Processing of Shape Defined by Disparity in Monkey Inferior Temporal Cortex

HIROKI TANAKA,[1,2] TAKANORI UKA,[2,3] KENJI YOSHIYAMA,[3] MAKOTO KATO,[2,3] AND ICHIRO FUJITA[1–3]

[1]Laboratory for Cognitive Neuroscience, Division of Biophysical Engineering, Graduate School of Engineering Science, Osaka University; [2]Core Research for Evolutional Science and Technology, Japan Science and Technology Corporation, Osaka 560-8531; and [3]Department of Cognitive Neuroscience, Osaka University Medical School, Osaka 565-0871, Japan

Tanaka, Hiroki, Takanori Uka, Kenji Yoshiyama, Makoto Kato, and Ichiro Fujita. Processing of shape defined by disparity in monkey inferior temporal cortex. *J Neurophysiol* 85: 735–744, 2001. Neurons in the monkey inferior temporal cortex (IT) have been shown to respond to shapes defined by luminance, texture, or motion. In the present study, we determined whether IT neurons respond to shapes defined solely by binocular disparity, and if so, whether signals of disparity and other visual cues to define shape converge on single IT neurons. We recorded extracellular activity from IT neurons while monkeys performed a fixation task. Among the neurons that responded to at least one of eight random-dot stereograms (RDSs) containing different disparity-defined shapes, 21% varied their responses to different RDSs. Responses of most of the neurons were positively correlated between two sets of RDSs, which consisted of different dot patterns but defined the same set of eight shapes, whereas responses to RDSs and their monocular images were not correlated. This indicates that the response modulation for the eight RDSs reflects selectivity for shapes (or their component contours) defined by disparity, although responses were also affected by dot patterns per se. Among the neurons that showed selectivity for shapes defined by luminance or disparity, 44% were activated by both cues. Responses of these neurons to luminance-defined shapes and those to disparity-defined shapes were often positively correlated to each other. Furthermore the stimulus rank, which was determined by the magnitude of responses to shapes, generally matched between these cues. The same held true between disparity and texture cues. The results suggest that the signals of disparity, luminance, and texture cues to define the shapes converge on a population of single IT neurons to produce the selectivity for shapes.

INTRODUCTION

Binocular disparity is a positional difference between the left and right retinal images of an object. When binocular disparity in a retinal region differs from that in its surroundings, shape is perceived binocularly, even though it cannot be from monocular images (Julesz 1971). This indicates that binocular disparity is a sufficient cue for the perception of shape.

In studies on binocular vision, binocular disparity has been focused on as a cue for the perception of depth rather than that for shapes (Regan 1991; Wheatstone 1838). Previous physiological studies aimed at determining the neural mechanisms of stereopsis, which computes depth from binocular disparity

cues, have revealed that many visual cortical areas in primates, including areas V1, V2, V3, V4, VP, MT, and MST, areas in the posterior parietal cortex, and the inferior temporal cortex (IT) contain disparity-selective neurons (Burkhalter and Van Essen 1986; Felleman and Van Essen 1987; Hubel and Wiesel 1970; Janssen et al. 1999; Maunsell and Van Essen 1983; Poggio and Fischer 1977; Poggio et al. 1988; Roy et al. 1992; Sakata et al. 1997; Uka et al. 2000; Watanabe et al. 2000), and examined the roles of these neurons in depth perception (Bradley et al. 1998; Cumming and Parker 1997, 2000; DeAngelis et al. 1998; Prince et al. 2000) or representation by these neurons of three-dimensional surface structure (Janssen et al. 1999, 2000a,b; Shikata et al. 1996; Taira et al. 2000; Uka et al. 1997). On the other hand, studies on the neural processing of two-dimensional shape based on disparity cues are very limited. von der Heydt et al. (2000) reported that some disparity-selective V2 neurons respond to edges defined by disparity in an orientation-selective manner. The edge information may be integrated into shape information in some area of the brain, although it is not known how neurons represent shapes defined by disparity.

The IT, the final stage of the ventral visual pathway in monkeys, is considered as being critically involved in shape processing (Mishkin et al. 1983). Many neurons in this area respond selectively to shapes; they prefer some shapes over others (Desimone et al. 1984; Gross et al. 1972; Tanaka et al. 1991). Sáry et al. (1993) showed that a population of IT neurons responded not only to shapes defined by luminance but also to those defined by texture or motion. They also showed that the shape selectivity of these neurons was similar for the cues that define shape. In the present study, we attempted to determine whether IT neurons responded to shapes defined by disparity and, if so, whether signals of disparity and other visual cues to define shape converge on the single IT neurons. We recorded extracellularly the responses of neurons to eight random-dot stereograms (Julesz 1971), which contained different shapes defined by disparity, and to the same sets of shapes defined by luminance or texture. We show that some IT neurons were selective for shapes defined by disparity. Shape selectivity of these neurons tended to be similar between disparity and other cues, suggesting that signals of different cues

to define shapes converge on these neurons to show their shape selectivity. Preliminary results have appeared elsewhere (Tanaka et al. 1999).

METHODS

Subjects and surgery

Two male monkeys (*Macaca fuscata*, 9 and 5 kg body wt) were used. In the first surgery, a scleral search coil was implanted under the conjunctiva of one eye to monitor eye position (Judge et al. 1980), and a head post was attached to the skull using acrylic screws and dental cement to allow head fixation. After a recovery period of >2 wk followed by 2–3 mo of training in a fixation task, an eye coil was implanted in the other eye, and a recording chamber was attached to the right side of the skull over the temporal cortex. After the monkeys received supplementary training, neuronal recordings were started. In *monkey 1,* another recording chamber was attached to the skull over the left temporal cortex after the recordings from the right IT were completed. All surgical procedures were performed under surgical anesthesia (pentobarbital sodium, 35 mg/kg ip) and aseptic conditions. After each surgery, the monkeys were administered an antibiotic (piperacillin sodium, 30 mg/kg im), analgesic (ketroprofen, 0.5 mg/kg im), and corticosteroid (dexamethasone sodium phosphate, 0.1 mg/kg im) to minimize potential inflammation. Surgical procedures and animal care conformed to the Guidelines of the National Institutes of Health for the Care and Use of Laboratory Animals (1996) and were approved by the animal experiment committee of Osaka University Medical School.

Task and stimulus presentation

The monkeys were trained in a fixation task controlled by a computer (PC486FS: EPSON, Suwa, Japan). They were seated on a primate chair facing a 15-in color CRT monitor (frame rate, 70 Hz; size, 260 × 195 mm; resolution, 1,024 × 768) placed 57 cm away. The monkey's head was fixed by screwing the head post to the chair. Stimuli were presented on the monitor using a computer (Asus Computer International, San Jose, CA). For *monkey 1,* the positions of both eyes were sampled at a rate of 100 Hz using the search coil technique (Judge et al. 1980) and stored for off-line analysis, although the position of one eye was monitored on-line. For *monkey 2,* the position of only one eye was sampled and monitored. Each trial was started with the presentation of a fixation spot (0.2 × 0.2°) at the center of the monitor. The monkeys were trained to fixate on it for 500 ms (fixation window: 2 × 2°). A visual stimulus was then presented for 2 s, and the monkeys were rewarded with a drop of water if they maintained their fixation for this duration. Otherwise the trial was aborted the moment they broke their fixation. After the monkeys were returned to their cages, they received an adequate amount of fruit. During the training and experimental sessions, the monkeys were deprived of water but were allowed dry food ad libitum in their cages.

Visual stimuli

A static random-dot pattern consisting of bright and dark dots covered the entire screen of the monitor. Fifty percent of the dots were bright. Each dot occupied 2 × 2 pixels subtending a visual angle of 0.05 × 0.05°. The luminances of the bright and dark dots were 38 and 0.2 cd/m^2, respectively. The stimulus set consisted of eight shapes (Fig. 1*A*) subtending a visual angle of 2°. Each shape was defined by a difference in disparity, luminance, or texture between the shape region and its surroundings (Fig. 1*B*).

RANDOM-DOT STEREOGRAMS (RDSs). Disparity-defined shapes were generated by adding a crossed disparity of 0.2° to the shape region (Fig. 1*B, top* and *middle*). Dots in the shape region of the left-eye image were horizontally shifted relative to those of the right-eye image. The right-eye
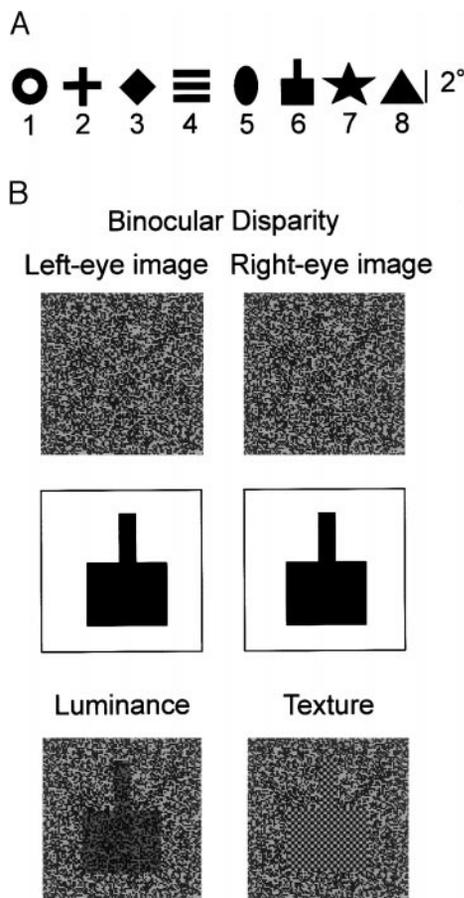

FIG. 1. Stimulus set. *A*: the 8 shapes used in the present experiment. The numbers under the shapes are used to aid identification of the shapes in the following figures. *B*: 3 different cues to define shape. The schematic illustration under "Binocular Disparity" shows the area where a horizontal disparity (0.2°) was added in random-dot steriograms (RDSs). Luminance-defined shapes (LUMs) and texture-defined shapes (TEXs) did not have an area added with crossed disparity. They contained only 0 disparity.

image was identical for all stimuli. A liquid crystal stereoscopic modulator and polarized glasses were used for dichoptic stimulation (Tektronix SGS610). In control experiments where we examined whether the differential responses of IT neurons to the different RDSs were due to response selectivity for shapes defined by disparity or caused by slightly different dot patterns in the left-eye images, only the left-eye images were presented to the left eye.

LUMINANCE-DEFINED SHAPES (LUMs). Luminance-defined shapes were constructed by making the bright dots inside the shapes darker than those in the surroundings (Fig. 1*B, bottom left*). The luminance of the bright dots in the shape region was 20 cd/m^2, yielding a contrast between the shape region and its surroundings of 50%.

TEXTURE-DEFINED SHAPES (TEXs). The dots in the shape region were four times as large as those in the surroundings (0.1 × 0.1°) and were arranged in a regular checkerboard pattern (Fig. 1*B, bottom right*). The average luminance of the shape region was the same as that of its surroundings. LUMs and TEXs did not have crossed disparity of 0.2° but contained only 0 disparity.

When visual stimuli were presented, the dot pattern in the central 7.5 × 5° rectangular region containing the shape region was changed with no correlation of dot position between the prestimulus and stimulus patterns. The shape disappeared when the arrangement of dots in this central region was returned to the original pattern. This procedure was adopted to avoid apparent motion of dots, something that could be an additional cue for the perception of shapes.

To distinguish whether neuronal responses to the visual stimuli were induced by the disparity-defined shapes or by the change of dot patterns in the central rectangular region, we recorded the neuronal responses when the rectangular region of dots contained no shape (referred to as the "no-shape pattern"). The no-shape pattern was identical to the dot pattern of the central $7.5 \times 5°$ region of the RDSs presented to the right eye. This dot pattern was also identical to that of the central $7.5 \times 5°$ region of all stimuli for LUMs and TEXs, except for the shape region. Therefore any difference between the neuronal responses to a visual stimulus and the no-shape pattern was considered as being evoked by the shape contained in the visual stimulus.

All the neurons, except for two, which we tested with only LUMs and RDSs, were tested with 24 stimuli (8 shapes $\times$ 3 cues) and the no-shape pattern shown in a random sequence. Additional control stimuli were then presented as long as the recording remained stable.

Extracellular recording

The activity of IT neurons (mostly single units and some multiunits) was recorded extracellularly from three hemispheres of the two monkeys using glass-coated elgiloy electrodes (tip, 15–40 μm; impedance, 2–3 MΩ at 1 kHz). The electrodes were controlled by a microdrive (MO-95 s, Narishige, Tokyo) attached to the recording chamber. The electrodes penetrated the dura mater into the lateral surface of the IT. The recorded signals were amplified, band-pass-filtered, and fed to a window discriminator and an oscilloscope. The triggered spikes were sampled in a computer (PC486FS: EPSON, Suwa, Japan). A rastergram and a peristimulus time histogram were plotted and displayed on-line. The timings of spike occurrence and the behavioral responses were stored for off-line analysis.

Histology

After the neuronal recording sessions were completed, *monkey 1* was anesthetized with an overdose of pentobarbital sodium (60 mg/kg ip), the chest cavity was opened, and heparin (200 IU/kg) was injected into the heart. The animal was transcardially perfused with 500 ml of phosphate-buffered saline (PBS, 37°C), followed by perfusion of 2 l of ice-cold 4% paraformaldhyde in 0.1 M PBS. We implanted two pins in the brain at the anterior and posterior edges of the recording chambers on both the right and left sides. The brain was then removed, photographed, blocked, postfixed overnight in the fixative, and immersed in 0.1 M PBS containing a graded series of sucrose (10–30%). The location of the implanted pins was verified for reconstruction of the recording area. *Monkey 2* is still alive and participating in a different experiment.

Data analysis

The spontaneous firing rate was calculated from the average spike count over 500 ms before stimulus onset. The net response was calculated by subtracting the spontaneous firing rate from the firing rate during a 2-s period starting 80 ms after stimulus onset. The net response was averaged across stimulus repetitions (5–10 times). This value was used as a measure of the neuronal responses to the stimulus.

To examine whether the strength of the responses to the stimulus was significantly different from spontaneous firing, we used a *t*-test (2-tailed, $P < 0.05$). We also used a *t*-test to determine whether the responses to shape stimuli were significantly different from those to the no-shape pattern. To determine whether response modulation to different shapes was statistically significant, one-way ANOVA was performed. Other analyses will be mentioned where relevant. The significance level in all tests was $P < 0.05$.

RESULTS

Recording sites

Histological analysis showed that our recording site in *monkey 1* was in the central portion of the dorsal IT (striped area in Fig. 2). The recording region included area TEd and possibly the most anterior part of area TEO. The two dots in Fig. 2 show the location of the pins implanted as histological landmarks.

Monkey 2 is still alive. Since the recording chamber was attached to a position on the skull just over the temporal cortex, similar to that in *monkey 1*, we believe that the recordings were made from a similar portion of the IT.

Database

We recorded from 225 units in three hemispheres of the two monkeys ($n = 166$ in *monkey 1*, $n = 59$ in *monkey 2*). All the units were tested with RDSs, LUMs, and TEXs except two, which were tested with only the first two. We sampled neurons that responded to at least one of the 24 stimuli (i.e., 8 shapes $\times$ 3 cues). If a neuron responded to the no-shape pattern and the response was not different from any of the responses to the visual stimuli, the neuron was considered to respond to the dot pattern and was discarded from further analysis. Using this criterion, 201 units responded to at least one of the RDSs, LUMs or TEXs. Of these, 105 units responded to RDSs, 142 responded to LUMs, and 138 responded to TEXs. All these stimuli generally evoked excitatory responses, although the ratio of excitatory to inhibitory responses was slightly different across the three cues (Table 1).

In Table 2, we classified the 201 units into the following seven groups according to the cues that evoked responses: units that responded to RDSs alone, LUMs alone, TEXs alone, both RDSs and LUMs, both RDSs and TEXs, both LUMs and TEXs, and RDSs, LUMs and TEXs. Sixty-eight percent of these units responded to more than two cues, and 24% responded to three cues.

Of the 105 units responding to the RDSs, 22 (21%) showed statistically significant response modulation (or, response selectivity) for different RDSs. Eighteen were single isolated neurons, while the other four were multiple neurons. Of the 142 units responding to LUMs, 66 (46%) showed response selectivity to LUMs (53 single neurons). Of the 138 units responding to TEXs, 50 (36%) showed response selectivity to TEXs (39 single neurons).
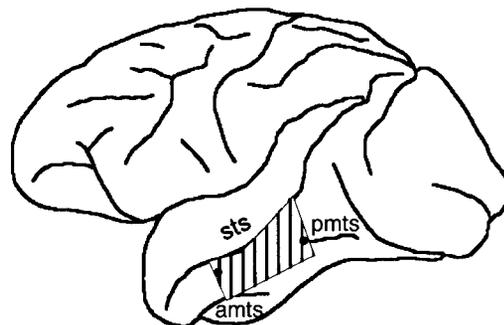

FIG. 2. Recording site (▨) shown in the lateral view of a left cortex (*monkey 1*). The 2 dots indicate the location of the implanted pins (see METHODS). amts and pmts, anterior and posterior middle temporal sulcus, respectively; sts, superior temporal sulcus.

TABLE 1. *Response distribution of the recorded cells*

Cue	Excitation	Inhibition	No Response
RDS	82 (41)	23 (11)	96 (48)
LUM	119 (59)	23 (11)	59 (30)
TEX	115 (57)	23 (12)	63 (31)

Each entry shows the number of units and its percentage (in parentheses) among the 201 units that responded to ≥1 of 24 stimuli. RDS, random-dot stereogram (disparity); LUM, luminance; TEX, texture.

Response selectivity for RDSs

Figure 3 shows an example of an IT neuron that responded to RDSs in a stimulus-selective manner. This neuron showed excitatory responses to five of the eight stimuli with a latency of ~150 ms. The magnitude of responses differed across different shapes (ANOVA, $P < 0.01$) with the maximum response to a square (*stimulus 3*; 10.8 spikes/s). Since this neuron showed a similar response modulation for LUMs (see Fig. 6), the response modulation for RDSs was considered to be largely due to the differences of the shapes defined by disparity and not due to the differences of the dot patterns in different RDSs.

The preferred shape and the degree of tuning differed among the units that showed response selectivity for RDSs. *Neuron A* in Fig. 4 responded only to a doughnut shape (*stimulus 1*), and little or no response was evoked by the other shapes. ANOVA revealed a highly significant modulation of responses by different shapes ($P < 0.0005$). *Neuron B,* responding to six different stimuli, showed significant, but rather broad, response selectivity (ANOVA, $P < 0.01$). To quantify the sharpness of the response selectivity for RDSs, we calculated the tuning width. This was defined as the number of stimuli to which a neuron responded with more than half the maximum response magnitude (taking a value from 1 to 8). The tuning width of *neuron A* was 1, indicating that this neuron belonged to the group of neurons showing the sharpest response selectivity. The tuning width of *neuron B* was 5. The median of the tuning width for the 17 single neurons that were selective for RDSs with excitatory responses was 2. Comparison of the distribution of the tuning widths for RDSs with that for LUMs (median = 3 for 47 single neurons that were selective for LUMs with excitatory responses) or TEXs (median = 2 for 31 single neurons that were selective for TEXs with excitatory responses) revealed no statistically significant differences (Mann-Whitney *U* test: $P > 0.9$ for LUMs vs. RDSs; $P > 0.9$

TABLE 2. *Classification of neurons according to effective visual stimuli*

	RDSs	LUMs	TEXs	Total
1 cue	11 (5)	26 (13)	28 (14)	65 (32)
	RDSs & LUMs	RDSs & TEXs	LUMs & TEXs	
2 cues	26 (13)	20 (10)	42 (21)	88 (44)
		RDSs & LUMs & TEXs		
3 cues		48 (24)		48 (24)

Each entry shows the number of units and its percentage (in parentheses) among the 201 units that responded to ≥1 of 24 stimuli.

for TEXs vs. RDSs). Thus the frequency distribution of tuning widths for RDSs was comparable with that for LUMs or TEXs.

Neurons A and B in Fig. 4 also exhibited a contrast in their response strength. While *neuron A* responded rather weakly, *neuron B* showed vigorous and sustained responses. The maximum response magnitude of *neuron B* was 29.2 spikes/s. This was the strongest response among the units selective for the RDSs. The average of the maximum response magnitude to the RDSs among the 17 single neurons with excitatory responses was 9.4 ± 6.1 (SD) spikes/s. The average of the maximum response magnitude to LUMs among the single neurons which were selective for LUMs with excitatory responses was 13.6 ± 8.5 ($n = 47$) and that to TEXs was 11.3 ± 8.4 spikes/s ($n = 31$). When we compared the response magnitude to RDSs with those to LUMs and TEXs, the magnitudes of the responses to LUMs tended to be higher than those to RDSs, although the difference between the maximum response magnitudes to LUMs and RDSs fell slightly short of the significance level (Mann-Whitney *U* test, $0.05 < P < 0.1$). Thus under the present stimulus condition, luminance tends to act as a stronger
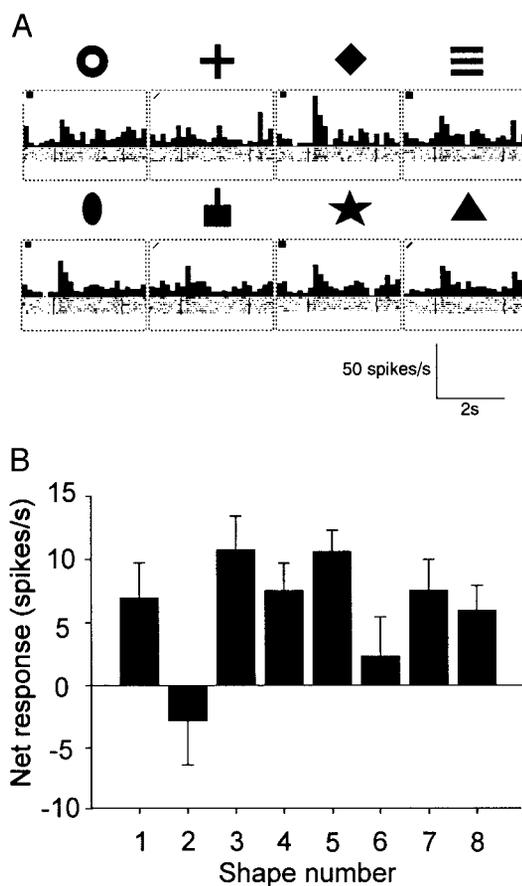

FIG. 3. An example of a neuron exhibiting varied responses to RDSs. *A*: peristimulus time histograms (PSTHs) and rastergrams showing the neural responses to 8 different RDSs. The 1st vertical line in the rastergrams indicates the onset of the stimuli, the 2nd line indicates the offset of the stimuli. A dot in the rastergrams indicates the occurrence of spikes in 30 ms. The width of a bin of the PSTHs corresponds to 150 ms. The small square at the *top left* indicates that the neuron exhibited a significant excitatory response compared with the neuronal activity during the 500 ms immediately before stimulus presentation (paired *t*-test, $P < 0.05$, $n = 7$). This neuron responded to 5 of the 8 RDSs. *B*: average net responses of the neuron shown in *A* to the 8 RDSs (error bar: SE). This neuron responded most strongly to the square (*stimulus 3*) and most weakly to the cross (*stimulus 2*).
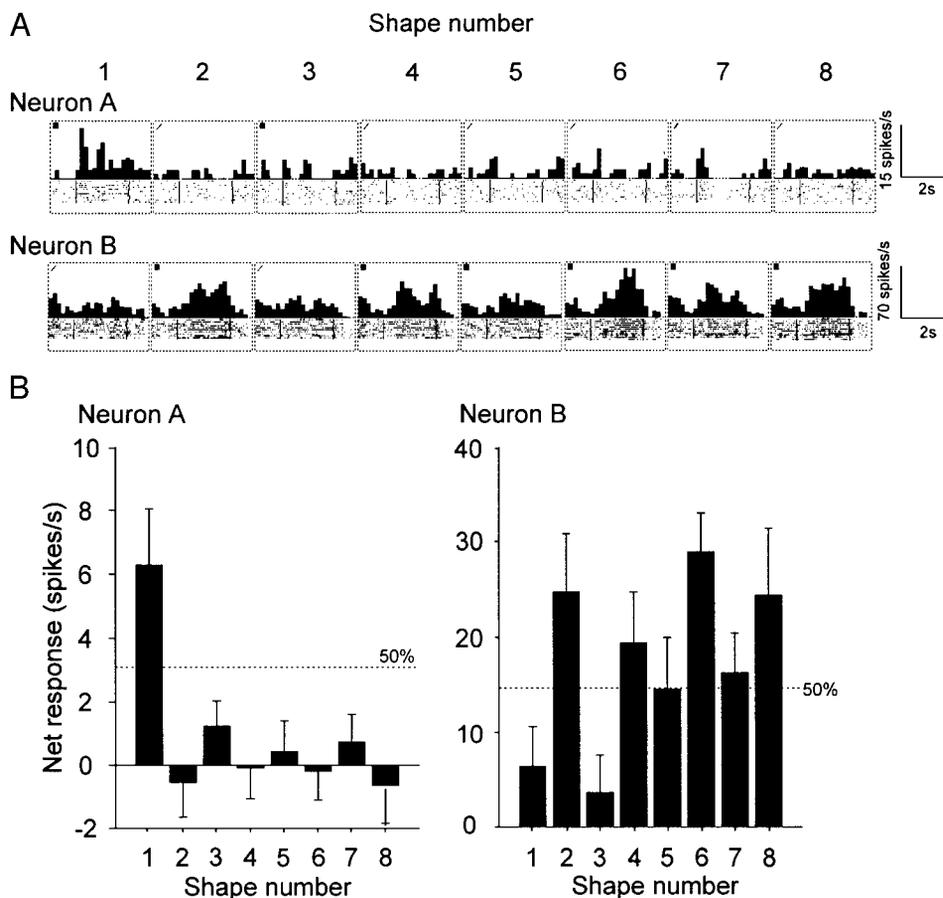
FIG. 4. Two examples of neurons which responded to RDSs with sharp (*neuron A*) and broad tuning (*neuron B*). *A*: PSTHs and rastergrams showing neural responses to 8 different RDSs. The format of the PSTHs and rastergrams is the same as that in Fig. 3. *Neuron A* responded only to the doughnut shape (*stimulus 1*), while *neuron B* responded to 6 different RDSs. These 2 neurons also differed in the magnitude of the responses. While *neuron A* responded relatively weakly, *neuron B* responded strongly. *B*: average net responses of the 2 neurons shown above (error bar: SE). · · ·, 50% of the maximum response.

cue to evoke stimulus-selective responses in IT neurons than disparity cues, which is consistent with the observation that the number of single neurons selective for LUMs ($n = 53$) was about three times that for RDSs ($n = 18$).

Responses to monocular images and to new RDSs consisting of different dot patterns

Seventeen of the 22 units selective for RDSs were tested with monocular images of the RDSs. Only the left-eye images were used because the right-eye image was identical across the eight RDSs (see METHODS). Figure 5*A* compares the responses of a single neuron to the RDSs (□) with those to the left-eye images (▲). The responses to the RDSs and the monocular images are sorted according to the magnitudes of the responses to the RDSs. The response profile to the monocular images was markedly different from that to the RDSs. Therefore the response modulation for the RDSs of this neuron was not explained by the responses to slightly different dot patterns in the monocular images. To evaluate the similarity between the two response profiles, we calculated Pearson's correlation coefficient between the two sets of eight responses (referred to as "response correlation"). In the 17 units tested, there was no response correlation between the RDSs and the monocular images (-0.01 ± 0.38, $n = 17$, $P > 0.9$, sign test). Hence, in general, the monocular responses to the dot patterns do not account for the response modulation for the RDSs.

Next we examined whether responses to the routinely used set of RDSs were correlated with those to a new set of RDSs consisting of totally different dot patterns, but defining the same set of eight shapes. Although the response magnitude to the new RDSs (Fig. 5*A*, ●) was lower than that to the original RDSs, the two response profiles showed strong positive correlation ($r = 0.76$, $P < 0.05$). Together with the finding that the response profile to the monocular images was different from that to the RDSs, the results showed that the response *profile* for the eight RDSs of this neuron was mainly based on the shape defined by disparity, although the response *magnitude* was affected by the dot pattern.

Of the 12 units tested with the two sets of RDSs, we calculated the ratio of the response magnitude for the more-effective RDSs to that for the less-effective RDSs. The average of the ratio was 0.62 ± 0.2. This indicates that, for many neurons, the dot pattern affects the response magnitude to the RDSs. However, response correlation between the two sets of RDSs was positively distributed (Fig. 5*B*, —, mean: 0.48, $P < 0.01$, sign test). The same was true when only single neurons were selected for analysis (- - -, $n = 9$, mean: 0.56, $P < 0.005$, sign test). We consider that the positive shift of the distribution reflects similarity between responses of a single neuron to the two sets of RDSs because such positive shift of the distribution of correlation coefficient was observed only when the two sets of data were from the same neuron. The correlation coefficient between responses to the original RDSs of one neuron and those to the new RDSs of the same or a different neuron was distributed around 0 (· · ·, mean: 0.09, $P > 0.2$, sign test, $n = 78$ pairs). There was a statistically significant difference between this distribution and the distribution of the response correlation in the same neuron (i.e., · · · vs. —, $P < 0.005$,
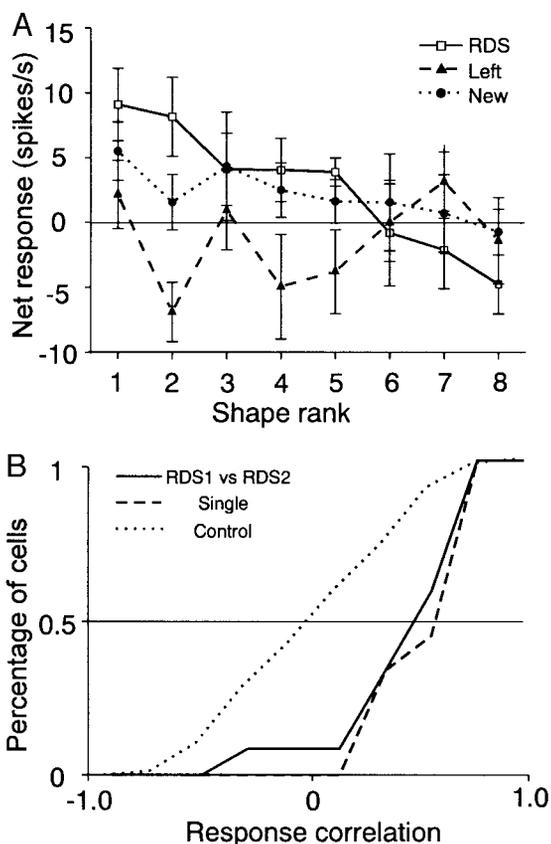
FIG. 5. Responses to the monocular images of RDSs and to a new set of RDSs consisting of a different dot pattern, but defining the same set of shapes as the original RDSs. *A*: an example of the neural responses to the routinely used (original) RDSs (□ — □), the left-eye images (▲ - - - ▲), and the new RDSs (● · · · ●). Error bars indicate SE. The responses to the 8 stimuli are sorted according to the magnitude of the responses to the original RDS. The neuron showed similar response modulation for the 2 sets of RDSs, but the response modulation for the monocular images was dissimilar to that for RDSs. *B*: cumulative histograms of response correlation. —, response correlation between 2 sets of RDSs for the 12 units that showed shape selectivity to RDSs. That for only single neurons is shown in dashed line ($n = 9$). · · ·, (control) response correlation between responses to the original RDSs of 1 neuron and responses to the new RDSs of the same or a different neuron among the 12 units ($n = 78$).

Mann-Whitney U test). In addition, there was no difference between the distribution of response correlation between two sets of RDSs in the same neuron (—) and the distribution of correlation coefficient between responses to LUMs and those to TEXs (- - - in Fig. 7*B*, $P > 0.1$, Mann-Whitney U test) to which IT neurons have been shown to exhibit similar shape selectivity (Sáry et al. 1993). These results support the views that modulation of responses to the eight RDSs, at least partially, represents response selectivity for the shapes defined by disparity.

Responses to the no-shape pattern

Because the central $7.5 \times 5°$ region changed its dot pattern at stimulus presentation, the change of the dot pattern could also affect the neuronal activity in addition to the shapes defined by disparity. However, because we analyzed only neurons that did not respond to the no-shape pattern or neurons whose responses to the shapes were different from their responses to the no-shape pattern, these factors were considered

not to have a substantial effect on the responses of the 22 units that were selective for RDSs. The average responses to the no-shape pattern among the 22 units was 0.77 spikes/s. This value was less than one-tenth the average maximum response magnitudes for the RDSs (8.5 spikes/s, $n = 22$), and the difference between these two values was significant (Wilcoxon paired-rank test: $P < 0.0001$, $n = 22$).

Vergence eye movement

We analyzed vergence eye position of *monkey 1* during recordings of activities of 16 neurons that were selective for RDSs. Except for one neuron, there was no statistically significant differences in the average vergence eye position during 2 s of stimulus presentation across different RDSs (ANOVA, $P < 0.05$).

Vergence eye movement around the onset of the visual stimuli was calculated by subtracting the average eye position over 500 ms before stimulus presentation from that during 2 s of stimulus presentation. The vergence eye movement was, on average, $-0.08 \pm 0.1°$, $n = 16$. We did not find statistically significant differences in the average vergence eye movement across different RDSs except for one neuron (ANOVA, $P < 0.05$). Thus it is unlikely that the neuronal selectivity for RDSs was caused by vergence eye movement.

Cue-invariant shape coding between disparity and other cues

We next examined whether responses to disparity-defined shapes and those to shapes defined by other cues were similar to each other. Pearson correlation coefficient was used to evaluate the similarity. Only single neurons were used for the analysis in the following text.

Luminance versus disparity

Figure 6 shows an example of a neuron whose shape selectivity to RDSs and LUMs was similar. This is the neuron shown in Fig. 3. The responses to LUMs were nearly twice as strong as those to RDSs for most of the stimuli (Fig. 6*B*). However, the response profiles were similar in that the stimulus rank was largely preserved between the two cues, and the two response curves were strongly correlated ($r = 0.83$, $P < 0.01$).

Sixty-three neurons showed shape selectivity for LUMs or RDSs. Of these, 28 (44%) were activated for both cues. The distribution of the response correlation for LUMs and RDSs of this cue-convergent group of neurons is shown in a cumulative histogram (Fig. 7*A*, —). The distribution was shifted toward positive values (mean = 0.26, $n = 28$) compared with that for responses to RDSs of one neuron and those to LUMs of the same or a different neuron among this group (· · ·, mean = 0.03, $n = 406$, Mann-Whitney U test, $P < 0.005$). Furthermore the distribution was similar to the distribution for LUMs and TEXs (- - -, mean: 0.27, $n = 33$ single neurons that showed shape selectivity for LUMs or TEXs and showed excitatory responses for both, Mann-Whitney U test, $P > 0.9$) to which IT neurons have previously been shown to respond in a cue-invariant manner (Sáry et al. 1993). The results indicate that shape selectivity tends to be similar for luminance and disparity cues among the cue-convergent group of neurons.
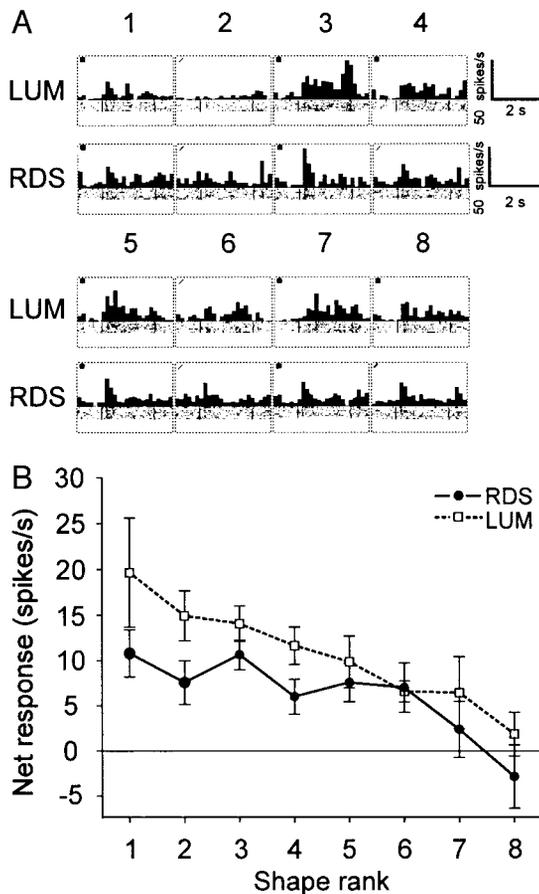
FIG. 6. An example of a neuron whose response profiles were similar for LUMs and RDSs. *A*: PSTHs and raster-histograms. This is the neuron shown in Fig. 3. This neuron responded most strongly to the square (*stimulus 3*) both for LUMs and RDSs. *B*: net average responses of the neuron for LUMs and RDSs (error bar: SE). The responses are sorted according to the response strength for LUMs. The 2 response profiles showed a strong correlation ($r = 0.83$, $P < 0.01$).

To further evaluate to what extent shape selectivity to LUMs and RDSs matched among the cue-convergent group, we calculated the average responses to the RDSs as a function of the shape rank determined by the responses to the LUMs (Fig. 7*B*). For each neuron, shapes were ranked according to the magnitude of the responses to the LUMs, and the responses to the RDSs were sorted in this rank order. Then the responses to the RDSs were normalized by the maximum response to the RDSs for each neuron. Finally the responses to RDSs were averaged over the population of neurons in each rank. The average rank response to the RDSs decreased almost monotonically as the rank order determined by the responses to LUMs became lower. The correlation between the rank and the average rank response to the RDSs was significant [Spearman's correlation coefficient: -0.23, $P < 0.001$, $n = 224$ (28 neurons × 8 shapes)]. Therefore shape selectivity for RDSs and LUMs, on average, matched to the extent that the ranks of shape preference were maintained.

Texture versus disparity

Figure 8 shows an example of a neuron whose shape selectivity for TEXs and RDSs was similar. This is *neuron A,* which was shown in Fig. 4. Although the response correlation for the

two cues did not reach a significance level ($r = 0.60$, $0.05 < P < 0.1$), this neuron responded predominantly to the doughnut shape for both cues.

Forty-eight neurons showed shape selectivity for TEXs or RDSs. Among them, 20 (42%) showed excitatory responses to both TEXs and RDSs. Again this cue-convergent group tended to show a positive response correlation for these cues (Fig. 9, —; mean, 0.33). The distribution of response correlation for TEXs and RDSs among the cue-convergent group was similar to that for LUMs and TEXs (- - -, Mann-Whitney *U* test, $P > 0.9$). The distribution was also shifted toward positive values compared with that between responses to TEXs of one neuron and responses to RDSs of the same or a different neuron among
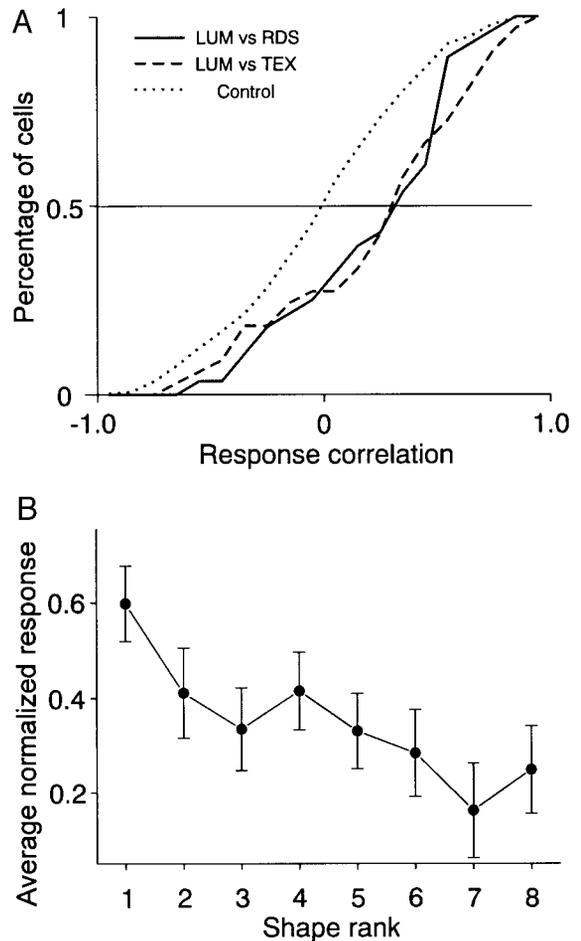

FIG. 7. *A*: cumulative histograms of response correlation. —, response correlation between LUMs and RDSs. The samples consisted of 28 single neurons that showed shape selectivity for LUMs or RDSs and excitatory responses for both cues (cue-convergent group). · · ·, (control) response correlation between responses to RDSs of 1 neuron and responses to LUMs of the same or a different neuron among this group ($n = 406$). - - -, response correlation between LUMs and TEXs ($n = 33$). *B*: average responses to RDSs across the 28 neurons in the cue-convergent group for LUMs and RDSs (error bar: SE). For each neuron, the shape rank was determined according to the magnitude of the responses to LUMs; i.e., the luminance-defined shape that activated a neuron most strongly was designated as rank 1, and the shape which activated the neuron least strongly was designated as rank 8. The responses to RDSs were sorted according to this rank order. The responses were then normalized by the maximum responses to RDSs. Finally, the responses on each rank were averaged across the 28 neurons in the cue-convergent group. The response curve decreases monotonically to the right, indicating that the shape preference for RDSs and LUMs was, on average, matched in this group of neurons.
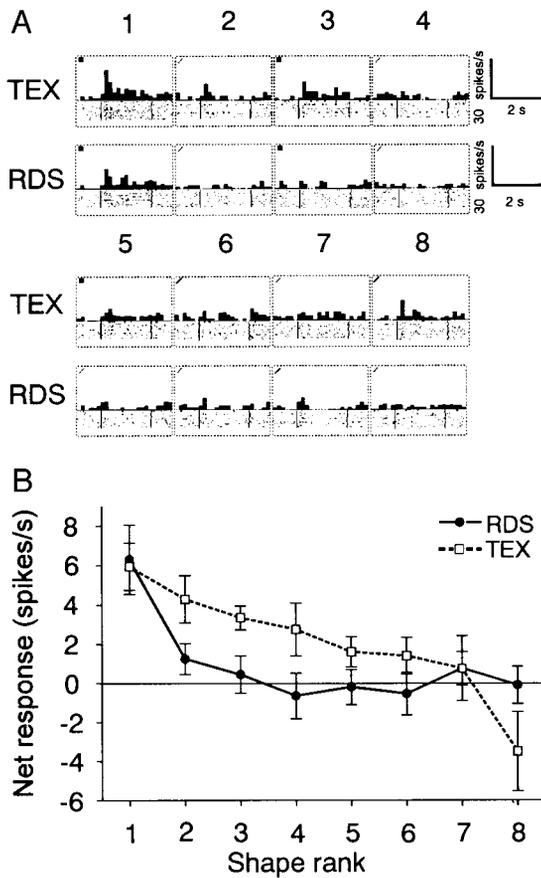
FIG. 8. An example of a neuron whose response profiles were similar for TEXs and RDSs. *A*: PSTHs and raster-histograms. This neuron is *neuron A* shown in Fig. 4. This neuron responded predominantly to the doughnut shape (*stimulus 1*) for both TEXs and RDSs. *B*: net average responses of the neuron in *A* to TEXs and RDSs (error bar: SE). For each cue, the responses to the 8 shapes are sorted according to the response strength for TEXs.

the cue-convergent group (· · ·; mean = 0.02, *n* = 210, Mann-Whitney *U* test, *P* < 0.005). The results indicate that shape selectivity tend to be similar for texture and disparity cues.

The average response to the RDSs decreased roughly monotonically as the rank of shape preferences determined by responses to TEXs became lower (Fig. 9*B*). The correlation between the rank order and the average response to RDSs was significant [Spearman's correlation coefficient: −0.32, *P* < 0.0001, *n* = 160 (20 neurons × 8 shapes)]. Thus the shape rank was maintained between RDSs and TEXs for the IT neurons.

DISCUSSION

A population of IT neurons showed differential responses to RDSs. Results of monocular test as well as test with a new set of RDSs indicate that the differential responses to the eight RDSs containing different shapes largely or at least partially represent selectivity for shapes defined by disparity. Shape selectivity tended to be similar for disparity and luminance cues in the neurons that showed excitatory responses for both cues. The same held true for disparity and texture cues. This indicates that signals of disparity cue and signals of luminance or texture cues to define shape converge on single IT neurons to show their shape selectivity. It should be noted, however,

that the response magnitude was different between the two sets of RDSs. The response magnitude was also dependent on the dot pattern of the RDSs.

Selectivity for RDSs

In the present experiment, 21% of the neurons that respond to at least one of eight RDSs showed response selectivity for the RDSs. Before we discuss the neural processing of shape defined by disparity, we consider possible alternative interpretations for what the response selectivity reflects.

First, there is a possibility that the differential responses for RDSs were caused by different dot patterns in the left-eye images. The monocular images of different RDSs look all alike perceptually, but dots in the shape region of the left-eye images were shifted relative to those of the right-eye images. IT neurons might be sensitive to the subtle differences in the dot patterns in the left-eye images. This possibility is, however,
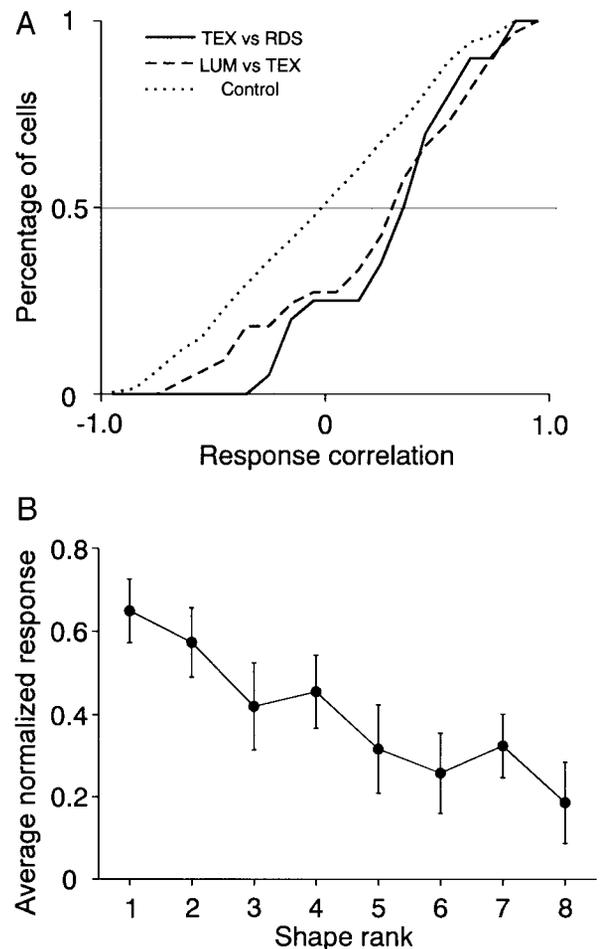

FIG. 9. *A*: cumulative histograms of the response correlation. —, response correlation between RDSs and TEXs. The samples consisted of 20 singles neurons that showed shape selectivity for TEXs or RDSs and excitatory responses to both (cue-convergent group). · · ·, (control) response correlation between responses to RDSs of a neuron and responses to TEXs of the same or a different neuron in these neurons (*n* = 210). - - -, response correlation between LUMs and TEXs (*n* = 33). *B*: average responses to RDSs across the 20 neurons in the cue-convergent group for TEXs and RDSs (error bar: SE). The rank order was determined according to the magnitude of responses to TEXs. The average responses decreased monotonically with a decrease in rank order.

ruled out because responses to RDSs and those to the left-eye images were dissimilar. In addition, responses for the two sets of RDSs, which consisted of totally different dot patterns but contained the same set of shapes, were positively correlated (Fig. 5B).

Second, it is possible that the neurons responded to the components of the shapes such as disparity of dots inside shape or contours defined by disparity. The first possibility is unlikely because responses to RDSs and those to shapes defined by other cues (which contained only 0 disparity) were correlated. On the other hand, it is difficult to determine strictly whether the neurons respond to the global shape or the component contours of the shape. For example, *neuron B* in Fig. 4 may have responded to a horizontal edge defined by disparity, which was common to shapes 2, 4, 6, and 8. To determine this, the "reduction process," in which a critical stimulus feature essential for neuronal activation is determined by stepwise decomposition of the effective stimulus, is necessary (Fujita et al. 1992; Tanaka et al. 1991). It remains to be established whether neurons selective for disparity-defined shapes responded to the global shapes or their component contours.

Shape from disparity

The neural processing of shape has been studied mostly using luminance-defined bars or shapes as stimuli. These studies have shown that local edge information is first detected in V1, and more complex features are then processed along the ventral visual pathway (Gallant et al. 1993; Hedgé and Van Essen 2000; Hubel and Wiesel 1968; Kobatake and Tanaka 1994).

Several studies have examined the mechanism for processing of shape based on texture or motion cues. Lesions in V4 of monkeys resulted in deficits in the discrimination of the orientation of texture-defined gratings (De Weerd et al. 1996), and some IT neurons were shown to respond to texture-defined shapes (Sáry et al. 1993). These results suggest that the ventral visual pathway processes shape based on texture cues. In regard to motion cues, it was found that some V4 and IT neurons responded to motion-defined gratings and shapes, respectively (Logothesis and Charles 1990; Sáry et al. 1993). Lesions in these areas impaired the ability to discriminate these stimuli (Britten et al. 1992; De Weerd et al. 1996; Schiller 1993). On the other hand, motion signals themselves are mainly processed in the dorsal visual pathway. Lesions in MT, which belongs to the dorsal pathway, also caused moderate deficits in the ability to discriminate shapes defined by motion (Schiller 1993). Hence both the ventral and dorsal visual pathways are thought to be involved in the analysis of shape based on motion cues.

Compared with studies on shape defined by texture and motion cues, even fewer studies have addressed the question of how and where shape based on disparity is processed. We showed in the present study that some IT neurons were selective for shapes defined by disparity. It remains unclear whether shape based on disparity cues is processed only along the ventral visual pathway because to our knowledge, no study has addressed this issue in areas in the dorsal visual pathway. von der Heydt et al. (2000) reported that some V2 cells were orientation selective for disparity-defined edges, but it was yet unclear whether such neurons were found predominantly in the

thin stripes or interstripes, which project to area V4. Since disparity-sensitive neurons are abundant in areas along the dorsal visual pathway (Maunsell and Van Essen 1983; Roy et al. 1992; Sakata et al. 1997), this pathway may also contribute to the processing of shape based on disparity cues. There are two possibilities of such contribution. First, this pathway merely provides disparity information to the ventral visual pathway. Second, neurons in the dorsal pathway per se represent shape defined by disparity. Recent reports have shown that some neurons in the posterior parietal cortex, the final stage of the dorsal pathway, responded to shapes defined by luminance (Murata et al. 1996; Sereno and Maunsell 1998; Taira et al. 1990). It is worthwhile examining whether these neurons also respond to shapes defined by disparity.

In the present study, we investigated the neural representation by IT neurons of two-dimensional "flat" shape defined by disparity. The stimuli we used are different from three-dimensional (3-D) "slanted" or "curved" surfaces defined by disparity gradient. Neurons selective for such disparity-defined surface-slant have been found in the caudal part of the lateral bank of the intraparietal sulcus (Shikata et al. 1996; Taira et al. 2000). It has recently been shown that IT neurons also respond to disparity-defined 3-D surface structures. Janssen et al. (2000b) reported that many neurons in the lower bank of the superior temporal sulcus respond to disparity-defined curved or slanted shape. Uka et al. (1997) have also reported that a population of IT neurons discriminated the depth order of two superimposed surfaces, irrespective of the type (i.e., crossed or uncrossed) of disparities added.

Convergence of signals of disparity and other cues on IT neurons

Sáry et al. (1993) found that shape selectivity of a population of IT neurons was similar for luminance, texture, and motion cues. In the present study, response selectivity for disparity-defined shapes tended to be similar to that for shapes defined by luminance or texture cues in the cue-convergent group of neurons.

However, not all IT neurons show such cue invariance. In our study, 56% of the neurons, which were shape-selective for disparity or luminance cues, were activated by only one of these cues. Similar results were observed between disparity and texture cues. This is not surprising if we consider that each visual attribute is, to some extent, processed separately in earlier visual cortical areas. Then how do signals of disparity and other cues to define shape converge on IT neurons? One possibility is that cue-dependent IT neurons converge onto another neuron in the IT. That is, an IT neuron that responds to shapes defined by disparity and another IT neuron that prefers similar shapes defined by other cues converge onto a third IT neuron. Another possibility is that the orientation selectivity for an edge is already cue-invariant in earlier visual cortices, and such cue-invariant information is conveyed to IT. A recent report by von der Heydt et al. (2000) showed evidence that supports the latter possibility. They reported that the orientation selectivity of some V2 neurons was invariant for disparity and luminance cues. If such neurons are abundant in V2 interstripes and V4 also contains such neurons, it is highly possible that the cue-invariant edge information created in these early areas is conveyed to IT. Further work is necessary

to reveal the underlying basis for cue invariance in shape representation in the visual system.

REFERENCES

BRADLEY DC, CHANG GC, AND ANDERSEN RA. Encoding of three-dimensional structure-from-motion by primate area MT neurons. *Nature* 392: 714–717, 1998.

BRITTEN KH, NEWSOME WT, AND SAUNDERS RC. Effects of inferotemporal cortex lesions on form-from-motion discrimination in monkeys. *Exp Brain Res* 88: 292–302, 1992.

BURKHALTER A AND VAN ESSEN DC. Processing of color, form and disparity information in visual areas VP and V2 of ventral extrastriate cortex in the macaque monkey. *J Neurosci* 6: 2327–2351, 1986.

CUMMING BG AND PARKER AJ. Responses of primary visual cortical neurons to binocular disparity without depth perception. *Nature* 389: 280–283, 1997.

CUMMING BG AND PARKER AJ. Local disparity not perceived depth is signaled by binocular neurons in cortical area V1 of the macaque. *J Neurosci* 20: 4758–4767, 2000.

DEANGELIS GC, CUMMING BG, AND NEWSOME WT. Cortical area MT and the perception of stereoscopic depth. *Nature* 394: 677–680, 1998.

DESIMONE R, ALBRIGHT TD, GROSS CG, AND BRUCE C. Stimulus-selective properties of inferior temporal neurons in the macaque. *J Neurosci* 4: 2051–2062, 1984.

DE WEERD P, DESIMONE R, AND UNGERLEIDER LG. Cue-dependent deficits in grating orientation discrimination after V4 lesions in macaques. *Vis Neurosci* 13: 529–538, 1996.

FELLEMAN DJ AND VAN ESSEN DC. Receptive field properties of neurons in area V3 of macaque monkey extrastriate cortex. *J Neurophysiol* 57: 889–920, 1987.

FUJITA I, TANAKA K, ITO M, AND CHENG K. Columns for visual features of objects in monkey inferotemporal cortex. *Nature* 360: 343–346, 1992.

GALLANT JL, BRAUN J, AND VAN ESSEN DC. Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex. *Science* 259: 100–103, 1993.

GROSS CG, ROCHA-MIRANDA CE, AND BENDER DB. Visual properties of neurons in inferotemporal cortex of the macaque. *J Neurophysiol* 35: 96–111, 1972.

HEDGÉ J AND VAN ESSEN DC. Selectivity for complex shapes in primate visual area V2. *J Neurosci* 20: RC61, 2000.

HUBEL DH AND WIESEL TN. Receptive fields and functional architecture of monkey striate cortex. *J Physiol (Lond)* 195: 215–243, 1968.

HUBEL DH AND WIESEL TN. Stereoscopic vision in macaque monkey. *Nature* 225: 41–42, 1970.

JANSSEN P, VOGELS R, AND ORBAN GA. Macaque inferior temporal neurons are selective for disparity-defined three-dimensional shapes. *Proc Natl Acad Sci USA* 96: 8217–8222, 1999.

JANSSEN P, VOGELS R, AND ORBAN GA. Selectivity for 3D shape that reveals distinct areas within macaque inferior temporal cortex. *Science* 288: 2054–2056, 2000a.

JANSSEN P, VOGELS R, AND ORBAN GA. Three-dimensional shape coding in inferior temporal cortex. *Neuron* 27: 385–397, 2000b.

JUDGE SJ, RICHMOND BJ, AND CHU FC. Implantation of magnetic search coils for measurement of eye position: an improved method. *Vision Res* 20: 535–538, 1980.

JULESZ B. *Foundation of Cyclopean Perception.* Chicago: University of Chicago Press, 1971.

KOBATAKE E AND TANAKA K. Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J Neurophysiol* 71: 856–867, 1994.

LOGOTHETIS NK AND CHARLES ER. V4 responses to gratings defined by random textured motion (Abstract). *Invest Ophthalmol Visual Sci* 31, *Suppl*: 444, 1990.

MAUNSELL JH AND VAN ESSEN DC. Functional properties of neurons in middle temporal visual area of the macaque monkey. II. Binocular interactions and sensitivity to binocular disparity. *J Neurophysiol* 49: 1148–1167, 1983.

MISHKIN M, UNGERLEIDER LG, AND MACKO KA. Object vision and spatial vision: two cortical pathways. *Trends Neurosci* 6: 414–417, 1983.

MURATA A, GALLESE V, KASEDA M, AND SAKATA H. Parietal neurons related to memory-guided hand manipulation. *J Neurophysiol* 75: 2180–2186, 1996.

POGGIO GF AND FISCHER B. Binocular interaction and depth sensitivity in striate and prestriate cortex of behaving rhesus monkey. *J Neurophysiol* 40: 1392–1405, 1977.

POGGIO GF, GONZALEZ F, AND KRAUSE F. Stereoscopic mechanisms in monkey visual cortex: binocular correlation and disparity selectivity. *J Neurosci* 8: 4531–4550, 1988.

PRINCE SJ, POINTON AD, CUMMING BG, AND PARKER AJ. The precision of single neuron responses in cortical area V1 during stereoscopic depth judgments. *J Neurosci* 20: 3387–3400, 2000.

REGAN D. Spatial vision for objects defined by colour contrast, binocular disparity, and motion parallax. In: *Spatial Vision. Vision and Visual Dysfunction.* London: Macmillan Press, 1991, vol. 10, p. 135–178.

ROY JP, KOMATSU H, AND WURTZ RH. Disparity sensitivity of neurons in monkey extrastriate area MST. *J Neurosci* 12: 2478–2492, 1992.

SAKATA H, TAIRA M, KUSUNOKI M, MURATA A, AND TANAKA Y. The parietal association cortex in depth perception and visual control of hand action. *Trends Neurosci* 20: 350–357, 1997.

SÁRY G, VOGELS R, AND ORBAN GA. Cue-invariant shape selectivity of macaque inferior temporal neurons. *Science* 260: 995–997, 1993.

SHIKATA E, TANAKA Y, NAKAMURA H, TAIRA M, AND SAKATA H. Selectivity of the parietal visual neurons in 3D orientation of surface of stereoscopic stimuli. *Neuroreport* 7: 2389–2394, 1996.

SCHILLER PH. The effects of V4, and middle temporal (MT) area lesions on visual performance in the rhesus monkey. *Vis Neurosci* 10: 717–746, 1993.

SERENO AB AND MAUNSELL JHR. Shape selectivity in primate lateral intraparietal cortex. *Nature* 395: 500–503, 1998.

TAIRA M, MINE S, GEORGOPOULOS AP, MURATA A, AND SAKATA H. Parietal cortex neurons of the monkey related to the visual guidance of hand movement. *Exp Brain Res* 83: 29–36, 1990.

TAIRA M, TSUTSUI K-I, JIANG M, YARA K, AND SAKATA H. Parietal neurons represent surface orientation from the gradient of binocular disparity. *J Neurophysiol* 83: 3140–3146, 2000.

TANAKA H, UKA T, YOSHIYAMA K, KATO M, AND FUJITA I. Shape defined by disparity evokes responses in monkey inferior temporal neurons. *Soc Neurosci Abstr* 25: 918, 1999.

TANAKA K, SAITO H, FUKADA Y, AND MORIYA M. Coding visual images of objects in the inferotemporal cortex. *J Neurosphysiol* 66: 170–189, 1991.

UKA T, TANAKA H, KATO M, AND FUJITA I. Disparity sensitivity and representation of 3-dimentional surface structure in monkey inferior temporal cortex. *Soc Neurosci Abstr* 23: 2063, 1997.

UKA T, TANAKA H, YOSHIYAMA K, KATO M, AND FUJITA I. Disparity selectivity of neurons in monkey inferior temporal cortex. *J Neurophysiol* 84: 120–132, 2000.

VON DER HEYDT R, ZHOU H, AND FRIEDMAN H. Representation of stereoscopic edges in monkey visual cortex. *Vision Res* 40: 1955–1967, 2000.

WATANABE M, TANAKA H, UKA T, AND FUJITA I. Disparity selective neurons in area V4 of macaque monkeys. *Soc Neurosci Abstr* 26: 954, 2000.

WHEATSTONE C. Contributions to the physiology of vision. I. On some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philos Trans R Soc Lond* 2: 371–393, 1838.