{(10,2131)} = 3.49$; $p < 0.0003$).

When the same cell (Fig. 3) was tested with two stimuli present, one effective and one noneffective for the cell, and the effective stimulus was not the target to be touched, the firing rate when the effective stimulus was being looked at (Fig. 3, *middle top*, firing rate close to the fovea) was nevertheless very similar to its value when only the effective stimulus was shown. The same firing rate was obtained for the effective stimulus when it was not the target for action in both the plain and complex natural scenes. Thus, the data show that even when an effective stimulus is not the target for action (and in this sense attention is not being paid to it), there is nevertheless a large firing rate response of the inferior temporal cortex neuron to the stimulus provided that it is being fixated. The effect of attention does become evident however when we consider the firing rates when the monkey is fixating more than a few degrees away from the effective stimulus. Under these conditions, the firing rate drops markedly as a function of the fixation distance away from the object (Fig. 3, *middle*). Indeed, the radius (measured by the half-amplitude width) of the receptive field of the neuron under these conditions was 9°. Thus, attention (in this case making a different stimulus the target for action) influenced the size of the receptive field of the inferior temporal cortex neuron but not its firing rate when an effective but nontarget stimulus was being fixated.

This point is also established by the data obtained in the third block of trials in which the same two stimuli were being shown, but now the effective stimulus for the neuron was the target to be

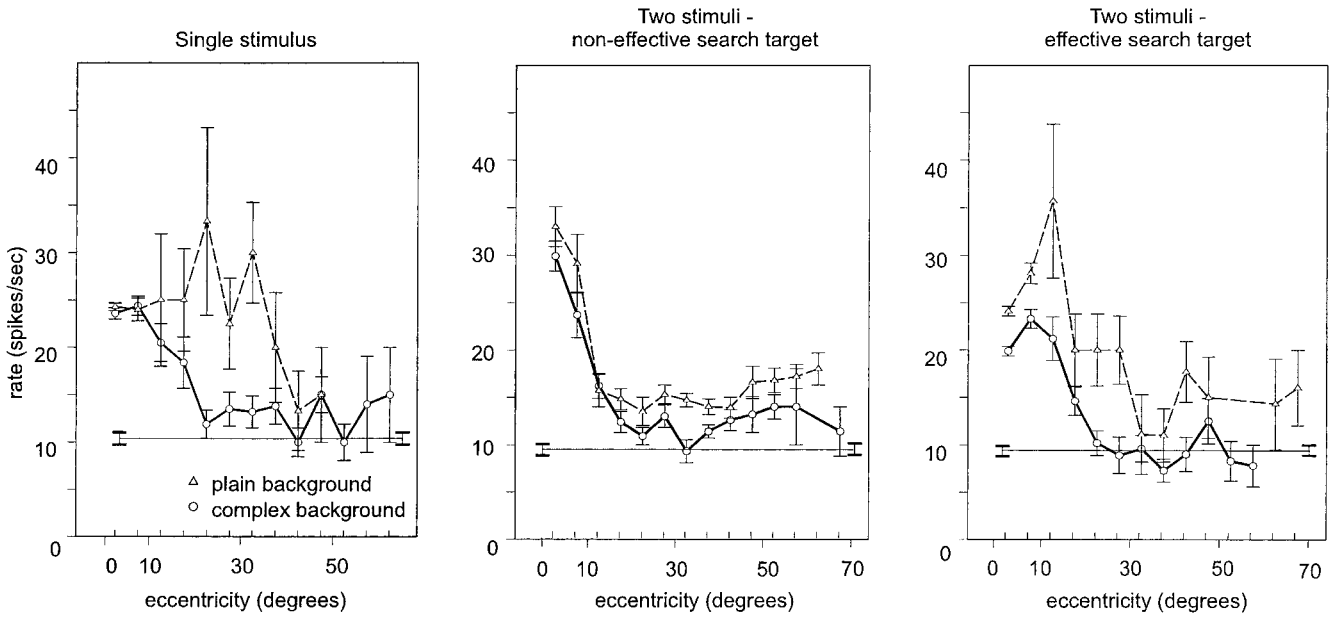

Figure 3. Firing of an inferior temporal cortex neuron (bj168) in the visual search tasks when 9° stimuli were used. Separate graphs are shown for the response in either a complex (natural scene) background or on a blank background to an effective stimulus for the neurons as a function of the distance of the fovea from the center of the effective stimulus (eccentricity degrees). *Left*, The neuronal responses to the effective stimulus object when a single object, which was the target for action, was present. *Middle*, The neuronal responses to the effective stimulus object when two objects were present and the noneffective object was the target for action. *Right*, The neuronal responses to the effective stimulus object when two objects were present and the effective object was the target for action. In all cases in these experiments, the ineffective stimulus produced no difference of the firing rate from the spontaneous value, which is indicated. The data are illustrated (because large, 9° stimuli were used) binned into 5° bins (the first at 0–5° from the center of the effective stimulus, the second at 5–10°, etc.), and the mean and SE of the neuronal response data points in 100 msec epochs of steady fixation within each binned limit are shown. The number of data points within each bin was often several hundred, although there were fewer data points (although still typically ≥20) in the bins far from the target in the blank background condition (and hence the larger SEs), because the monkey could typically find the object quickly in the blank scene and did not make as many saccades with eye positions away from the target while looking for it. *Horizontal lines* show the means, and error bars indicate the SEM of the baseline spontaneous firing rate of the neuron.

touched (Fig. 3, *right*). The effect of this was to increase the receptive-field size of the neuron (to 29°) in the two-stimulus display with a plain background (Fig. 3, compare *right* with *middle*). The effect of making the effective stimulus a target to be selected in the two-stimulus display was also to increase the receptive-field size slightly in the complex background (to 17°). Thus, object-based attention in the visual search task had a minor effect in a complex natural scene of increasing the receptive-field size for a target stimulus compared with the condition when the effective stimulus for the neuron was not a target, but the effect was much smaller in the complex natural scene than when two stimuli were shown in a plain background. The latter is the normal condition in which experiments on attention have been performed previously (Chelazzi et al., 1993, 1998; Desimone and Duncan, 1995; Chelazzi, 1998; Chelazzi and Corbetta, 2000).

A comparison of the data shown in Figure 3, *top left* and *top right*, showed that this neuron had the same firing rate to the effective stimulus when fixated in the blank screen independently of whether it was shown alone or with a second stimulus. However, simply having a second stimulus present, although it was not a target for action, reduced the size of the receptive field of the neuron somewhat. Thus, just having a second stimulus present had some effect on receptive-field size, but the effect was much less than that produced by a complex natural scene (Fig. 3, compare *top right blank screen* with *top left complex natural scene*).

The data for the 9° stimuli for the nine neurons in which testing in all conditions was completed are shown in Figure 4 and Table 1. Figure 4 shows the mean firing rate for the nine neurons in the six conditions tested. To combine the data for different neurons, the firing rates were first expressed as the firing rate for that neuron relative to the condition when the stimulus was being fixated (i.e., as a percentage). The interaction term showing that the population of neurons responded differently as a function of eccentricity in, for example, the condition with one stimulus present (trial block 1) was significant ($F_{(7,56)} = 2.53$; $p < 0.025$). Furthermore, a one-way ANOVA showed that the receptive-field sizes were different in the different conditions of trial block 1 (with a single stimulus present) ($F_{(5,48)} = 10.27$; $p < 0.00002$). Table 1 shows the receptive-field sizes for the neurons in the six conditions tested. The results for the population of neurons show effects that are very similar to those described for neuron bj168 in Figure 3 (*top*). In particular, with one object in the scene (Fig. 4, *left*), the receptive fields are large with a plain background (averaging 71.6° for the half-amplitude width) and very much smaller in a complex natural scene (where they average 25.6° for the half-amplitude width) (*post hoc* test; $p < 0.002$; trial block 1). When there are two objects in a plain scene compared with one object, the receptive fields of the neurons are generally smaller (averaging 55.6° when the effective stimulus is the search target). Thus, just introducing a second stimulus into a display, even when it is to be ignored, does reduce the size of inferior temporal cortex neurons somewhat (Fig. 4, compare *left* and *right*, plain background condition). However, attentional effects on receptive-field size are most clearly evident when there are two stimuli in an otherwise plain background, because in this condition the receptive field is small (29.6°) when an effective stimulus is not the target for action (Fig. 4, *middle*, plain background condition) and is larger (55.6°) when the effective stimulus is the target for action (Fig. 4, *right*, plain background condition) (*post hoc* test; $p < 0.04$; comparing trials from blocks 2 and 3 for the plain background condition). (It should

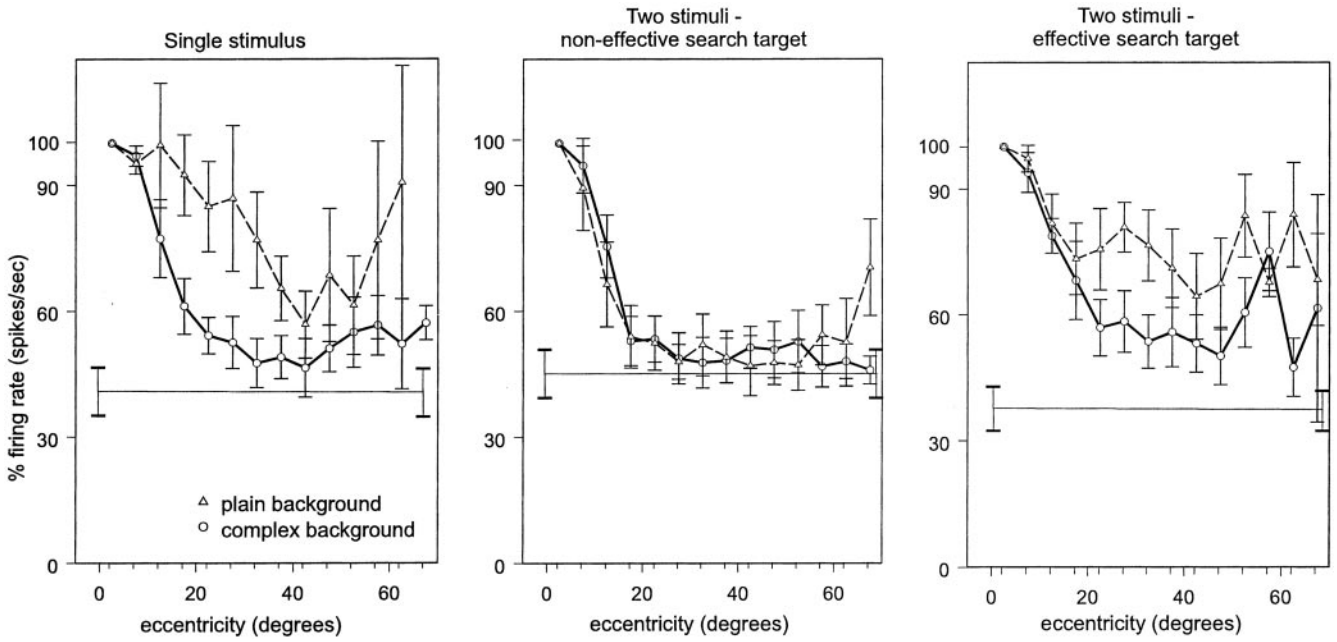

Figure 4. The mean firing rate for the nine neurons tested with 9° stimuli (together with the SEMs) in the six conditions tested, normalized across neurons in each case by calculating the neuronal response as a percentage of its value when the effective stimulus was being fixated in the blank scene. The rates are shown as a function of the eccentricity of the stimulus from the fovea. In all cases, the response shown is to the effective stimulus. *Left*, The neuronal responses to the effective stimulus object when a single object, which was the target for action, was present. *Middle*, The search target was the noneffective stimulus for the cell being recorded. *Right*, The search target was the effective stimulus for the cell being recorded. *Horizontal lines* show the means, and error bars indicate the SEM of the baseline spontaneous firing rate of the neurons.

Table 1. Inferior temporal cortex: receptive field size (radius, in degrees)

	Radius of receptive field (in degrees)					
	One stimulus, target for action (one stimulus, S+/0)		Unattended, not target for action (two stimuli, S−/S+)		Attended, target for action (two stimuli, S+/S−)	
Stimulus diameter	Plain	Complex	Plain	Complex	Plain	Complex
9° stimuli (mean ± SE)	35.8 ± 5.0	12.8 ± 1.5	14.8 ± 1.3	10.2 ± 1.3	27.8 ± 4.8	12.7 ± 1.3
5° stimuli (mean ± SE)	38.8 ± 4.2	11.0 ± 1.0	8.6 ± 1.6	7.8 ± 1.6	23.5 ± 3.7	9.6 ± 0.8

be noted that the receptive-field diameters given include the 9 × 7° stimulus, so that a receptive-field width of 20° extends only 5.5° beyond the edge of the stimuli on each side.) In the two-stimulus condition, as soon as a complex natural background is present, the receptive fields become small. Under these conditions of a natural background, attention has some effect in increasing the receptive-field size for a target stimulus, but the effect is rather minor (compare Fig. 4, *right* and *middle*). [The increase is from 20.4 to 25.4° diameter, as shown in Table 1, and this difference is not significant (Wilcoxon $T = 8.5$; $n = 9$).]

The values for the firing rates when the effective 9 × 7° stimulus is being fixated under the different conditions were very similar to each other and were not statistically significantly different. First, provided that an object is being fixated, there is little difference in the firing rate in the blank scene and complex natural scene conditions [e.g., 48.9 vs 44.7 spikes/sec with one object present in trial block 1 (p values not significant; $t = 1.50$; df = 8)]. Thus, inferior temporal cortex neurons can respond as well to objects when the objects are shown in complex scenes as they do when the objects are shown in plain backgrounds, provided that the object is fixated. Given that the neurons did not respond at all to the noneffective stimulus, this finding shows that the tuning of the neurons is not affected, and remains tuned to the same effective stimuli, when they are shown in real scenes as when they are shown against blank backgrounds. Thus, inferior temporal cortex neurons code for objects even when they have to be segmented out of complex backgrounds. Second, it is shown that even if an effective stimulus for a cell is not a target for action, then the neuronal response is as large to the stimulus as when it is a target for action (54.6 vs 48.9 spikes/sec in a plain background and 43.8 vs 44.7 spikes/sec in a complex background). Thus, attention per se (defined by whether the object in the scene is to be selected for action in a visual search task) makes little difference to the firing of inferior temporal cortex neurons, provided that the stimulus is being fixated.

In addition to the results described for neurons tested in all six experimental conditions with the 9° stimuli, additional results, which confirm those already described, were obtained in an additional seven neurons in the main experimental condition, with one object that was the target for visual search in the blank versus complex natural scene conditions. In these seven additional experiments (which were performed in two monkeys), the mean ± SEM firing rates when the effective object was being looked at in the blank and complex scenes were similar (53.3 ± 13.1 vs 53.1 ± 15.4 spikes/sec), and the receptive-field diameters were reduced from a mean of 58.8 ± 10.0° in the blank scene to 29.4 ± 4.8° in the complex natural scene.

In general, comparable results for the second subset of eight

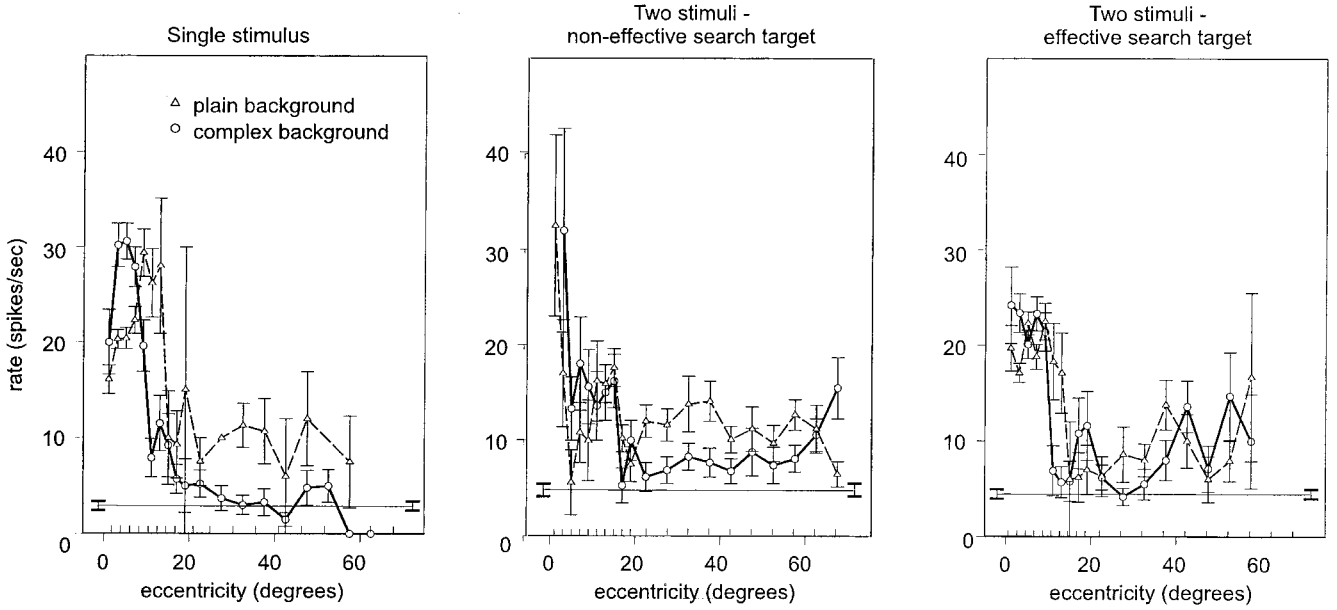

Figure 5. Firing of an inferior temporal cortex neuron (bj346) in the visual search tasks when 5° stimuli were used. Separate graphs are shown for the response in either a complex (natural scene) background or on a blank background to an effective stimulus for the neurons as a function of the distance of the eyes from the center of the effective stimulus. Conventions are as in Figure 3, except that the data are illustrated (because small stimuli were used) with 2° binning for bins in the range 0–20°.

neurons tested with 5 × 3.5° stimuli were found, except that the receptive-field sizes were slightly smaller overall than with the 9° stimuli, reflecting, as predicted, the smaller stimulus size. Figure 5 shows the data from a single cell tested with the small (5°) stimuli (compare with Fig. 3 for the larger stimuli). The small receptive field in, for example, the condition in which the effective stimulus is not the target for the search is shown in Figure 5 (*middle*). Table 1 shows the receptive-field sizes in the six test conditions; the firing rates are indicated in Figure 6. The firing rate data show that when the effective stimulus was being fixated, there was little effect of whether the stimulus was shown in a plain background versus a complex natural scene (32.6 vs 33.9 spikes/sec when one stimulus was on the screen). The main effect found was, as before, on the size of the receptive field, which was large with one stimulus present in the plain screen condition (77.6°) and small (22.0°) with the complex background (Table 1). As shown in Table 1, if two stimuli were present, then the receptive fields were still quite large in the plain background if the object was the target of the search (47° diameter) and were quite small (17.2°) if the object was not the target for the search ($p < 0.01$; one-tailed; Wilcoxon $T = 1.5$; $n = 8$). Thus, attention defined operationally in this way did have clear effects on receptive-field size in the plain background. However, in the complex background, the receptive-field size was rather small not only when the object was not the target of the search (15.6° diameter) but also when it was the target of the search (19.2°; Wilcoxon $T = 9.5$; $n = 8$; NS) (Table 1). Thus, there was little effect of attention defined operationally in this way with the complex background. The smallest receptive-field diameter found, 15.6°, extended ~5.3° beyond the edge of the small (5 × 3.5°) stimuli.

The main effects for the first subset of cells tested with 9 × 7° stimuli and for the second subset of eight cells tested with 5° stimuli are summarized in Table 1. Some of the important comparisons have been described above. A diagram that schematizes to scale the results found with the 5 × 3.5° stimuli is provided in Figure 6.

Recording tracks in both monkeys were made over an extensive portion of the inferior temporal cortex, from the upper and lower banks and fundus of the superior temporal sulcus, through the middle temporal gyrus to just lateral of the middle temporal sulcus. The recording sites of the cells reported here are illustrated in Figure 7. As can be seen in Figure 7, the cells are distributed in a region that extends from the gyrus lateral to the middle temporal sulcus to the lower bank of the STS, and the investigated area of cortex is indicated by the *boxed area* contained in the *bottom left* coronal section.

Discussion

These experiments show that inferior temporal cortex neurons respond to objects very well when they are shown in complex natural scenes, provided that the object is fixated. In addition, in natural scenes the receptive fields of inferior temporal cortex neurons become much smaller than they are to objects shown in blank scenes. The output of the inferior temporal cortex thus reflects best the single object that is being fixated in complex natural scenes. This property of inferior temporal cortex neurons helps to make the output of the inferior temporal visual cortex easily decoded and interpreted, for there is no confusion about which object in a complex scene with multiple objects the output is about. The output is about the object being fixated, and the receptive field includes or is close to the fovea.

There has been little previous investigation of IT neuronal activity when objects are shown in natural scenes. During a visual fixation task used to control fixation and with one object in a complex natural scene or against a plain background, Rolls et al. (2000) showed that the tuning of neurons to effective stimuli was similar in complex scenes and in plain backgrounds and that the magnitude of the neuronal responses was also similar, but that the receptive-field sizes were reduced in the complex natural scene. Sheinberg and Logothetis (2001) found, in a task in which a single small target object (1.5°) was presented in a complex natural scene 20 × 20° in a task in which the monkey had to move a lever right or left to different objects, that IT neurons had similar tuning to that in a plain background. (They did not compare

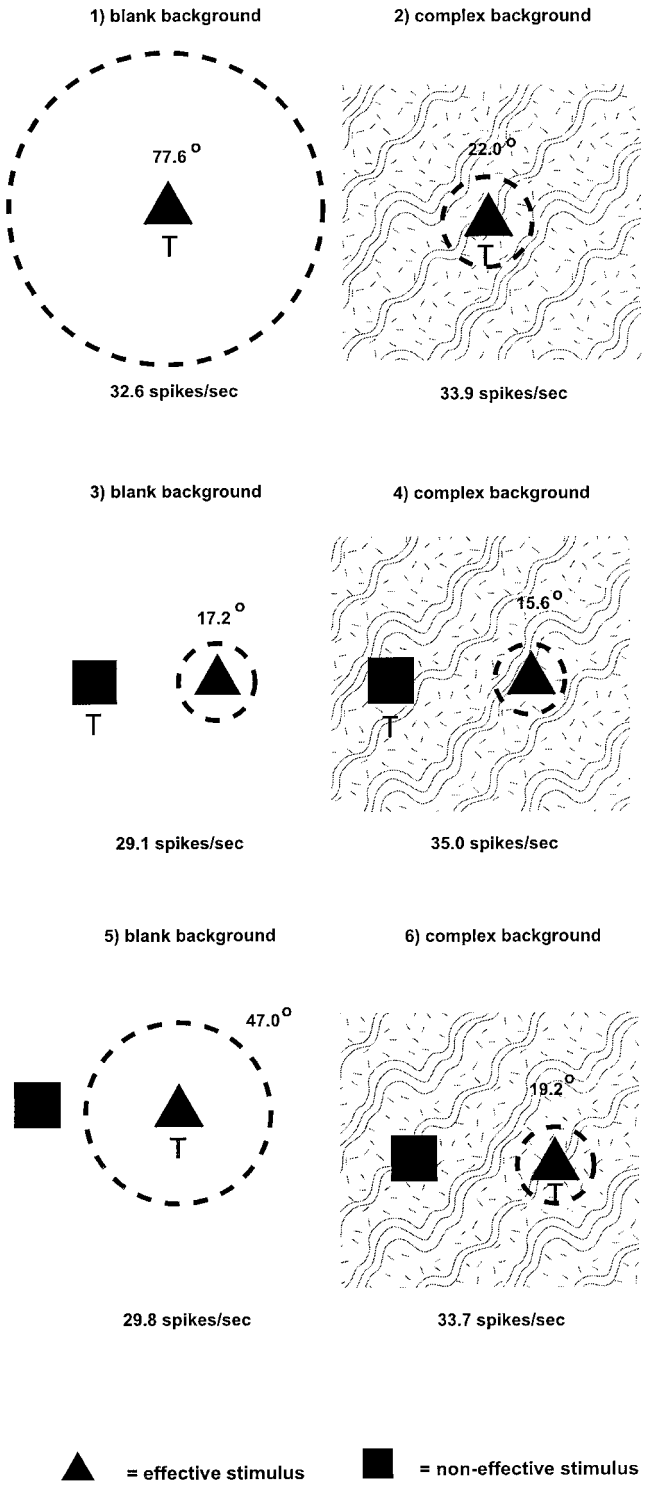

Figure 6. Summary of the receptive-field sizes of inferior temporal cortex neurons to a 5° effective stimulus presented in either a blank background (*1, 3, 5*) or in a natural scene (*complex background*; *2, 4, 6*). The stimulus that was a target for action in the different experimental conditions is marked by *T*. When the target stimulus was touched, a reward was obtained. The mean receptive-field diameter of the population of neurons analyzed and the mean firing rate in spikes per second are shown. The stimuli subtended 5 × 3.5° at the retina and occurred on each trial in a random position in the 70 × 55° screen. The *dashed circle* is proportional to the receptive-field size. *Top*, Responses with one visual stimulus in a blank (*left*) or complex (*right*) background. *Middle*, Responses with two stimuli when the effective stimulus was not the target of the visual search. *Bottom*, Responses with two stimuli when the effective stimulus was the target of the visual search.

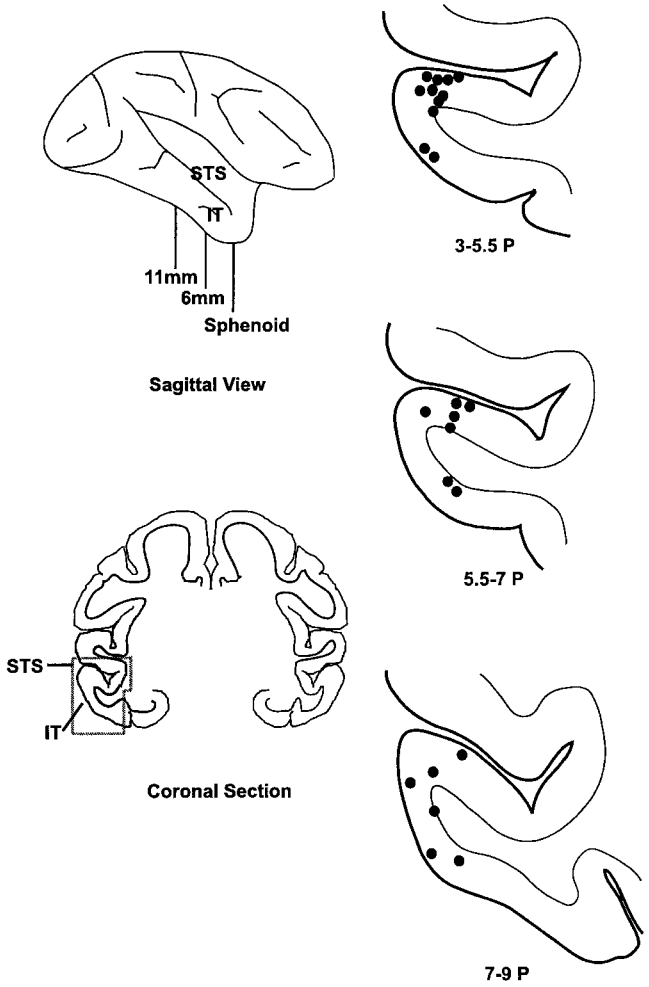

Figure 7. Reconstructed histological coronal sections show by *filled circles* the sites at which the neurons analyzed in this study were recorded. *Numbers below* the sections indicate the distance (in millimeters) posterior (*P*) to the sphenoid bone reference point (which is at approximately the anteroposterior level of the anterior commissure), and these distances are also illustrated in the *topleft* of the figure in the lateral view. A full coronal section is illustrated at the *top right* of the figure, and the area of cortex investigated in this study is indicated by the *boxed area* encompassing the STS and the lateral portion of the IT.

the sizes of the receptive fields in blank and natural scene conditions, and they had only one object in the display; thus, they did not study the effects of object-based attention.)

Attention can be operationally defined in the experiment described here by whether an object is the target for action, and is an example of object-based attention (Rolls and Deco, 2002). In terms of this definition, attentional effects were found in these experiments in that the receptive fields were larger (23.5° radius) in the plain background for the object that was the search target than for the object that was to be ignored (8.6° radius, data for 5° stimuli as shown in Table 1). However, the magnitude of this object-based attentional effect was very small in the complex natural scene, in that the receptive fields of inferior temporal cortex neurons were only slightly larger (by 1.8°, 9.6 vs 7.8° radius) for the target than for the object to be ignored in the complex scene (with the 5° stimuli). Space-based attention (in which attention is directed to a particular location) might show a larger effect on the firing rate to the attended versus unattended location, but we note that additional investigations would be needed to show that even this is the case in natural scenes, because most previous

studies of attention have been performed when two small objects are present in a plain background (Chelazzi et al., 1993, 1998; Desimone and Duncan, 1995; Chelazzi, 1998; Chelazzi and Corbetta, 2000).

It might be suggested that in fact attention is needed in the complex scene to process the distractor object, and that this is why the inferior temporal cortex neurons responded to fixated objects in the complex scene even when they were not the target of the visual search. However, we note that the same high firing rate when an object is fixated, and reduction in receptive-field size, occur when objects are passively being viewed in complex scenes during the performance of a visual fixation task (Rolls et al., 2000; E. T. Rolls and M. C. A. Booth, unpublished observations). Thus, the high firing rates for fixated distractor objects and reduction in receptive-field size found in the experiments described here in natural scenes primarily reflect the operation of the inferior temporal cortex neurons in the presence of a complex background, and not that the monkey was paying special attention to the distractor to determine whether it was the target for action or was not to be selected for action. Thus, the mechanism by which the reduction in receptive-field size occurs appears to be related to competition between different features in the visual scene. The result of the competition is that whatever object is at the fovea appears to be given preference in determining the output of inferior temporal cortex neurons. The evidence that this is the case is that the neurons always fired when the monkey fixated the objects, and that the firing decreased monotonically as a function of the distance from the fovea. Models that use the higher cortical magnification factor for the fovea can account for these findings (Rolls and Deco, 2002; Trappenberg et al., 2002).

It is interesting and important that when objects are shown in complex scenes, not only the magnitude of IT neuronal responses but also the tuning of the neurons to the objects remains relatively unaffected (Sheinberg and Logothetis, 2001). This is shown in the present study by the fact that there is little if any reduction in the firing rate to an effective stimulus when a complex natural background is introduced (relative to the rate in the blank background) (Figs. 3–6), and that the neurons did not start to respond in the complex scene to the stimulus that was ineffective when tested in the blank screen. This finding might be called "background invariance," to capture the point that the tuning of many inferior temporal cortex neurons is invariant when the stimuli are shown against a background. This particular result complements the findings of DiCarlo and Maunsell (2000) and Missall et al. (1999) that inferotemporal cortex neurons respond similarly to an effective shape stimulus for a cell even if some distractor stimuli are present a few degrees away. This result was also found by Rolls and Tovee (1995), who showed in addition that with two stimuli present in the visual field, the anterior inferior temporal neuronal responses were weighted toward the stimulus that was closest to the fovea. In early visual cortical areas (V1–V4), free viewing in natural scenes may produce some reduction in neuronal responses (Livingstone et al., 1996; Gallant et al., 1998), although the testing conditions were so different from those described here that a direct comparison is not realistic.

The results described here confirm the finding by Rolls et al. (1977) that inferior temporal cortex neurons respond to visual stimuli independently of their reward or punishment association; they show in addition that the inferior temporal cortex provides an output that can be unambiguous to receiving stages because it is about one object, that at the fovea. Succeeding stages can then easily use a pattern associator to determine whether that stimulus is associated with taste reward or punishment (Rolls, 1990, 1999; Rolls and Treves, 1998; Rolls and Deco, 2002).

Both object-based and space-based attentional processes can be understood quantitatively by a model of ventral and dorsal visual stream processing in which during covert visual search (i.e., without eye movements) competition in early common visual areas is biased by top-down object bias applied to the end of the ventral stream or by top-down spatial bias applied to the end of the dorsal stream (Rolls and Deco, 2002). However, the fundamental points made here are that search for objects in complex natural scenes is performed more by overt processes (moving the eyes around the scene) and is facilitated by the much reduced receptive-field sizes of inferior temporal cortex neurons in natural scenes that typically include the fovea attributable in part to its large cortical magnification factor (Rolls and Deco, 2002; Trappenberg et al., 2002). Moreover, the results described in this paper indicate that the coordinates of the object in space that is to be the target for action are passed to the motor system by virtue of the facts that the object represented in the inferior temporal cortex in complex scenes is at the fovea and that the dorsal visual system that executes the actions has information about eye gaze position (cf. Ballard, 1991; Rolls and Deco, 2002).

References

Ballard DH (1991) Animate vision. Artif Intell 48:57–86.

Baylis GC, Rolls ET, Leonard CM (1987) Functional subdivisions of temporal lobe neocortex. J Neurosci 7:330–342.

Booth MCA, Rolls ET (1998) View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. Cereb Cortex 8:510–523.

Chelazzi L (1998) Serial attention mechanisms in visual search: a critical look at the evidence. Psychiatry Res 62:195–219.

Chelazzi L, Corbetta M (2000) Cortical mechanisms of visuospatial attention in the primate brain. In: The new cognitive neurosciences, Ed (Gazzaniga MS, ed), pp 667–686. Cambridge, MA: MIT.

Chelazzi L, Miller EK, Duncan J, Desimone RE (1993) A neural basis for visual search in inferior temporal cortex. Nature 363:345–347.

Chelazzi L, Duncan J, Miller EK, Desimone RE (1998) Responses of neurons in inferior temporal cortex during memory-guided visual search. J Neurophysiol 80:2918–2940.

Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. Annu Rev Neurosci 18:193–222.

DiCarlo JJ, Maunsell JHR (2000) Form representation in monkey inferotemporal cortex is virtually unaltered by free viewing. Nat Neurosci 3:814–821.

Feigenbaum JD, Rolls ET (1991) Allocentric and egocentric spatial information processing in the hippocampal formation of the behaving monkey. Psychobiology 19:21–40.

Gallant JL, Connor CE, Van Essen D (1998) Neural activities in areas V1, V2,and V4 during free viewing of natural scenes compared to controlled viewing. NeuroReport 9:2153–2158.

Gross CG, Rocha Miranda CE, Bender DB (1972) Visual properties of neurons in inferotemporal cortex of the macaque. J Neurophysiol 35:96–111.

Gross CG, Desimone R, Albright TD, Schwartz EL (1985) Inferior temporal cortex and pattern recognition. Exp Brain Res [Suppl]11:179–201.

Hubel DH, Wiesel TN (1982) Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. J Physiol (Lond) 160:106–154.

Judge SJ, Richmond BJ, Chu FC (1980) Implantation of magnetic search coils for measurement of eye position: an improved method. Vision Res 20:535–538.

Livingstone MS, Freeman DC, Hubel DH (1996) Visual responses in V1 of freely viewing monkeys. Cold Spring Harbor Symp Quant Biol 61:27–37.

Meddis R (1984) Statistics using ranks. a unified approach. Oxford: Blackwell.

Missall M, Vogels R, Chao-Yi L, Orban GA (1999) Shape interactions in inferior temporal neurons. J Neurophysiol 82:131–142.

Panzeri S, Rolls ET, Battaglia F, Lavis R (2001) Speed of information retrieval in multilayer networks of integrate-and-fire neurons. Network 12:423–440.

Perrett DI, Rolls ET, Caan W (1982) Visual neurones responsive to faces in the monkey temporal cortex. Exp Brain Res 47:329–342.

Rolls ET (1990) A theory of emotion, and its application to understanding the neural basis of emotion. Cognit Emot 4:161–190.

Rolls ET (1992) Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas. Phil Trans R Soc Lond B Biol Sci 5:11–21.

Rolls ET (1999) The brain and emotion. Oxford: Oxford UP.

Rolls ET (2000) Functions of the primate temporal lobe cortical visual areas in invariant visual object and face recognition. Neuron 27:205–218.

Rolls ET, Baylis GC (1986) Size and contrast have only small effects on the responses to faces of neurons in the cortex of the superior temporal sulcus of the monkey. Exp Brain Res 65:38–48.

Rolls ET, Deco G (2002) Computational neuroscience of vision. Oxford: Oxford UP.

Rolls ET, Tovee MJ (1995) Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. J Neurophysiol 73:713–726.

Rolls ET, Treves A (1998) Neural networks and brain function. Oxford: Oxford UP.

Rolls ET, Judge SJ, Sanghera M (1977) Activity of neurones in the inferotemporal cortex of the alert monkey. Brain Res 130:229–238.

Rolls ET, Sanghera MK, Roper-Hall A (1979) The latency of activation of neurons in the lateral hypothalamus and substantia innominata during feeding in the monkey. Brain Res 164:121–135.

Rolls ET, Yaxley S, Sienkiewicz ZJ (1990) Gustatory responses of single neurons in the orbitofrontal cortex of the macaque monkey. J Neurophysiol 64:1055–1066.

Rolls ET, Webb B, Booth MCA (2000) Responses of inferior temporal cortex neurons to objects in natural scenes. Soc Neurosci Abstr 26:1331.

Sheinberg DL, Logothetis NK (2001) Noticing familiar objects in real world scenes: the role of temporal cortical neurons in natural vision. J Neurosci 21:1340–1350.

Siegel S, Castellan NJ (1988) Nonparametric statistics. New York: McGraw-Hill.

Tovee MJ, Rolls ET, Treves A, Bellis RP (1993) Information encoding and the responses of single neurons in the primate temporal visual cortex. J Neurophysiol 70:640–654.

Tovee MJ, Rolls ET, Azzopardi P (1994) Translation invariance in the responses to faces of single neurons in the temporal visual cortical areas of the alert macaque. J Neurophysiol 72:1049–1060.

Trappenberg TP, Rolls ET, Stringer SM (2002) Effective size of receptive fields of inferior temporal cortex neurons in natural scenes. In: Advances in Neural Information Processing Systems 14. Cambridge, MA: MIT, in press.