Reprint Series
14 May 1993, Volume 260, pp. 995–997

SCIENCE

Cue-Invariant Shape Selectivity of Macaque Inferior Temporal Neurons

Gyula Sáry, Rufin Vogels,* and Guy A. Orban

Cue-Invariant Shape Selectivity of Macaque Inferior Temporal Neurons

Gyula Sáry, Rufin Vogels,* Guy A. Orban

The perception of shape is independent of the size and position of the shape and also of the visual cue that defines it. The same shape can be recognized whether defined by a difference in luminance, by motion, or by texture. Experiments showed that the shape selectivity of individual cells in the macaque inferior temporal cortex did not vary with the size and position of a shape and also did not vary with the visual cue used to define the shape. This cue invariance was true for static luminance and texture cues as well as for relative motion cues—that is, for cues that are processed in ventral and dorsal visual pathways. The properties of these inferior temporal cells meet the demands of cue-invariant shape coding.

Recently, it has been shown that the direction selectivity of cells in the primate extrastriate middle temporal area (MT) is generally form-cue–invariant (*1*). This may underlie the form-cue invariance in the perception of motion direction. Here, we present neurophysiological evidence that a population of cells in the macaque inferior temporal cortex (IT) (*2*) forms the neural correlate of a different type of perceptual invariance: the visual cue invariance of shape perception (*3*) (Fig. 1).

Neuropsychological observations in humans as well as in animals have shown that IT lesions cause severe impairments in visual shape discrimination (*4*). Also, single-cell recording studies have revealed that the responses of IT cells can be highly selective for visual shape, preferring some shapes over others (*5*, *6*). It has been shown that IT neurons keep their shape selectivity irrespective of changes in retinal image size, contrast sign, or position (*7*). These neuronal response invariances match the invariance in size, contrast, and position of shape perception, which suggests that the IT is involved in the processes that underlie the recognition of shapes and objects. These experiments have used shapes defined by a single visual cue, luminance contrast. We stimulated IT cells using shapes defined by one of three visual cues and determined whether their shape selectivity was cue-independent.

Neurons were recorded in the IT cortex of two male rhesus monkeys performing a fixation task (*8*). Sets of eight figures (Fig. 2) created by random dots were used as stimuli (*9*). Each figure could be defined by one of three cues: luminance difference, relative motion, or texture difference (Fig. 1). In the case of the luminance-defined

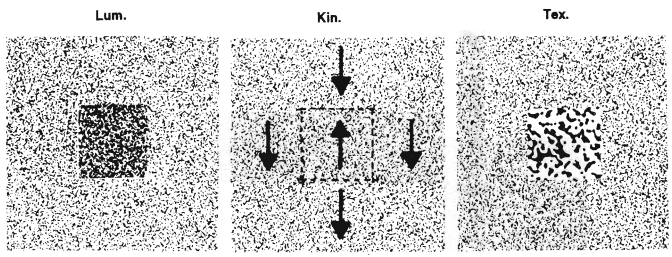

Fig. 1. Shape perception is cue-invariant. A square can be defined by a difference in luminance between figure and background (Lum.), by relative motion of the dots (up versus down) of the figure and background (Kin.), or by a difference in dot size (texture) between figure and background (Tex.). For the middle figure (Kin.), the arrows show the direction of motion of the dots and the hatching indicates the virtual borders, which are rendered visible solely by virtue of relative dot motion.

figures, dots inside the figure were darker than those outside the figure. For the motion-defined figures, dots inside the figure moved with the same speed but in a direction opposite that of the dots outside the figure. In the texture-defined figures, the dot size of the figure was three times the dot size of the background, but there was no difference in average luminance. A fourth set of figures in which the dots inside the figure moved with the same velocity as the background dots was used as a control for other dynamic cues, such as dynamic occlusion (*10*).

We tested IT cells (261) with each of the three types of figures (*11*). Sixty-five percent of the cells responded to each of the three types of cues, and 28% of these responsive neurons were shape-selective for each of the three cues. The cell shown in Figs. 2 and 3 preferred the three star-like figures out of the set of 24 luminance-, motion-, and texture-defined figures (Fig. 2). The same cell was also tested with the control figures (Fig. 3): The preference was the same, but its response was about half that for the kinetic figures, which indicates that the response to the kinetic stimulus arises at least partly from the figure defined by relative motion. This was a general finding because the average net response was larger for the kinetic than for the control figures in 26 neurons that had been tested with both control and kinetic figures and that had responded significantly to either stimulus (*12*). Also, this shape selectivity was not a result of some local peculiarity of the cell's receptive field because presentation of the figures at 4° eccentricity (*13*) had no effect on its stimulus selectivity for the motion- and luminance-defined figures (Fig. 3).

For the 47 cells that showed shape selectivity for the three visual cues, the average figure rank, determined by the response strength to the figures, was similar for the three visual cues (Fig. 4), which indicates that for this population of IT cells the preference and selectivity for shape is invariant with respect to the defining visual cue (*14*). For some of these cells, we determined whether the shape selectivity, in addition to being cue-invariant, was also independent of changes in the position and size of the shape. The results of these stimulus-invariance tests confirmed those of previous studies that reported position and size invariance of shape selectivity in the IT (*7*): All seven cells tested with peripheral presentation (4° to 5° eccentricity) of the figures showed the same selectivity as with foveal figure presentation, and for three out of four cells tested the shape selectivity was not affected by a 16-fold variation of stimulus area.

We do not know whether these cells respond to a component of the figure or to the overall shape ("gestalt"). Moreover, these results should not be interpreted as providing evidence for grandmother cell–like "figure detectors" or "shape-component detectors" because the response level of the cells is usually not only dependent on the shape but also on (the saliency of) the defining cue, the position, or size of the figure (*7*, *15*). However, a population of single units, such as those we recorded

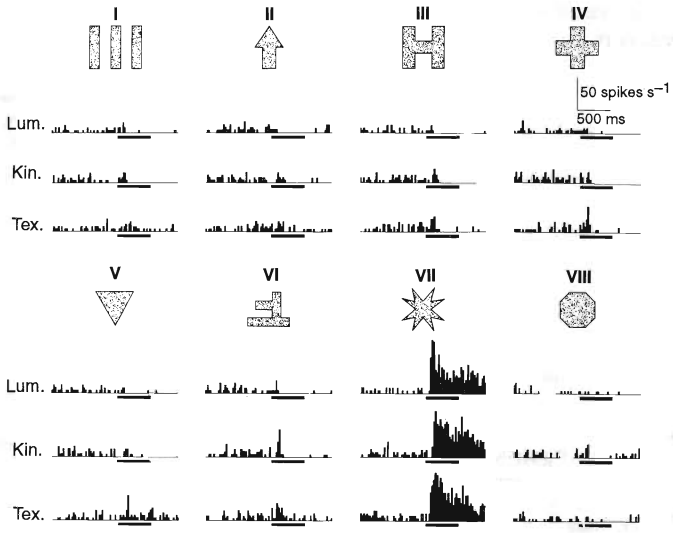

Fig. 2. Cue-invariant shape selectivity of a single inferior temporal unit. The figures used in the experiment are illustrated on the rows above the peristimulus time histograms. Lum., luminance-defined figures; Kin., relative motion–defined figures; and Tex., texture-defined figures. The duration of the stimulus presentation is indicated by the horizontal bar below each histogram (average of at least ten trials).

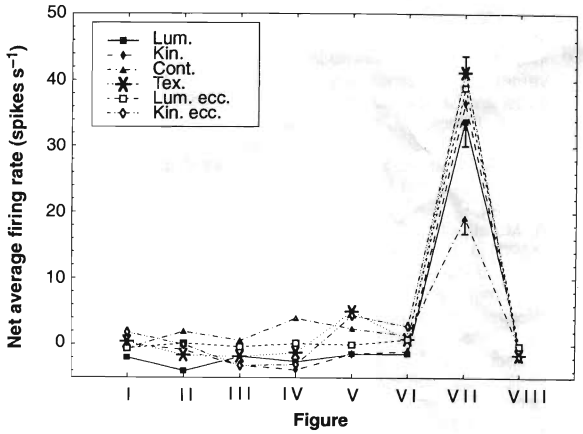

Fig. 3. Shape selectivity of a single inferior temporal unit. The figures have the same numbers as those in Fig. 2. Lum., foveal presentation of luminance-defined shapes; Kin., foveal presentation of motion-defined shapes; Cont., foveal presentation of control figures; Tex., foveally presented texture-defined shape; Lum. ecc., presentation of the luminance-defined figures at an eccentricity of 4° on the horizontal meridian in the contralateral visual field; Kin. ecc., peripheral presentation of the figure defined by relative motion. Each point corresponds to the mean net firing rate averaged over at least ten trials. All standard errors had a magnitude similar to the three shown.

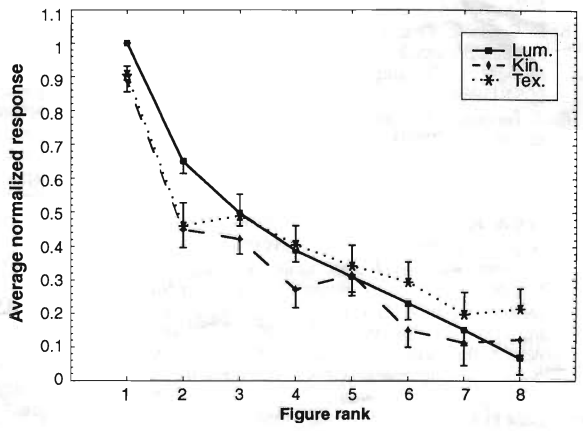

Fig. 4. The average normalized response and standard errors plotted as a function of the figure rank for those cells (40 cells of monkey A and 7 cells of monkey I), showing shape selectivity for each visual cue. For each cell and visual cue, we normalized the net responses by expressing them as a fraction of the maximal net response for that cue. The figures were ranked according to the cell's response strength for the luminance-defined figure, and this ranking was applied to all three types of figures. For each figure rank and cue, the normalized net responses were then averaged for the 47 cells.

Laboratorium voor Neuro- en Psychofysiologie, Faculteit der Geneeskunde, Katholieke Universiteit Leuven, Campus Gasthuisberg, B-3000 Leuven, Belgium.

*To whom correspondence should be addressed.

from, each having a different preference, can code unambiguously the shape of a figure independent of its retinal size, position, contrast, and defining cue.

Our results provide further evidence for the role of the IT in the representation of shapes, even for shapes defined by relative motion. The latter is an important point because the IT is part of the ventral stream of the visual pathway, whereas motion analysis is traditionally assumed to be part of the dorsal pathway, which consists of visual areas such as the MT and the medial superior temporal area. Our results, however, show that the IT processes shapes even in those cases in which the shape can be computed from motion information only, which indicates a high degree of convergence of information from dorsal and ventral visual areas in the IT area (*16*). Hence, our finding of cue invariance for shape selectivity in IT neurons shows that single IT cells can abstract invariant shape properties from widely varying stimulus conditions.

REFERENCES AND NOTES

1. T. D. Albright, *Science* **255**, 1141 (1992). However, in some instances the direction selectivity can strongly depend on the way the moving figure is defined [J. F. Olaveria, E. A. DeYoe, J. J. Knierim, J. F. Fox, D. C. Van Essen, *J. Neurophysiol.* **68**, 164 (1992)].
2. The IT is the last unimodal visual area of the ventral cortical processing stream [L. G. Ungerleider and M. Mishkin, in *Analysis of Visual Behavior*, D. J. Ingle, M. A. Goodale, R. J. Mansfield, Eds. (MIT Press, Cambridge, MA, 1982), pp. 549–580] and receives its main visual input from extrastriate areas V4 and TEO [R. Desimone, J. Fleming, C. G. Gross, *Brain Res.* **184**, 41 (1980); A. Morel and J. Bullier, *Visual Neurosci.* **4**, 555 (1990); J. S. Baizer, L. G. Ungerleider, R. Desimone, *J. Neurosci.* **11**, 168 (1991)]. The MT, on the other hand, is part of the dorsal cortical processing stream [L. G. Ungerleider and R. Desimone, *J. Comp. Neurol.* **248**, 190 (1986)].
3. A shape can be defined by any one of several static cues such as brightness, color, texture, or depth. Furthermore, recent perceptual studies show that discrimination of a shape defined solely by relative motion, a dynamic cue, can be as acute as discrimination of the same, luminance-defined shape [D. Regan and S. Hamstra, *Perception* **20**, 315 (1991)].
4. P. Dean, *Psychol. Bull.* **83**, 41 (1976).
5. C. G. Gross, C. E. Rocha-Miranda, D. B. Bender, *J. Neurophysiol.* **35**, 96 (1972); R. Desimone, T. D. Albright, C. G. Gross, C. Bruce, *J. Neurosci.* **4**, 2051 (1984).
6. K. Tanaka, H. K. Fukuda, Y. Moriya, *J. Neurophysiol.* **66**, 170 (1991).
7. T. Sato, T. Kawamura, E. Iwai, *Exp. Brain Res.* **38**, 313 (1980); E. L. Schwartz, R. Desimone, T. D. Albright, C. G. Gross, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 5776 (1983); E. T. Rolls and G. C. Baylis, *Exp. Brain Res.* **65**, 38 (1986).
8. The monkeys were trained to fixate a small spot. Eye movements were recorded with the scleral search coil technique [S. J. Judge, B. J. Richmond, F. C. Chu, *Vision Res.* **20**, 535 (1980)]. After 1000 ms of fixation during which a random-dot field was presented, the figure appeared for 500 ms. The figure was presented foveally, except in some controls. If the monkey maintained fixation, the animal was rewarded with apple juice immediately after figure offset. Animal care was in accordance with the guidelines of the Belgian Fund for Scientific Research. Aseptic surgeries were performed under full anesthesia.
9. The eight figures were chosen from the literature (*5*, *7*). The stimuli were created with hardware developed in the Utrecht Biophysics Institute (J. J. Koenderink) and displayed on a BARCO CD233 monitor (frame rate 50 Hz; P22 phosphor) (BARCO Video and Communications, Kortrijk, Belgium). The random-dot patterns had a 50% density and a dot size of 3 arc min. The average luminance and contrast were 26 cd m^{-2} and 95%, respectively. The speed of the moving random-dot fields was 3° per second. The direction of motion (upward and downward) was randomized over trials. All figures covered the same area (a square 3° on a side). The average luminance of the figure was lower (16 cd m^{-2}) than the background for the luminance-defined figures. For the motion-defined and texture-defined figures, the luminance of the dots outside the figure was the same as that of the dots inside the figure so that there was no difference in average luminance between figure and background [measured with a Minolta LS-100 luminance meter (0.01 cd m^{-2} precision)]. On each trial, the random-dot patterns were renewed so that any spurious differences in luminance between figure and background were averaged out. For luminance-defined and texture-defined figures, the random-dot patterns were stationary, and before figure onset the background random-dot field was displayed. For the motion-defined figures, we presented a moving random-dot field having the same velocity as the background before figure onset to eliminate responses to motion.
10. In the relative motion–defined figures, flicker is present at those borders that are not parallel to the direction of motion. This is a result of the appearance and disappearance of the dots at the virtual border of the figure (dynamic occlusion cue). The contribution of this dynamic occlusion cue to the kinetic figures responses was determined in control tests with figure stimuli in which the direction of motion for figure and background was the same, yielding flickering dots at those borders that were not parallel to the direction of motion. The control stimuli differed in only one respect from the kinetic stimuli: the motion direction of the dots inside the figure relative to that of the background dots. These tests also controlled for another dynamic cue present in all stimuli, static as well as kinetic. Indeed, correlated with the appearance of the figure was an inevitable one-frame decorrelation of the dots inside the figure because of the transition from a frame containing only background to a frame containing figure and background. This transient onset cue was also present in the control stimuli: the random-dot pattern within the figure was renewed in the frame marking onset of the figure. Hence, the contribution of all dynamic cues unrelated to the velocity difference between figure and background could be determined directly by a comparison of the responses to the control and relative motion figures.
11. Single cells were recorded with glass-coated tungsten microelectrodes in daily sessions. The figures were presented monocularly in a randomized order, and at least ten correct trials per figure were collected for each cue. Only those cells (175 cells in monkey A and 86 cells in monkey I) that responded to at least one type of figure were fully tested. Spike counts were computed trialwise with a 500-ms bin that started 50 ms after figure onset. Net responses were calculated by trialwise subtraction of the activity during a 500-ms fixation period just preceding figure onset. We used analysis of variance (ANOVA) [R. E. Kirk, *Experimental Design: Procedures for the Behavioral Sciences* (Brooks/Cole, Belmont, CA, 1968)] to test the significance of the responses to any figure and the significance of shape selectivity. Tests were classified as significant when the corresponding type I error was smaller than 0.05. The histology of the brain of monkey A confirmed that the recordings were made in the middle part of the inferior temporal cortex, or area TE3 [B. Seltzer and D. N. Pandya, *Brain Res.* **149**, 1 (1978)]. Monkey I is still engaged in experiments, so its histology is not yet available.
12. Paired t test: $t(25) = 11.6$; $P < 0.001$. Fifteen of the 26 cells were shape-selective for each of the three visual cues. For each of these 15 cells, the response to the relative motion figures was larger than the response to the control figures (the average net response for relative motion figures was 29 spikes per second and for control figures was 11 spikes per second; $t(14) = 4.45$; $P < 0.001$). This shows that the presence of any dynamic cue other than relative motion cannot explain the responses to the relative motion figures. For the same cells, the responses to the luminance-defined (average net response was 35.7 spikes per second) and the texture-defined figures (average net response was 25.0 spikes per second) were also significantly larger than the responses to the control figures, which indicates that the responses for the former stimuli cannot be explained by the transient onset cue (*10*). Hence, we conclude that this population of IT cells responds to shapes defined by texture, relative motion, and luminance.
13. A shift of 4° in eccentricity was sufficient to stimulate nonoverlapping parts of the visual field, given the size and shape of the figures.
14. The only criterion used to select the 47 cells on which this analysis is based was a pure statistical one: for each of the three types of cues, ANOVA tests had to show a statistically significant effect of figure selectivity.
15. For the 47 cells selective for the three defining cues, the response was usually larger for the preferred luminance-defined figure than for the same motion-defined figure (median ratio: 1.4; first quartile: 1.0; third quartile: 1.9) or texture-defined figure (median ratio: 1.5; first quartile: 1.0; third quartile: 2.3). Therefore, cue invariance implies cue-invariant shape selectivity but not cue-invariant response strength, because the response strength of a cell depends on many other parameters such as dot contrast, dot velocity, and dot size. This is similar to the "ambiguity" problem encountered at earlier levels of the visual system—that is, the response strength of V1 cells may depend on the orientation, spatial frequency, contrast, and size of a grating pattern.
16. Our finding that single inferior temporal units are selective for shapes defined by relative motion is in agreement with recent lesion work in monkeys [K. H. Britten, W. T. Newsome, R. C. Saunders, *Exp. Brain Res.* **88**, 292 (1992)]. Lesion studies suggest that the dorsal area MT is necessary for providing the velocity signal of the local motion components [V. L. Marcar and A. Cowey, *Eur. J. Neurosci.* **4**, 1228 (1992)], but MT cells are not sensitive to the orientation of kinetic boundaries [V. L. Marcar *et al.*, *Soc. Neurosci. Abstr.* **17**, 525 (1991)]. However, the velocity information from the MT area may lead to the responses to kinetic boundaries observed in area V2 [V. L. Marcar *et al.*, *ibid.* **18**, 1275 (1992)] and ventral area V4 [N. K. Logothetis and E. R. Charles, *Invest. Opthalmol. Vis. Sci. Suppl.* **31**, 444 (1990)]. The latter units may provide the input to the shape-selective IT cells we recorded from.

17. Supported by European Specific Program for Research in Information Technology grants BR action 3001 (Insight) and IUAP22 (Vision and Memory) (G.A.O.) and grant FGWO 9.0039.90 (R.V.). R.V. is a research associate of the Belgian Fund for Scientific Research. Gy.S. was on leave from the Department of Physiology of the Albert Szent-Györgyi Medical University, Szeged, Hungary. We thank P. Kayenbergh, G. Meulemans, and G. Vanparrijs for technical assistance, A. Coeman and C. Fransen for assisting with parts of the data analysis, W. Spileers for help with the eye coil surgery, and B. De Bruyn, K. Lauwers, S. Raiguel, and A. A. Schoups for comments on the manuscript.

31 August 1992; accepted 4 March 1993