The role of temporal cortical areas in perceptual organization

(electrophysiology/extrastrate/binocular vision/monkeys/visual perception)

D. L. SHEINBERG* AND N. K. LOGOTHETIS†

Division of Neuroscience, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030

Communicated by F. Crick, Salk Institute for Biological Sciences, San Diego, CA, January 24, 1997 (received for review December 2, 1996)

ABSTRACT The visual areas of the temporal lobe of the primate are thought to be essential for the representation of visual objects. To examine the role of these areas in the visual awareness of a stimulus, we recorded the activity of single neurons in monkeys trained to report their perceptions when viewing ambiguous stimuli. Visual ambiguity was induced by presenting incongruent images to the two eyes, a stimulation condition known to instigate binocular rivalry, during which one image is seen at a given time while the other is perceptually suppressed. Previous recordings in areas V1, V2, V4, and MT of monkeys experiencing binocular rivalry showed that only a small proportion of striate and extrastriate neurons discharge exclusively when the driving stimulus is seen. In contrast, the activity of almost all neurons in the inferior temporal cortex and the visual areas of the cortex of superior temporal sulcus was found to be contingent upon the perceptual dominance of an effective visual stimulus. These areas thus appear to represent a stage of processing beyond the resolution of ambiguities—and thus beyond the processes of perceptual grouping and image segmentation—where neural activity reflects the brain’s internal view of objects, rather than the effects of the retinal stimulus on cells encoding simple visual features or shape primitives.

Neurons in the visual areas of the anterior temporal lobe of monkeys exhibit pattern-selective responses that are modulated by visual attention and are affected by the stimulus in memory, suggesting that these areas play an important role in the perception of visual patterns and the recognition of objects (1, 2). To understand the role of these areas in perception and object vision, we conducted combined psychophysical and electrophysiological experiments in monkeys experiencing binocular rivalry. Binocular rivalry refers to the stochastic changes of perception when one is viewing two different patterns dichoptically. We have recently shown that the perceived image during rivalry is independent of which eye it is seen through (3), a finding that suggests that binocular rivalry may be the result of competition between different stimulus representations throughout the visual cortex, rather than between the two monocular channels early in striate cortex (for review see ref. 4). The study of cell activity during binocular rivalry may therefore provide us with significant insights regarding the neural sites and mechanisms underlying the perceptual multistability experienced when one is viewing any ambiguous figures, such as the well studied figure-ground reversals, and may lead to a better understanding of the principles of perceptual organization.

METHODS

Two animals (Macaca mulatta) participated in the experiments reported in this paper. After the monkeys were familiarized with the laboratory environment and the experimenter, they underwent an aseptic surgery (5, 6). After recovery, the monkeys were trained to fixate a light spot and to perform a categorization task by pulling one of two levers attached to the front of their primate chair. They were taught to pull and hold the left lever whenever a sunburst-like pattern (left-object) was displayed and to pull and hold the right lever upon presentation of other figures, including images of humans, monkeys, apes, wild animals, butterflies, reptiles, and various manmade objects (right-objects). In addition, they were trained not to respond or to release an already pulled lever upon presentation of a physical blend of different stimuli (mixed-objects).

Example stimuli are shown in Fig. 1. The patterns were generated using a graphics computer (Indigo2, Silicon Graphics) and were presented on a display monitor placed 97 cm away from the animal. Stereoscopic presentations were accomplished using a liquid crystal polarizer (NuVision SGS19S) that allowed alternate transmission of images with circularly opposite polarization at the rate of 120 frames per sec (60 frames per sec for each eye). Polarized glasses were worn to allow the passage of only every other image to each eye.

During the behavioral task, individual observation periods consisted of random transitions between presentations of left-, right-, and mixed-objects. Juice reward was delivered only after the successful completion of an entire observation period. However, negative feedback was always given to the monkeys in the form of aborting an observation period following an incorrect response. Once the animals had learned to classify the different object types rapidly and accurately, periods of rivalrous stimulation (7–20 sec) were introduced in observation periods lasting 15–30 sec. During rivalrous periods, no feedback was given to the monkeys. Eye position was constantly monitored and stored. Excursions of the eyes outside of a ±0.75° window surrounding the fixation spot automatically aborted the observation period.

Single-cell activity was recorded in both monkeys in the upper and lower banks of superior temporal sulcus (STS) and the inferior temporal cortex (IT) using of a chamber consisting of a ball-and-socket joint with a 18-gauge stainless steel tube passing through its center (7). The base of the well was secured to the skull using small skull-screws and bone cement. The position of the guide-tube could be varied before each experimental session in any direction using a calibration device, attachable to the outer part of the ball-and-socket joint. The placement of the chambers was aided by a set of x-ray images combined with a set of magnetic resonance images (2.4-Tesla Magnet; Bruker, Billerica, MA) acquired before the head-post surgery of each monkey. We recorded from three hemispheres.

Abbreviations: STS, superior temporal sulcus; IT, inferior temporal cortex; SDF, spike density function.

*e-mail: sheinb@bcm.tmc.edu.

†To whom reprint requests should be addressed. e-mail: nikos@bcm.tmc.edu.

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

Copyright © 1997 by THE NATIONAL ACADEMY OF SCIENCES OF THE USA

0027-8424/97/043408-6$2.00

PNAS is available online at http://www.pnas.org.

Vol. 94, pp. 3408–3413, April 1997
Neurobiology
in two monkeys with the chambers placed at AP = 20, L = 20; AP = 19, L = 20; and AP = 19, L = 19, respectively. By swiveling the guide tube, different sites could be accessed within an $8 \times 8$ mm$^2$ cortical region. Since both monkeys are still alive and participating in similar experiments, the recording areas were estimated from the stereotaxic coordinates of the guide tube and the white-to-gray matter transitions expected from magnetic resonance images. According to these estimates, the recording sites were probably in areas TPO1, TPO2, and TEa and in the gyral portion of IT, most likely areas TEM, TE1, and TE2.

**RESULTS**

Because the interpretation of the neurophysiological data of this study strongly depended on the reliability of the animals' behavioral responses, special care was taken to ensure that the monkeys were reporting their perceptions accurately, rather than alternately pulling the levers in a random fashion. To encourage reliable performance, each observation period consisted of randomly intermixed periods of rivalrous and nonrivalrous stimulation, during which left-objects and right-objects were displayed monocularly. The slightly lustrous appearance of a monocularly viewed image served to maximize the similarity of percepts elicited by nonrivalrous and rivalrous stimulation and to reduce the chances of the monkey adopting different behavioral strategies in the two different stimulation conditions. Moreover, to train the monkey to report only exclusive visibility of a figure, mixed-objects, mimicking piecemeal rivalry, were randomly intermixed within each observation period. The monkeys reliably withheld response during these mixed periods, even when such periods constituted an entire observation period.

Finally, we systematically compared the monkeys' psychophysical performance with that of humans in the same tasks. During binocular rivalry, the time for which different stimuli are perceived depends strongly on the images' relative stimulus strength, a term specifying the combined effect of such stimulus parameters as luminance, contrast, spatiotemporal frequency, and amount of contour per stimulus area (8). For our task, we varied stimulus strength by changing the spatial frequency content of one image in the stimulus pair by lowpass filtering it. In humans, limiting the spatial frequency content of an image has been shown to decrease the stimulus' predominance (9), where predominance of a stimulus is typically defined as the percentage of the total viewing time during which this stimulus is perceived (8). Since our stimuli were large enough ($2.5 \times 2.5$) to often instigate piecemeal rivalry, predominance of the stimulus was defined to be the ratio of the time for which one stimulus was exclusively visible to the total time for which either stimulus was exclusively visible.

Fig. 2 shows the remarkable similarity in the dependency of predominance of a visual pattern on its spatial frequency content in both monkeys and humans. We take the consistency in both sets of data as strong evidence for the reliability of the monkeys' behavior.

Following the initial behavioral training, we began the combined psychophysical-physiological experiments. We isolated 159 visually responsive single units. Responsiveness was determined by presenting stimuli from a battery of hundreds of visual images. The selectivity of these cells was tested by repeatedly presenting a subset of the available visual stimuli in pseudorandom order in search of one or more effective stimuli, while the monkey fixated a central light spot.

Example responses of an IT neuron are shown in Fig. 3. The cell discharges action potentials upon presentation of the effective stimuli, here images of particular butterflies, and responds minimally to all other tested stimuli (including the sunburst pattern). Of the visually responsive neurons, 50 were found to be selective enough to be tested during the object classification task under both rivalrous and nonrivalrous conditions. The rivalry stimuli were created by presenting the effective stimulus to one eye and the ineffective stimulus (i.e.,

---

**Fig. 1.** Example stimuli used during the experiments. Stimuli consisted of geometrical sunburst patterns (left-objects), images of animate objects (right-objects), and physical blends of images that were used to mimic piecemeal rivalry (mixed-objects). The monkeys were trained to pull the left lever whenever the left-objects were visible, the right-lever whenever right-objects were visible, and neither lever when mixed-objects were visible.

**Fig. 2.** Behavioral verification of monkey's performance during rivalry. (A) Each pair of images depicts a stimulus condition, wherein the image of the face remained unchanged while that of the sunburst was blurred, to various degrees, by lowpass filtering. Filtering was achieved by multiplying the amplitudes of forward Fourier transformed images by an exponential gain and then converting back to the space domain. The lowpass cutoffs shown below each image refer to the frequency at which the exponential filter was equal to 1/e. (B) Predominance of a stimulus as function of spatial frequency bandwidth. Predominance is defined here as $T_{sunburst}/(T_{sunburst} + T_{face})$, where $T_{sunburst}$ and $T_{face}$ are the time durations for which the sunburst and the face were exclusively visible. (Left) Data from monkeys and (Right) data from experimentally naive human subjects. Note that predominance is systematically related to the spatial frequency content of the sunburst pattern for both monkeys and humans; as the sunburst is blurred to greater extents, it is perceived dominant for a decreasing proportion of time.
the sunburst) to the other. Fig. 3B shows two observation periods during this task, one from each monkey. Each plot illustrates the stimulus configuration, the neuron’s activity, and the monkey’s reported percept throughout the entire observation period. In both cases, the neuron discharged only before and during the periods in which the monkey reported seeing the effective stimulus. During rivalrous stimulation, the stimulus configuration remained constant, but significant changes in cell activity were accompanied by subsequent changes in the monkeys’ perceptual report.

The neural activity was further analyzed by constructing average spike density functions (SDFs), sorted by the monkey’s perceptual reports. Fig. 4A shows these data for the same cell depicted in the Fig. 3B Upper. Fig. 4A Upper and Lower show responses in nonrivalrous and rivalrous conditions, respectively. As shown in Fig. 3, this neuron fired vigorously when the monkey reported seeing the cell’s preferred pattern in both the nonrivalrous and rivalrous conditions. However, when the monkey reported seeing the ineffective stimulus, the cell response was almost eliminated, even when the effective stimulus was physically present during rivalry.

To increase the instances of exclusive visibility of one stimulus, and to further ensure that the monkey’s report accurately reflected which stimulus he perceived at any given time, we also tested the psychophysical performance of the monkeys and the neural responses of STS and IT cells using the flash suppression paradigm (10). In this condition, one of the two stimuli used to instigate rivalry is first viewed monocularly for 1-2 sec. Following the monocular preview, rivalry is induced by presenting the second image to the contralateral eye. Under these conditions, human subjects invariably perceive only the newly presented image and the previewed stimulus is rendered invisible. Previous studies have shown that the suppression of the previewed stimulus is not due to forward masking or light adaptation (10) and that instead it shares much in common with the perceptual suppression experienced during binocular rivalry (11). In our experiments, the monkeys, just like the human subjects,
consistently reported seeing the stimulus presented to the eye contralateral to the previewing eye during the flash suppression trials.

To confirm that the animals responded only when a flashed stimulus was exclusively dominant, catch trials were introduced in which mixed stimuli were flashed, after which the monkey was required to release both levers. Performance for both animals was consistently >95% for this task. Fig. 4B shows the activity of an STS neuron in the flash suppression condition. Fig. 4B Upper shows the cell responses for monocular presentations, and the Fig. 4B Lower shows the neuron’s activity at the end of the monocular preview (to the left of the dotted vertical line) and when perceptual dominance is exogenously reversed as the rival stimulus is presented to the other eye (to the right of dotted vertical line). The cell fires vigorously when the effective stimulus dominates perception and ceases firing entirely when the ineffective stimulus is made dominant. To better understand the differences between the temporal areas and the prestriate areas, recordings were also performed in area V4 using the flash suppression paradigm (D. Leopold and N.K.L., unpublished observations). V4 neurons were largely unaffected by the perceptual changes during flash suppression. Presenting the ineffective stimulus after priming with the effective one caused no alteration in the firing rate of any of

Fig. 4. Cell activity sorted by the dominant percept during nonrivalrous and rivalrous conditions. (A Upper) Averaged responses to the monocularly presented ineffective and effective stimuli. Above each graph is a pictorial representation of the visual stimuli presented. At time zero, depicted by the dotted line, the stimulus changed from either a blank screen or a mixed-object (data not shown) to the ineffective (Left) or effective (Right) stimulus. The cell fired only in response to the butterfly pattern. Presentation of the sunburst had little or no effect on the neuron’s activity. (Lower) Response of the cell just before and after the onset of rivalrous stimulation, with the effective stimulus presented to one eye and the ineffective to the other. The data are sorted based the monkey’s perceptual report: trials in which the monkey first reported seeing the ineffective stimulus (Left) and those for which the monkey first reported seeing the effective stimulus (Right). In these conditions, the stimuli presented are identical, but the recorded cell response correlates well with the monkey’s reported percept. (B) Data collected using the suppression paradigm. The nonrivalrous trials (Upper) show that this cell consistently responded to the effective stimulus and not at all to the ineffective stimulus. The flash suppression trials are similar to the rivalry trials shown in A, except that preceding the rivalrous stimulation, either the effective stimulus (Lower Left) or ineffective stimulus (Lower Right) was previously presented monocularly. Rivalry onset, marked by the dotted vertical line, thus consisted of adding either the ineffective or effective stimulus to the rivalrous pair. Following rivalry onset, the monkey’s reported percept consistently switched to the newly presented stimulus, and the previewed stimulus was perceptually suppressed. Using this paradigm, phenomenal suppression was especially effective, and cell activity during the onset of rivalrous stimulation closely mirrored that during the nonrivalrous controls.

Neurobiology: Sheinberg and Logothetis
the cells; presenting the effective stimulus after priming with the other had an weak effect on a small percentage of V4 neurons.

Across the population of cells from which we recorded, we found significant differences in the temporal structure of individual neural responses. Some neurons responded in a sustained fashion, while others exhibited a periodic burst or very transient response (Fig. 5A). We were concerned that typical methods of characterizing cell response, such as counting the number of spikes occurring within a fixed time window, would ignore these potentially informative variations. We thus characterized the entire spike waveforms for each trial using a well established method of dimensionality reduction and then applied multivariate statistical tests on the data to test for

**FIG. 5.** (A) Examples of different response types of neurons in STS and IT. While some cells elicited relatively sustained responses to visual stimuli (e.g., cells r115 and n034), others exhibited a periodic bursting behavior (e.g., n039 and r105), or a highly transient response (e.g., r027 and n083). (B) Mean normalized cell responses in the nonrivalrous (30 cells), rivalrous (24 cells), and flash suppression (33 cells) conditions, for the effective (solid line) and ineffective (dotted line) trials. In all conditions, average cell response in the effective trials was elevated over response during the ineffective trials. (C) Scatter diagram of average cell responses for all tested neurons. For visualization purposes, only projections of the response vectors onto the two first components, C1 and C2, are presented. Each marker represents the mean of all ineffective (○) and effective (□) trials for a given cell. The distance (exemplified by the solid lines for three of the cells) between almost all pairs of responses are statistically significant (see below). (D) Separation of mean responses to the perceived effective and ineffective stimuli for the three stimulation conditions. Each individual response is represented by an eight-element vector. Separation is given by the Mahalanobis distance (12), \( D = \sqrt{(\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2)} \), where \( \mu_1 \) and \( \mu_2 \) are the mean response vectors for the effective and ineffective trials, respectively, and \( \Sigma \) is the covariance matrix of the eight-dimensional response vectors. Because the two response types usually had different variances, \( \Sigma \) was replaced by its unbiased estimate, \( \Sigma = ((n_1 + n_2)/n_1)\Sigma_1 + ((n_1 + n_2)/n_2)\Sigma_2 \), where \( n_1 \) and \( n_2 \) are the number of presentations of each stimulus type, and \( n = n_1 + n_2 \). The significance of this variance-weighted distance was assessed by means of the Hotelling’s \( T^2 \) statistic (13) given by \( (n_1\Sigma_1/n_1)(\mu_1 - \mu_2)'\Sigma_1^{-1}(\mu_1 - \mu_2) + n_2\Sigma_2^{-1} = (n_1\Sigma_1/n_1)D^2 \), where \( p \) stands for the dimensionality of the response space. The numbers in the top right of each plot show the proportion of cells for which the response in the ineffective and effective trials was significantly different at the \( \alpha = 0.05 \) level. It should be noted that the high percentages of modulating neurons reported here was due to the specific multivatiate analysis. Similar results, in terms of proportions of significantly modulating cells, were obtained by the more traditional analysis of counting the number of spikes occurring in individual trials. Computing mean rates, however, requires arbitrary decisions pertaining the time window over which these rates must be computed when neurons show highly variable temporal modulations.
differences in cell response between the ineffective and effective trials (13). A detailed description of the analysis methods is given elsewhere (14). Briefly, the spike train for each trial was defined as a discrete function over the interval

\[ [0, N - 1], \]

where \( N \) was the number of points in the peristimulus time window (for population analysis, 800 points spaced 1 msec apart). The spike function takes the value 1 if a spike occurs at point \( t \), with \( t \in [0, N - 1] \), and zero otherwise. Each trial’s SDF was computed using the adaptive-kernel estimation process (15). These SDFs were subjected to principal components analysis (16), which is an orthogonal transform that typically results in a description of the data in a response space with strongly reduced dimensionality, and whose basis vectors, called the principal components, are uncorrelated (and can thus be studied independently of one another) and ordered to represent decreasing proportions of the total variance of the data. In this study, the principal components of cell responses were extracted using the variances and covariances of subsampled (every 5 msec) SDFs, after centering the data (the mean SDF for a stimulus or report condition was subtracted from individual SDFs). Response vectors for individual trials were calculated by projecting a given SDF onto each of the leading principal components. For these data, a maximum of eight components was required to explain at least 75% of the cumulative response variance, and thus an eight-dimensional space was used to represent each cell’s response to the two different perceptual conditions.

Fig. 5B shows that, on average, cell response was consistently higher for those trials in which the effective stimulus dominated perception compared with trials in which the ineffective stimulus dominated. Fig. 5C depicts each cell’s mean response for the ineffective and effective trials, as projected into the first two dimensions of the eight-dimensional space used to analyze the data. In these graphs, each cell is represented twice, once for effective trials and once for ineffective trials. The separation of these populations at the individual cell level is further quantified in Fig. 5D, which shows a histogram of separations, in units of standard deviation, of each cell’s ineffective and effective response vectors. Overall, \( \approx 90\% \) of the recorded cells in STS and IT were found to reliably predict the perceptual state of the animal. The proportion of cells showing statistically significant separations between the effective and ineffective conditions are shown in the top right of each plot in Fig. 5D.

The reliability of a given response pattern in predicting the animal’s perceived stimulus was also tested by comparing the performance of a statistical pattern classifier with that of the monkey. Two eight-dimensional subspaces were generated by extracting the principal components of the responses to the effective and ineffective stimulus in the nonrivalrous trials. Individual responses in the rivalry trials were then assigned to one or the other subspace by using a minimum-distance statistical pattern classifier. On average, 78.5% (range = 66–91%) of the monkey’s reported percept was predicted by this trial by trial classification method.

**DISCUSSION**

These results show that the activity of the vast majority of studied temporal cortex neurons is contingent upon the perceptual dominance of an effective visual stimulus. Neural representations in these cortical areas appear, therefore, to be very different from those in striate and early extrastriate cortex. Only 18% of the sample in striate cortex (5) and \( \approx 20\% \) and 25% of the cells in areas MT and V4, respectively (5, 6), were found to increase their firing rate significantly when their preferred stimulus was perceived. Moreover, one-fifth of the studied MT neurons and 13% of V4 neurons responded only when the effective stimulus was phenomenally suppressed, while other cells showed response selectivity only during perceptual rivalry and not while the animal was involved in passive fixation. The different response types in these areas may be the result of the feedforward and feedback cortical activity that underlies the processes of grouping and segmentation—processes that are probably perturbed when ambiguous figures are viewed. If so, the areas reported here may represent a stage of processing beyond the resolution of ambiguities, where neural activity reflects the integration of constructed visual percepts into those subsystems responsible for object recognition and visually guided action.

It is worth considering how the present data can be interpreted in light of the growing body of literature concerning so-called attentional modulation of cortical activity (1, 17, 18). Indeed, paradigms employed in studies of visual selective attention bear great similarity to the rivalry paradigm, in that more than one competing stimuli is generally presented to the subject and the effects of this competition are closely monitored. These experiments have often found that the activity of cells in visual cortex is both a function of the visual stimulus and of the animal’s set or state, indicating that other neural processes—generally referred to as attention—can influence cell activity above and beyond that which can be explained by the visual stimulus alone. Our view is that the phenomenon of binocular rivalry is also a form of visual selection, but that this selection occurs between competing visual patterns in the absence of explicit instructions to attend to one stimulus or the other. Decades of research have failed to reliably demonstrate that the perceptual alternations experienced during rivalry are under the direct control of voluntary attention. As such, we believe that rivalry accentuates the selective processing that underlies basic perceptual processes including image segmentation, perceptual grouping, and surface completion. In this view, the modulation of cortical activity reported here may be of distinct origin from the modulatory effects reported for tasks in which attention is overtly directed to one stimulus or another. Nonetheless, it is striking that both the effects of modulation due to rivalry and to attention have been reported in many of the same visual cortical areas. It will be of great interest to see if and how the same neurons participate in both phenomena.

We thank David Leopold for many useful suggestions and comments on the manuscript. This research was supported by the National Institutes of Health Grants NIH 1R01EY10089-01 to N.K.L. and NRSA 1F32EY086624 to D.L.S.