Sensory Modality Specificity of Neural Activity Related to Memory in Visual Cortex

JAY R. GIBSON AND JOHN H. R. MAUNSELL

1 Department of Neuroscience, Brown University, Providence, Rhode Island 02912; and 2 Division of Neuroscience, Baylor College of Medicine, Houston, Texas 77030

Gibson, Jay R. and John H. R. Maunsell. Sensory modality specificity of neural activity related to memory in visual cortex. J. Neurophysiol. 78: 1263–1275, 1997. Previous studies have shown that when monkeys perform a delayed match-to-sample (DMS) task, some neurons in inferotemporal visual cortex are activated selectively during the delay period when the animal must remember particular visual stimuli. This selective delay activity may be involved in short-term memory. It does not depend on visual stimulation; both auditory and tactile stimuli can trigger selective delay activity in inferotemporal cortex when animals expect to respond to visual stimuli in a DMS task. We have examined the overall modality specificity of delay period activity using a variety of auditory/visual cross-modal and unimodal DMS tasks. The cross-modal DMS tasks involved making specific long-term memory associations between visual and auditory stimuli, whereas the unimodal DMS tasks were standard identity matching tasks. Delay activity existed in auditory/visual cross-modal DMS tasks whether the animal anticipated responding to visual or auditory stimuli. No evidence of selective delay period activation was seen in a purely auditory DMS task. Delay-selective cells were relatively common in one animal where they constituted up to 53% neurons tested with a given task. This was only the case for up to 9% of cells in a second animal. In the first animal, a specific long-term memory representation for learned cross-modal associations was observed in delay activity, indicating that this type of representation need not be purely visual. Furthermore, in this same animal, delay activity in one cross-modal task, an auditory-to-visual task, predicted correct and incorrect responses. These results suggest that neurons in inferotemporal cortex contribute to abstract memory representations that can be activated by input from other sensory modalities, but these representations are specific to visual behaviors.

INTRODUCTION

Inferotemporal cortex (IT) is part of the ventral pathway in primate visual cortex (Ungerleider and Mishkin 1982) and plays an important role in pattern recognition (Cowey and Gross 1970; Dean 1976; Desimone et al. 1984; Gross et al. 1972; Iwai and Mishkin 1969). IT contains neurons that respond selectively to complex stimuli, such as faces and hands (Kobatake and Tanaka 1994), suggesting that behaviorally relevant visual stimuli are represented explicitly in this cortical area.

Studies in behaving monkeys also have shown that neurons in many regions of visual cortex are affected by extraretinal inputs related to attention (Moran and Desimone 1985; Motter 1993; Treue and Maunsell 1996), motor signals (Andersen and Mountcastle 1983; Wilson et al. 1990), and short-term memory (Fuster and Jervey 1982; Miller et al. 1991). These extraretinal signals demonstrate that sensory processing in visual cortex depends on behavior. Characterizing these signals will be important for understanding how sensory information is used to guide behavior. We focus here on activity associated with short-term memory in IT and its relation to long-term memory.

We studied short-term memory in monkeys performing a delayed match-to-sample (DMS) task, in which animals are required to remember a stimulus during a short delay period and signal whether a later test stimulus is the same (match) or different (nonmatch). While animals perform this task, some single units in IT maintain firing during the delay (Fuster and Jervey 1982; Miyashita and Chang 1988). This firing, or delay activity, is selective in that activity is high while the animal is remembering one stimulus, but low while remembering another, and hence, it is correlated with short-term memory for specific stimuli. Selective maintained firing has been observed in many brain structures in the context of both stimulus memory and motor planning (Gnadt and Andersen 1988; Hikosaka et al. 1989; Koch and Fuster 1989; Kurata and Wise 1988; Mays and Sparks 1980; Niki 1974). It has not, however, been observed in all studies of short-term memory in visual cortex and thus may not be relevant to all forms of visual short-term memory (Baysis and Rolls 1987; Eskandar et al. 1992; Miller et al. 1993).

Although several studies have demonstrated memory-related activity in IT, the modality specificity of that activity remains in question. This issue is important for understanding the abstract nature of this neural activity and how it differs from sensory processing in unimodal sensory cortex. The activity of neurons in the ventral pathway in visual cortex has been shown to be affected by signals arising from somatosensory or auditory stimuli. Somatosensory stimuli can modulate selectively visual responses and elicit delay activity in V4 (Haenny et al. 1988; Maunsell et al. 1991). Colombo and Gross (1994) reported that auditory stimuli also can elicit delay activity in IT. In these experiments, the animal always expected, and responded to, a visual test stimulus. Activity in visual cortex has not been examined in a task in which an animal anticipates a nonvisual stimulus. Nor has a thorough analysis been performed on the sensory modality specificity of visual cortical delay activity.

Delay activity in IT also has been shown to display distinct long-term memory representations (Miyashita 1988; Sakai and Miyashita 1991). Miyashita and Sakai trained monkeys to recognize randomly assigned pairings of visual stimuli. After the associations were learned, some IT neurons were activated selectively during the delay periods after the pre-
sentation of either member of a particular pair. This pattern of activation presumably depended on the animal being trained to associate the particular stimuli and can be considered a correlate of long-term memory. We wanted to determine if such a representation could also exist for cross-modal associations.

Here, we present data addressing the issue of modality specificity of neuronal activity in IT related to short- and long-term memory. Some data has been presented previously in preliminary form (Gibson and Maunsell 1994).

METHODS

Behavioral tasks

Single units were recorded in two macaque monkeys while they performed delayed match-to-sample tasks. Animal 1 was a 7.5-kg male Macaca mulatta and animal 2 was a 7.2-kg male M. nemestrina. Water intake was controlled 5 days a week during training and recording. The animals worked for apple juice rewards that were delivered at the end of every correctly completed trial.

Animals were trained to perform four delayed match-to-sample tasks, which differed in the modality of the sample and test stimuli used: a visual matching task, also referred to as the visual-to-visual task, a visual-to-auditory cross-modal matching task, an auditory-to-visual cross-modal matching task, and an auditory-to-auditory task (Fig. 1). Each trial began with the appearance of a small red spot (0.25° square) on the display. Once the animal had fixated this spot and pressed a lever, a 300-ms presample period began, followed by the presentation of a sample stimulus (400–500 ms). After the sample stimulus was turned off, there was a delay period when only the fixation spot was present. The delay period usually lasted 2 s (range, 1.6–3 s) and ended with the presentation of a test stimulus. These were the maximum delays with which animals could perform effectively all four tasks. The unimodal tasks (Fig. 1, A and D) were conventional DMS tasks in which the matching test stimuli were identical to the sample stimuli. If the test stimulus matched the sample stimulus, the animal was required to release the lever. If it did not match, the animal had to hold down the lever until the test stimulus was turned off. The animal had to remember the sample stimulus during the delay period to perform the trial successfully. The fixation spot was present during the entire trial, and the animal had to maintain fixation within 1° of this spot or the trial was aborted without reward. Visual stimuli were centered around the fixation point. All trials were separated by ≥2 s.

In the cross-modal association tasks (Fig. 1, B and C), the test stimulus was a different sensory modality than the sample stimulus, and the animal looked for a ‘‘paired associate,’’ rather than an identical stimulus. The same paired associates were used in the visual-auditory and auditory-visual tasks; the tasks differed only in which stimulus was the sample and which was the test. The associations learned by animal 1 are in Fig. 1.

Monkeys have difficulty discriminating and remembering auditory stimuli in a task of this sort (Colombo and D’Amato 1986; D’Amato and Colombo 1985; D’Amato and Salmon 1984). Learning is slow and there are practical limits on the number of auditory stimuli that can be learned. For our purposes, however, auditory stimuli offered advantages. For example, one objective was to have associations that were as unaffected as possible by prior experience. Some physical features, such as orientation, are common to visual and tactile signals whereas no obvious features are linked naturally in a similar manner between auditory and visual stimuli. Auditory stimuli also are delivered easily.

We used only two samples in each of the DMS tasks (Fig. 1). Animals performed blocks of one task at a time, with each block consisting of eight trials. When one task block was completed successfully, another task commenced. The order in which the four tasks were presented was random, but all four were completed before the cycle repeated. The trials within a single task block were picked at random with the constraint that the animal successfully completed the same number of trials for each sample stimulus. Sounds from the auditory-to-auditory matching task were never used in the cross-modal tasks. This was done to isolate effects in each task. If the same stimuli were used in the auditory-to-auditory and visual-to-auditory tasks and similar short-term memory correlated effects were observed in both, one could not be sure if effects were due to processes in each task, individually, or just due to the involvement of the stimuli in just one of the tasks.

The visual stimuli were Fourier descriptors (Zahn and Roskies 1972), ~10° across, white on a black background. The auditory stimuli were as follows: low tone, high tone, a repeated tap that sounded like a snare drum, and another repeated tap sound higher in frequency than the snare, which sounded like clicks. These sounds were selected because they proved distinct enough for the animals to distinguish readily and because their repetitive nature allowed them to be presented for arbitrary lengths of time (See D’Amato and Salmon 1984). A visual mask stimulus was presented immediately after the sample stimulus (Fig. 1) in all four tasks. Mask stimuli erase sensory persistence in psychological studies (Averbach and Coriell 1961; Coltheart 1983; Sperling 1960). The mask permitted us to
examine the effect of intervening visual stimuli and accompanying visual sensory processing on neural signals during the delay period (Miller et al. 1993). Auditory masks were not used because auditory sensory processing probably does not occur in IT (Desimone and Gross 1979). The masks were arrays of nine Fourier descriptors centered around the fixation point and with each element the same size as the visual stimuli used in the task.

**Surgical procedures**

Surgical procedures were performed under full anesthesia in aseptic conditions. Heart rate, respiration rate, and temperature were monitored throughout the surgery. Antibiotics (Bactrim) were given the day before and for 7 days after surgery. Animals were initially prepared with ketamine (15 mg/kg), atropine (50 g/kg), and diazepam (50 g/kg). After these drugs took effect, the animal was intubated and a venous catheter was inserted, and then the animal was mounted in a stereotaxic head holder and put under isoflurane anesthesia. Analgesic was administered after surgery (Banamine, im).

The head post and recording chambers were secured to the skull by acrylic cemented to stainless steel plates, which were attached to the skull with screws (Synthes). All chambers were positioned to record from the central and anterior portions of inferotemporal cortex. In animal 1, the chamber was placed over its right hemisphere for recording in IT from 9 to 20 mm AP. Animal 2 had a chamber placed over each hemisphere. One on the left side for recording 11–22 mm AP in IT, and another on the right side for recording 2.5–13.5 mm AP.

**Single unit recording**

During training and recording each monkey sat in a primate chair facing a color video monitor (75 Hz, 40 × 31 cm) ×65 cm away. A small loud speaker was placed below the animals line of sight at a distance of 40 cm. The animals head was stabilized for eye position monitoring and neuronal recording by anchoring the headpost. Eye position was monitored using a scleral search coil system (Judge et al. 1980; Robinson 1963). A computer was used for all data collection, stimulus presentation, and behavioral control.

While animals performed the delayed match-to-sample tasks, action potentials were recorded extracellularly with a tungsten microelectrode (1–2 MΩ, ~30-μm tips, Microprobe, Type B). Signals from the electrode were amplified, filtered, and transformed into pulses by a window discriminator. Spike times were recorded with 1-ns resolution, and behavioral events were recorded with 5-ns resolution. All time bases were synchronized with the vertical retrace of the video monitor at the beginning of each trial.

A guide tube and grid system (Crist et al. 1988) was used to direct the electrode to recording sites. The tip of the guide tube usually was positioned at the ventral bank of the lateral sulcus or in the white matter just dorsal to the superior temporal sulcus.

**Data analysis**

Except where noted, data analysis was based on correctly completed trials. Behavioral responses made within 200 ms of test onset were counted as guesses and excluded from analysis. We restricted our analysis of delay-period activity to the 800 ms immediately before test stimulus onset. This period was chosen because it was unlikely to be contaminated by responses to the offset of the sample stimulus, and it was a period when the animal had to remember the sample stimulus. Delay activity was typically determined using 15 repetitions on each condition (minimum 7).

Delay activity was compared between the two sample conditions in each task using a paired t-test (P < 0.05). Thus significant differences in delay activity were calculated with respect to each individual task in isolation. This method was chosen instead of pooling all eight sample conditions in the four tasks because this study focused on measuring the strength of delay selectivity correlated with short-term memory for sample stimuli within each single task. Selectivity was measured as differences in firing rates between two trial conditions. Delay activity also was compared between tasks by averaging delay activity rates for each task and performing a one-way analysis of variance (ANOVA; P < 0.05).

Sensory responses were analyzed task by task to avoid any task specific confounds. Responses to sample stimuli were determined in a period 75–400 ms after sample onset and were compared with activity during the 250 ms immediately before sample onset. Cells were considered responsive if activity was significantly different for either of the sample stimuli, although we recognize that activity during this period might include extraretinal effects associated with recognizing or remembering the sample. A one-way ANOVA and a multiple comparisons method were used for this analysis (Zar 1984). The selectivity of responses was determined by comparing both sample responses in a task to see if these values differed significantly.

Responses to test stimuli also were analyzed. In an attempt to avoid behavioral response influences on neural activity, we measured test responses in a time window beginning 75 ms and ending at 275 ms after test onset. This window was well before the earliest average reaction time for both monkeys (345 ms). When test responses were compared with sample responses, sample responses were analyzed using the same time window as that for the test.

**Histology**

When recordings were completed, the animal was euthanatized with an overdose of barbiturates (Nembutal, iv), and perfused with a phosphate buffer rinse followed by paraformaldehyde fixative. The brain then was removed, blocked coronally, and equilibrated with 30% sucrose in phosphate buffer. Blocks were sectioned at 40 μm on a freezing microtome. Sections were then mounted on slides, stained in cresyl violet, and coverslipped. Recording locations were identified using fiducial pins inserted into the grid shortly before perfusion.

**RESULTS**

Quantitative data were collected from 230 cells in IT. Of these, 174 cells were recorded while the monkeys performed all four tasks. The remaining 56 units were recorded while they performed only the two cross-modal tasks. Many more isolated units were encountered, but only those whose activity was obviously modulated by some aspect of the task were tested. The percentage of correctly completed trials in each task for animal 1 and animal 2 during data collection were: visual-to-visual, 95 and 99%; visual-to-auditory, 94 and 99%; auditory-to-visual, 89 and 99%; auditory-to-auditory, 80 and 89%.

**Sensory responses**

As expected, most sensory responses in IT were visual. Figure 2 shows for each task the percent of cells with significant responses to the sample stimulus. In the tasks with visual sample stimuli, ~70% of the neurons responded (125/174 for visual-to-visual and 121/174 for visual-to-auditory) compared with 30 and 19% in each task with auditory sample stimuli (52/174 for auditory-to-visual and 33/174 for auditory-to-auditory). Sensitivity to one stimulus modality was uncorrelated with sensitivity to the other.
A prevalence of visual sensory responses is consistent with earlier reports (Desimone and Gross 1979; Gross et al. 1972). It should be noted, however, that activity during the presentation of the sample stimulus may not have been entirely sensory. For example, IT neurons can be activated in a nonspecific way by auditory stimuli (Iwai et al. 1987; Ringo and O’Neill 1993). Consistent with this, the number of neurons reaching statistical criterion for auditory sample selectivity was no more than the number expected by chance, whereas about a quarter of cells discriminated between the visual sample stimuli. Furthermore, any traces of selectivity to auditory sample stimuli may be a result of selective delay activity beginning before the stimuli disappear (see below).

Selective delay activity

Many neurons had selective delay activity that depended on the identity of the sample stimulus. These neurons were significantly more active during those delay periods after particular sample stimuli, and this selective activity generally persisted throughout the delay period. The data in Fig. 3 are from a neuron that showed selective delay-period activity during the two cross-modal tasks. In the visual-to-auditory task, this cell was more active after sample stimulus 1, whereas in the auditory-to-visual task, the cell was more active after sample stimulus 2. This selective activation was not sensory because sensory stimulation (i.e., fixation point, the background, and the visual mask) was identical during the delay period in all trials. Selective delay activity has been documented in IT previously with both simple visual matching and visual association tasks, and it has been suggested to play a role in short-term memory (Fuster and Jervey 1982; Miyashita and Chang 1988). Although the neuron was also active during the delay period of the visual-to-visual and auditory-to-auditory tasks, we focus here on activity that was selective, or differential, for the sample stimuli within tasks.

Many neurons had selective delay activity during the two cross-modal tasks. Selectivity strength was quantified for each cell by calculating the difference in delay period firing rate between the two sample conditions of each task. For each cell four values were calculated—one for each task. The distributions of these differences are plotted in Fig. 4. During the visual-to-auditory task, 26% (46/174) of cells had statistically significant differences in delay activity between the two sample conditions. A similar proportion of significant differences was seen for the auditory-to-visual task (23%, 40/174). Fourteen percent (25/174) of cells showed statistically significant delay selectivity during the visual matching task, a value that falls within the range reported previously in studies using a similar number of stimuli (10%, Fuster and Jervey 1982; 18%, Colombo and Gross 1994). The number of neurons reaching statistical criterion for selective delay period activity during the auditory matching task (5%, 8/174) was no greater than the number expected by chance.

Whereas neurons from animal 1 had significant differences in the strength of delay selectivity between tasks, animal 2 had too few delay selective neurons to make such a comparison (see Individual differences in delay activity). Applying a one-way ANOVA and multiple comparisons test (Zar 1984), animal 1 had greater selectivity in the two cross-modal tasks.
modal tasks than in either of the unimodal tasks. These data suggest that delay activity is more selective in cross-modal tasks than in purely visual matching tasks.

Cells with significant delay-period selectivity in the cross-modal tasks had modest differences in firing rates between sample conditions, but the average modulation between the two conditions was ~60% [\(\frac{A - B}{\text{min}(A,B)}\)]. Thus tasks involving auditory stimuli produced clear selective delay activity in visual cortex, but only in tasks that required a visual discrimination.

Seven cells from animal 1 that displayed delay selectivity in both cross-modal tasks also were tested under passive viewing of identical task stimulation. In this test, the lever was removed, and the animal was only required to fixate while stimuli from the cross-modal tasks were presented. Delay selectivity that existed while the animal actively performed the cross-modal association tasks disappeared in six of these cells during passive viewing (average loss in selectivity, 94%). Selectivity returned when the animal was required to start performing the tasks again.

Although the focus in this study was delay activity dependent on remembered stimuli, delay activity also varied between the four tasks the animals performed (e.g., Fig. 3). Hence, delay activity may be related not only to short-term memory for specific stimuli but also to other aspects that differ between tasks. Half of the neurons sampled (88/174) in both animals had delay activity that differed significantly from task to task (41/68 in animal 1 and 49/106 in animal 2). Significant differences in delay activity between the visual and auditory matching tasks were the most frequent (45/88), followed by differences between the visual matching and the visual-to-auditory cross-modal tasks (43/88). Differences between tasks are confounded with stimulus selectivity because tasks used different sample stimuli. Many neurons that did not have differences in delay activity between the two conditions within a task, nevertheless had strikingly different average delay activity between the four tasks.

**Individual differences in delay activity**

There was a substantial difference in the incidence of delay selective neurons between the two animals. Animal 1 had many cells with strong delay selectivity. In animal 2, a few neurons had clear selective delay-period activity (Fig. 5), but these were rare (Table 1). The low incidence of delay selectivity in animal 2 was not due to poor performance because it performed ~4% better in every task compared with animal 1. An example of a delay-selective cell from animal 2 is provided in Fig. 5. The difference between animals is also unlikely to stem from sampling different regions of IT. Data were collected first from animal 1 in a region in IT between 9 and 20 mm anterior in the right hemisphere (Fig. 6A). Almost all delay-selective cells in animal 1 were in a region of the ventral bank of the STS located 9–12 mm anterior. In this region, we estimate that ~5–10% of the cells encountered displayed delay selectivity, whereas outside of this region, delay-selective cells were not nearly as often encountered. Recordings from IT in animal 2 were made initially in the left hemisphere in the range from 11 to 22 mm anterior. When few units with selective delay-period activity were found after searching this region, we sampled an overlapping, but more posterior, range from the right hemisphere (9–22 mm anterior; see Fig. 6B and C). A total of 65 electrode penetrations separated by no more than 2 mm were made at 51 sites in the two hemispheres. Given this density of sampling, we think it is unlikely that we failed to sample the region corresponding stereotaxically to that which contained the delay-period activity found in animal 1. The pronounced difference in the number of delay-selective cells observed between the animals may reflect a gross difference in the functional organization of their cortices, or perhaps different behavioral strategies used by each animal might have resulted in different brain structures or neural mechanisms being used.

**Long-term memory representation**

Of the neurons with delay selectivity in at least one of the cross-modal tasks, almost half showed selectivity during both (26/60). For animal 1 there was a clear relationship in delay selectivity between the cross-modal tasks. When visual sample 1 in the visual-to-auditory task elicited more delay-period activity, auditory sample 2 generally produced greater delay-period activity in the auditory-to-visual matching task (e.g., Fig. 3). The converse pairing also was ob-

**TABLE 1. Percent of delay-selective cells by animal**

<table>
<thead>
<tr>
<th>Task</th>
<th>Animal 1, %</th>
<th>Animal 2, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual-to-Visual</td>
<td>24</td>
<td>8</td>
</tr>
<tr>
<td>Visual-to-Auditory</td>
<td>44</td>
<td>8</td>
</tr>
<tr>
<td>Auditory-to-Visual</td>
<td>53</td>
<td>9</td>
</tr>
<tr>
<td>Auditory-to-Auditory</td>
<td>3</td>
<td>6</td>
</tr>
</tbody>
</table>

Only cells that were tested on all four tasks were included: 68 from animal 1 and 106 from animal 2.
This type of long-term representation encoded in delay activity has been described by Sakai and Miyashita (1991) in the context of visual-visual associations using the same type of associational DMS task. The data here show cross-modal associations also can generate such a representation.

Paradoxically, in our results the long-term representation observed in animal 2 (Fig. 8B, \( R = -0.02, P > 0.50, n = 119 \) ) , even when analysis was restricted to cells with delay selectivity in at least one of the two cross-modal tasks (\( R = -0.08, P > 0.50, n = 21 \) ).

We interpret the correlation between delay-period specificity in the cross-modal tasks as a specific long-term representation of the novel associations that animal 1 learned.
Relationship between sensory responses and delay activity

Consistent with earlier studies, sample response selectivity and delay period selectivity were fairly independent of each other (Ferrera et al. 1994; Fuster and Jervey 1982; Maunsell et al. 1991; Miyashita and Chang 1988). Many cells were stimulus-response selective while not being delay selective, and many were delay selective while not response selective (Figs. 3, 5, 7, and 11).

Within a given task, only weak relationships were observed. For example, in the auditory-to-visual task, there was a weak correlation ($R = 0.35, P = 0.04, n = 61$ delay-selective cells). In Fig. 9A, one can see that the selectivity during the delay is much greater than that observed during the sample epoch. These cells had negligible responses to the auditory sample stimuli and any correlation may have been due to memory-related processes starting during the sample presentation. A similar weak correlation was observed for the visual-to-auditory task ($R = 0.37, P = 0.002, n = 69$ delay-selective cells), and again, this correlation may be due to the initiation of delay activity during the sample period.

In the visual-to-visual task, the correlation was strong ($R = 0.60, P = 0.002, n = 25$). Colombo and Gross (1994) similarly noted a strong correlation between sample responses and delay activity in an auditory-to-visual task relative to a visual-to-visual task. Thus there appears to be some connection between sensory and delay activity, at least in certain circumstances.

Delay activity before incorrect responses

Both animals occasionally made errors. In a separate analysis, we compared delay-period activity during incorrectly completed trials with that from correct trials. Analysis was restricted to cells that showed significant differences in delay activity in at least one of the cross-modal tasks. If the animal did not make at least two errors in each condition of a particular task while a cell was monitored, it was not used in the analysis of that task. Because animal 2 made so few errors and because there were only a few delay-selective cells found in this animal, only cells from animal 1 qualified for this analysis. The analysis could not be performed on either unimodal task either because of too few errors or lack of delay-period selectivity.

One cell’s activity during correctly and incorrectly completed trials in the auditory-to-visual task is illustrated in Fig. 10. During correctly completed trials, the neuron was most active after the sample 2 (snare). During incorrectly completed trials the neuron was more active after the sample 1. Thus high levels of delay-period activity were correlated...
reliably with responding to the paired associate for sample 2, correctly or not.

Although the data for the cell in Fig. 10 included enough incorrect trials to show a statistically significant reversal, most cells did not. Nevertheless, when the sample of 23 cells is analyzed together, the reversal is statistically significant. In Fig. 11, we plot the difference in delay-period firing rate during correctly completed trials against the difference during incorrectly completed trials for each cell. There was a strong anticorrelation between the correct and incorrect rate differences \((R = -0.80, P < 0.001, n = 23)\), which indicates that delay selectivity over the population predicted whether the animal would respond correctly.

Surprisingly, no such relationship existed for the visual-to-auditory task. Delay activity in this task was the same whether or not the animal made a correct or incorrect response. Among the cells that met criteria for analysis in this task, there was a strong positive correlation between correct and incorrect rate differences indicating that selectivity remained the same regardless of response accuracy \((R = 0.86, P < 0.001, n = 17)\). This difference may arise from differences in the monkeys ability to analyze visual and auditory stimuli (see DISCUSSION).

**Effects of the mask stimulus on delay activity**

A visual mask was presented after every sample stimulus. Although this mask was irrelevant to the matching tasks, it nevertheless evoked strong responses from some neurons in IT. Figure 12 shows the range of responses. Some cells were excited or inhibited by the mask (Fig. 12, A and C), whereas others showed no response (Fig. 12B). Using data from animal 1, where ample time before mask onset permitted analysis, delay-period selectivity was generally evident before and after the mask appeared \((R = 0.51, P < 0.001, n = 176)\), only instances of statistically significant delay activity included.

Earlier studies that used a sequential matching task in which several stimuli were presented before the match appeared, failed to see sustained delay-period activity in IT (Baylis and Rolls 1987; Eskandar et al. 1992; Miller et al. 1993). Miller et al. (1993) reported selective delay activity immediately after the sample stimulus, but this activity was erased by stimuli that intervened between the sample and the matching stimulus. In that study, the animal had to attend to the intervening stimuli. The visual mask in the current study differed from that of Miller et al. (1993) in that the animal could ignore this intervening stimulus and still successfully perform the task.

**Sensory response modulation related to short-term memory**

Many studies have reported modulation of sensory responses dependent upon recently presented stimuli and thus related to short-term memory (Fuster and Jervey 1982; Gross et al. 1979; Vogels et al. 1995). During a sequential DMS task, Miller et al. (1993) found cells in inferotemporal cortex that showed a decreased response to matching test stimuli relative to nonmatching test stimuli—a match suppression effect. Using a similar task, Ferrera et al. (1994) reported a similar “match suppression” in parietal visual cortex and area V4, although the effect was small.

Although cells did show response modulation based on whether a test stimulus was a match or nonmatch, we found no systematic match stimulus suppression. As in Miller et al. (1993), this analysis was restricted to cells exhibiting excitatory sensory responses, and a two-way ANOVA was used (stimulus against match/nonmatch). We observed significant match/nonmatch modulation in 14/87 cells in the visual-to-auditory task and 17/91 cells in the auditory-to-visual task. Modulation of match stimuli was seldom observed in the visual-to-auditory and auditory-to-auditory tasks (2/30 and 0/20). A match/nonmatch index (Miller et al. 1993) was calculated for cells in the visual-to-auditory and auditory-to-visual tasks and then tabulated in frequency histograms (Fig. 13).

\[
\text{match/nonmatch index} = \frac{(\text{avg. match response} - \text{avg. nonmatch response})}{(\text{avg. match response} + \text{avg. nonmatch response})}
\]

**FIG. 10.** Delay activity predicts behavior. Data from 1 cell collected during auditory-to-visual task are sorted with respect to correct and incorrect behavioral responses. Using a 2-factor analysis of variance, this cell’s delay activity had no significant effects with respect to sample stimulus or response accuracy (correct/incorrect) alone, but interaction term was highly significant \((P < 0.0001)\). This interaction represents the reversal of the delay selectivity between correct and incorrect responses. As in Fig. 5, a gap exists in histograms because spikes were collected from trials of varying delay length. First 1,600 ms and last 1,300 ms of trials are plotted. Rasters show neural activity on a trial-by-trial basis. Only every 7th trial is illustrated for rasters representing correct responses.
This index is a normalized difference between match and nonmatch responses. A positive index value means matching stimuli evoked a higher sensory response. In neither task did mean indices deviate appreciably from zero: $-0.014 \pm 0.014$ (mean $\pm$ SE) for the visual matching task and $-0.012 \pm 0.013$ for the auditory-to-visual task.

The discrepancy between these data and that of Miller et al. may arise from differences in behavioral paradigm used. They used novel stimuli in each recording session, and they used interfering stimuli that were relevant to the task. We also collected fewer responses to matching test stimuli, so small systematic effects on test responses may not have been detected.

Previous studies also have found that sensory responses depend on whether stimuli are a sample or a test—sample/test modulation (Eskandar et al. 1992; Gross et al. 1979; Mikami and Kubota 1980; Riches et al. 1991). Like these earlier studies, we compared sample and test responses for the visual matching task, whereas for the cross-modal tasks, visual sample responses from the visual-to-auditory task were compared with visual test responses in the auditory-to-visual task.

Sample/test modulation came in the form of both increases and decreases in the sensory response. Effects were observed in 34/125 and 33/121 visually responsive cells in the visual-to-visual and cross-modal tasks, respectively, and modulation was usually the same in the two tasks. For only 6/125 cells in the visual matching task and 14/121 cells for the cross-modal tasks, there was an effect of sample stimulus identity upon sample/test modulation (a statistically significant interaction between sample/test modulation and the stimuli in the task). Thus these few cells encoded information about the particular sample stimulus in a given trial. When this analysis was performed with respect to auditory stimuli, 20/52 and 17/33 responsive cells showed effects in the cross-modal and auditory-to-auditory tasks. Again modulation tended to be the same in these tasks. In summary, sample/test modulation in this study was not specific to the visual modality.

**DISCUSSION**

As in previous studies of IT, we observed selective neural activity during the delay period of delayed match-to-sample tasks that was correlated with short-term memory for specific stimuli (Fuster and Jervey 1982; Miyashita and Chang 1988). This activity was observed not only in a simple visual matching task, but also during cross-modal tasks in which auditory and visual stimuli were associated. Stimulus-specific delay activity could be evoked by auditory stimuli (in the auditory-to-visual task), confirming earlier studies showing that nonvisual stimuli can trigger memory-related neural activity in visual cortex (Colombo and Gross 1994; Haenny et al. 1988). Stimulus-selective delay activity also existed when animals expected to respond to an auditory stimulus but only when the task involved visual stimuli. Furthermore, we showed that delay activity could reveal long-term representations for learned cross-modal associations, and we showed that delay activity also could predict correct or incorrect responses in one behavioral task.

Delay activity did not exist for the purely auditory task. Thus it seems that memory representations in visual cortex involving nonvisual stimuli require that these stimuli be associated with visual stimuli. This could mean that memory representations in visual cortex, just as in sensory processing, are specific for visual stimuli. It is possible that we sampled regions that specialize in representing cross-modal associations, but the fact that sensory responses were solely visual and the existence of delay selectivity in the visual-to-visual task argue against this.

The presence of delay activity in both cross-modal tasks...
within a single cell provides a clue as to what this delay activity represents. If visual cortex mediates only visual representations, delay activity elicited by nonvisual stimuli (as in the auditory-to-visual task) could represent the anticipated visual stimuli that are being cued. However, this explanation will not suffice for the same cell’s selective delay activity in the visual-to-auditory task in which an auditory stimulus is anticipated. In that case, one might argue that delay activity is a persisting representation of the visual sample stimulus, but then the cells representation must change from test to sample between tasks. It is more parsimonious to view delay-period activity in IT as encoding pairings or groupings of associated objects. The frequent failure of neurons with selective delay-period activity to display selective sensory responses is consistent with this idea of not simply coding specific physical characteristics.

The hypothesis that delay activity in this study represented associations is consistent with our finding of a long-term representation encoded in delay activity in one subject of this study. Although Sakai and Miyashita (1991) report a similar phenomenon for unimodal visual associations in IT, we show that it can apply to cross-modal associations. Just as in Sakai and Miyashita, this representation probably did not exist before training.

Although Sakai and Miyashita (1991) observed that paired associates tended to elicit correlated levels of delay activity, paired associates in the current experiment produced negatively correlated activity. This was probably a consequence of having only two associated pairs in our tasks. We defined the correct response to associated stimuli (visual A and auditory A, or visual B and auditory B) as a lever release. With only two stimuli for each modality, however, it was possible for the animal to infer that the correct response to associated stimuli was to not release the lever, providing that it paired visual A with auditory B and visual B with auditory A. The task could be correctly completed using either set of rules. Because the delay period activity in animal 1 was consistent with the latter formulation, we infer that the animal settled on it. Based on studies showing consistent mnemonic effects produced by associated stimuli (Maunsell et al. 1991; Sakai and Miyashita 1991), if three or more pairs were used in this experiment, there likely would have been a straightforward correlation between operationally defined pairs.

Although both animals studied had delay selectivity in the cross-modal tasks, far fewer neurons with this property were found in animal 2. Despite our regular sampling intervals and fairly extensive sampling range in this animal, one possibility is that we undersampled the relevant region of cortex. IT is large, and there are no reliable neurophysiological markers for specific subdivisions within it. Miyashita and Chang (1988) have reported that delay-selective cells tend to cluster into localized spots, and Fuster and Jervey (1982) reported that most delay-selective cells resided locally in the ventral bank of the STS. We likewise observed such clustering in animal 1 and sampled most delay-selective neurons from that region. It is possible that the sampling in animal 2 may not have been adequate to find a similar focal point.

The difficulty in finding delay-selective cells representing associations is further demonstrated in Sakai and Miyashita (1991) who, using 12 pairs of associated visual stimuli, only found that 2% of their sample displayed this delay activity phenomenon. It is also possible that the neural representation of the task was different in animal 2, and did not involve neurons with selective delay-period activity in IT. Neurons with delay activity might have existed in a different brain region or the task might have been solved without delay period selectivity being obvious in the firing rates of individual neurons. A difference in neural representation for almost identical tasks, and its connection with behavioral strategy, has been reported by Miller and Desimone (1994).

The differences between the two subjects in the current study may be related to differences in the incidence of neurons with delay activity between previous studies of IT. Delay-selective cells can be difficult to find in IT, with some investigations failing to find any (Baylis and Rolls 1987; Eskandar et al. 1992; Mikami and Kubota 1980; Miller et al. 1993; Riches et al. 1991). Other reports that describe delay-period activity have involved large numbers of animals, which may be needed to make a sufficient number of observations (Fuster 1990; Fuster and Jervey 1982).

The apparent difficulty in finding delay-selective cells in these previous studies also might be due to differences in behavioral paradigm. Animals in these studies performed simple visual matching tasks, whereas we had animals performing a number of memory tasks, including difficult auditory tasks (Colombo and D’Amato 1986; D’Amato and Salmon 1984). Neuronal activity in visual cortex has been shown to be affected by task difficulty (Spitzer and Richmond 1991). Task difficulty could play a role in the strength...
of delay activity observed in various studies. Also, Miyashita (1988) found that stimulus familiarity may be important in determining whether delay activity is seen. He found that although familiar stimuli evoked selective delay activity, novel stimuli evoked little, even though the animal performed the task with the novel stimuli just as well.

Because delay activity is not seen in all studies, there is a question as to how relevant this phenomenon is to short-term memory. One way of showing relevance is by correlating delay activity with behavioral performance. For example, Colombo and Gross (1994) reported that the amount of delay activity in IT (both selective and nonselective) correlates with behavioral performance in a short-term memory task. Wilson et al. (1990) showed that delay activity in IT cells can predict behavioral responses, but this was in the context of a simple delayed response task and could have represented activity related to motor set (see Kurata and Wise 1988).

We demonstrated a connection between delay activity and short-term memory in two ways. First, just as Vogels et al. (1995) showed, delay activity disappeared with passive viewing of the exact same stimulus sequence that occurred in the tasks. Second, in the auditory-to-visual task, we found that delay activity could predict robustly correct and incorrect responses, a property that has not been previously demonstrated in sensory cortex for motor-independent memory. Studies have shown this property for DMS tasks in prefrontal cortex (Funahashi et al. 1989; Watanabe 1986) with isolated examples but not so clearly and systematically across a large population of cells. Because the passive viewing and incorrect trials data were only possible from one animal subject in this study, how this data applies to all animal subjects is not known.

The same neurons that predicted behavioral responses during the auditory-visual task did not predict responses during the visual-auditory task. One explanation for this difference is that the visual sample stimuli were processed more accurately and reliably than the auditory sample stimuli. Both the difficulty in training animals to do auditory tasks and their poorer performance in the auditory matching task support this idea (Colombo and D’Amato 1986; D’Amato and Salmon 1984). Mistakes in both the cross-modal tasks might occur mainly in evaluating the auditory stimuli. This would affect delay selectivity in the auditory-to-visual task because mistakes would have been made at the time the sample was presented. On the other hand, the visual samples would have been processed correctly in the visual-auditory task, with mistakes made when the auditory test stimulus was presented. Thus in the visual-to-auditory task, delay-period activity would reflect accurately the sample stimulus identity, and not predict the impending error.

The current data are also consistent with previous studies in showing how delay activity and sensory responses are two distinct processes. First, selectivity for the two occurred fairly independently (Colombo and Gross 1994; Ferrera et al. 1994; Fuster and Jervey 1982; Haenny et al. 1988; Maunsell et al. 1991; Miyashita and Chang 1988). Second, although selective sensory responses remained purely visual, selective delay activity could involve memory for both auditory and visual stimuli and thus represent a form of multimodal sensory integration in visual cortex (Haenny et al. 1988; Maunsell et al. 1991). Finally, we also show that a clear long-term memory representation for cross-modal associations observed in delay activity can exist without any similar representation in sensory responses.

One corollary of sensory processing and short-term memory processing being independent processes is that delay selectivity does not necessarily represent visual imagery. Support for this stems from animal I in which the long-term representation seen in delay activity in the cross-modal tasks had no consistent relationship with the sensory responses seen in those tasks. For instance, Sakai and Miyashita (1991) have shown that a sensory response to a visual stimulus correlates with delay activity evoked by the stimulus paired associate in a unimodal visual association task. No such relationship was observed for the cross-modal associations in this study. Thus even if the animal was using a form of imagery to mediate short-term memory, this imagery did not share the same representation as that evoked by viewed stimuli.

Although selective delay activity exists in IT, this does not mean that short-term memory is mediated there. Delay activity for visual-visual associations has been observed in prefrontal cortex (Funahashi et al. 1993; Goldman-Rakic 1991; Watanabe 1990) and for cross-modal associations in the supplementary motor area (Tanji and Kurata 1985). Thus delay activity in IT may just be part of a more distributed brain representation for short-term memory. Our results suggest that short-term memory representations, while potentially diffuse in nature, also can be specific for certain brain areas because no selective delay activity was observed in IT during the purely auditory task. Selective delay activity triggered by auditory sample stimuli in the cross-modal tasks could originate through pathways that include projections from multimodal association areas that border IT or from medial temporal structures that are thought to play a critical role in associational memory (Felleman and Van Essen 1991; Morel and Bullier 1990; Suzuki and Amaral 1994; Van Hoesen 1982).

We thank G. M. Ghose, C. E. Landisman, and C. J. McAdams for helpful comments on preliminary versions of the manuscript and B. Noerager for technical assistance.

This research was supported by National Eye Institute EY-05911 and grants from the Office of Naval Research and the McKnight Foundation.

Address for reprint requests: J. R. Gibson, Dept. of Neuroscience, Brown University, Providence, RI 02912.

Received 27 February 1997; accepted in final form 5 June 1997.

REFERENCES


BROWN, M. W., WILSON, F. A. W., AND RICHES, I. P. Neuronal evidence that inferomedial temporal cortex is more important than hippocampus in
certain processes underlying recognition memory. *Brain Res.* 409: 158–


Hikosaka, O., Sakamoto, M., and Usui, S. Functional properties of mon-


