NeuroReport 5, 829–832 (1994)

Neuronal tuning to learned complex forms in vision

Kuniyoshi Sakai[CA] **and**
Yasushi Miyashita

Department of Physiology, School of Medicine,
University of Tokyo, 7-3-1 Hongo, Bunkyo-ku,
Tokyo 113, Japan

[CA] Corresponding Author

WE tested the hypothesis that the fine form selectivity of anterior inferotemporal (AIT) neurones represents long-term memory of object forms. Prior to neuronal recording, two monkeys were trained to learn a set of computer-generated Fourier descriptors (FDs). A slight alteration of one FD parameter produces a form with similar global features. We found that these altered forms always elicited weaker neuronal responses than the learned forms. This effect was not due to nonsensory factors such as attention, because another sham manipulation, in which one FD was rotated or reflected, did not produce a comparable effect. These results demonstrate that AIT neurones acquire a preference for complex forms through learning by tuning mechanisms.

Key words: Visual memory; Associative learning; Single-neurone recording; Anterior inferotemporal cortex; Macaque monkey

Introduction

The anterior inferotemporal (AIT) cortex has been proposed to be the memory storehouse in object vision.[1–3] Along the visual pathway from the primary visual cortex (V1) to the AIT cortex, both the receptive field size and the complexity of neuronal processing increase.[4,5] Consequently, AIT neurones respond selectively to computer-generated complex patterns.[6,7] The conventional method in determining the response selectivity of a single neurone is to test its responses with a screening set of various object forms. It is a possible mistake to conclude that one tested form is the optimum stimulus for that cell, though it is most effective among a limited number of stimuli in the screening set. Another untested form, whose global feature is similar to the tested form, may have elicited much stronger responses. Therefore, it is necessary to analyse fine form selectivity with better resolution than previously tested. This problem originates from the complexity of form features, in contrast to the optimal features in V1 such as orientation. The analysis of fine form selectivity becomes particularly important in searching for memory traces of specific forms, which could be acquired through learning experience. From this viewpoint, we tested the hypothesis that the fine form selectivity of AIT neurones represents long-term memory of object forms.

Materials and Methods

Subjects: Two adult monkeys (*Macaca fuscata*) were trained in the pair-association task.[7,8] After the monkey had learned the task, four fixation bolts and a cylindrical chamber for microelectrode recording were attached to the skull under aseptic conditions and general anaesthesia with Nembutal® (30 mg kg^{-1}). The chamber was filled with sterile saline containing gentamycin. The monkey was given antibiotics and sufficient rest for recovery after surgery. Training for head restraint was achieved gradually, thereby using the monkey's operant behaviour as evidence of the absence of discomfort. Recording sessions usually lasted 3 h, during which time the animal consumed 400–500 ml of fruit juice. The care and use of animals conform with the guidelines of the National Institute of Health (revised 1985) and the Faculty of Medicine, University of Tokyo (1987).

Stimuli: Visual stimuli are computer-generated Fourier descriptors (FDs). FDs are appropriate for the purpose of this experiment, because they are specified by a set of parameters:[9] harmonic amplitude Ak and phase angle αk; $k = 1,2, \ldots$ We prepared a screening set of 24 FD patterns and sorted them randomly into pairs, numbering from 1 and 1′ to 12 and 12′ as in the previous report (Fig. 1*a* in Sakai and Miyashita[7]). The pair combinations in the screening set were fixed throughout the whole experiment. A slight alteration of one FD parameter from its original value produces a very similar pattern. This manipulation of an original pattern is called parametric transformation (Fig. 1A). The similarity of patterns with slightly altered FD parameters is validated by the fact that the position of point $Z(s)$ in the drawing plane, where s is arc length, is a continuous function of FD parameters (some examples are shown in Fig. 2B). We adopted another manipulation, geometric transformation, in which an original pattern was rotated about its centre with various rotation angles or was reflected as a mirror image. The geometric transformations show different views of the same form, whereas the parametric transformations show different forms with similar global features. Both types of transforms were treated as unlearned patterns, and the same procedure was used to incorporate both types of transforms. The pair-association task forces the monkey to discriminate unlearned transforms from the original pattern by imposing a new association (see below). The geometric transformation can serve as a sham manipulation to incorporate transformed patterns.

Task: In each trial of the pair-association task,[7,8] a cue stimulus is presented at the centre of the video monitor for 0.5 s. After a delay period of 5 s, two stimuli of choice, the paired associate (correct choice) and a distractor (incorrect choice), are shown. The monkey obtains a reward for touching the correct one within 1.2 s. This task paradigm can reliably demand the learning of a screening set of visual stimuli, because monkeys cannot select a paired associate correctly without memorizing and recalling pair combinations.

The procedure to incorporate an unlearned pattern in the task is the following. For each recorded neurone, we select an original pattern (for example No. 3′ shown in Fig. 1A, top) that elicits the best or the second best response among the 24 patterns during the cue presentation. Pattern $\hat{3}'$, transformed from 3′, is used to temporarily replace, not the original 3′, but one of the original patterns (for example No. 9′) that is randomly selected for each recording session. In this example, pattern $\hat{3}'$ and remaining 23 original patterns are used as cue and choice stimuli. This procedure has two advantages. First, a neuronal response to the original 3′ can be recorded and compared with a response to $\hat{3}'$ in the same block of trials. Secondly, a new association between two patterns ($\hat{3}'$ and 9 in this case) is imposed along with the original pair combination (3′ and 3) to discriminate $\hat{3}'$ from 3′.

Recording and analysis: Extracellular discharges of a single neurone were recorded by a glass-insulated tungsten microelectrode using standard physiological techniques.[7] The recording sites are within a similar region of the AIT cortex shown in the previous report (Fig. 2*a* in Sakai and Miyashita[7]). Statistical significance was evaluated by means of the one-tailed modified *t*-test.[10] In this paper, the critical comparison is the neuronal response to the original *vs* the transformed patterns during the cue period.

Results

Of 474 isolated anterior inferotemporal neurones in three hemispheres, 89 cells reproducibly showed pattern-selective responses during the cue period. The most effective stimuli for the 89 cells covered all patterns in the screening set. Out of the 89 cells, 40 were tested for transformed patterns in stable recording conditions. The intensity of the strongest neuronal responses to the screening set ranged from 25.0 to 77.5 Hz (44.9 ± 14.2 Hz, mean ± s.d.; $n = 40$). Figures 1B and 1C show data from one neurone with pattern-selective responses. Although both the original pattern (Fig. 1B) and the pattern with a slightly altered phase angle (Fig. 1C) elicited the cue response, the former had a stronger effect than the latter. This difference is statistically significant ($p < 0.05$, one-tailed modified *t*-test).

Since two or more transformed patterns were tested in each of 10 cells, we analysed 54 cases of transformations. Either the geometric transformation or the parametric transformation was analysed in the individual neurones. Figure 2A exemplifies the effect of the geometric transformation on responses of a single neurone. This cell gave an equal response to both rotated and original patterns, whereas it showed a weaker response to the reflected pattern. In most cells, only small response changes were observed for rotation and reflection.

In contrast, the parametric transformation caused a

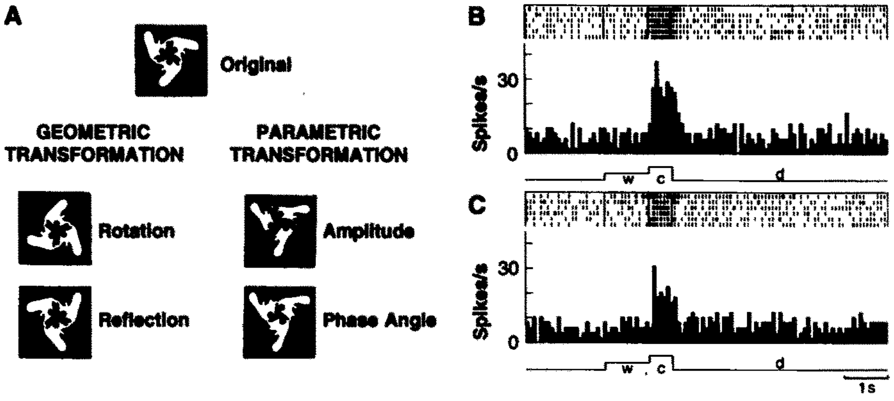

FIG. 1. (A) Two types of transformations of Fourier descriptors. Geometric transformations include rotation and reflection of an original pattern. Examples of rotation and reflection are shown on upper left and lower left, respectively. Parametric transformation is a slight alteration of one FD parameter. Examples of harmonic amplitude and phase angle alterations are shown on upper right and lower right, respectively. The actual images on a video monitor were yellow monochrome against a black field. (B) Neuronal response to the cue stimulus of the original pattern shown at the top in (A). Rastergrams of single-cell discharges in each trial (upper) and spike-density histograms (lower) are shown. Bin width: 80 ms. (C) Neuronal response to the cue stimulus of the pattern shown on the lower right in (A). The neurone illustrated in (B) showed weaker responses to this parametrically transformed pattern than to the original pattern during the cue period in each trial. w, warning period; c, cue period; d, delay period.

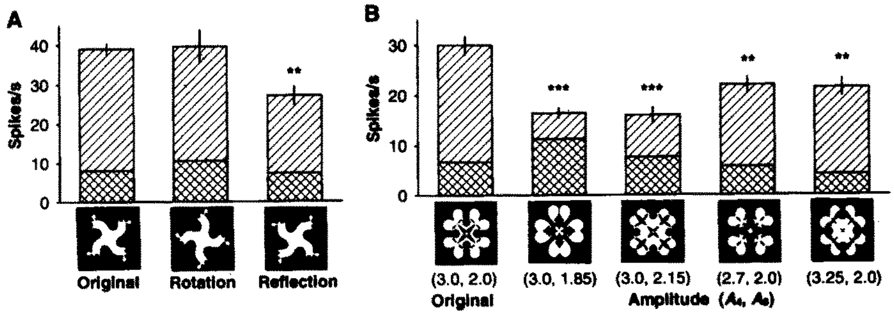

FIG. 2. **(A)** Responses of an exemplary neurone for the geometric transformation. Histograms show mean discharge rates (mean ± s.e.m.) for each cue presentation of the original pattern and patterns transformed from the original. The cross-hatched area in each histogram bar corresponds to the spontaneous discharge rate. **(B)** Responses of an exemplary neurone for the parametric transformation. The original pattern shown is specified by two harmonic amplitudes and two phase angles ($A_4 = 3.0$, $\alpha_4 = \pi/2$, $A_8 = 2.0$, $\alpha_8 = \pi/2$). A set of two numbers attached to each pattern denote its amplitude parameters (A_4, A_8). In each transformed pattern, one of these parameters is slightly altered from the original value. Note the close similarity between these four transformed patterns and the original. This neurone exhibited optimal responses to the original pattern rather than to the transformed patterns. These data were taken from the first five trials in which the original or transformed pattern was used as a cue stimulus. $^{**}p < 0.01$; $^{***}p < 0.001$ (modified *t*-test).

different effect. The original FD pattern shown in Fig. 2B has two amplitude parameters A_4 and A_8 that can take continuous values. We prepared four patterns with a slightly altered value of A_4 or A_8. It can be readily noted that the global features of these altered patterns are strikingly similar (Fig. 2B). All of the transformed patterns elicited significantly weaker responses than the original pattern. This observation was further confirmed by responses of other neurones.

Figure 3 summarizes all 54 cases of transformations. As to the geometric transformation, most of the response changes were within the range of ± 25% and were statistically insignificant ($p > 0.05$, modified *t*-test). In contrast, the parametric transformation resulted in a significantly weak neuronal response for most cases. Moreover, we did not observe any case in which responses to transformed patterns surpassed those to the original patterns. The difference between the two distributions of neuronal responses for the geometric and parametric transformations was statistically significant ($p < 0.01$, Mann–Whitney *U*-test).

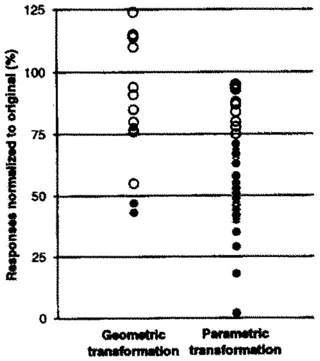

FIG. 3. The different effects of geometric and parametric transformations. The cue response of each neurone to a transformed pattern was compared with that to a corresponding original pattern, which was normalized to 100%. Each circle represents one case. Filled circles denote significantly weak neuronal responses at $p < 0.05$, while open circles denote responses insignificant at $p > 0.05$ (modified *t*-test). A clear distinction was observed between the two distributions of neuronal responses for the geometric transformation ($n = 14$) and the parametric transformation ($n = 40$). These data were based on the comparison between cue responses to an original pattern in three trials and those to a transformed pattern in the first two trials in all 54 cases. In each comparison, mean discharge rates during the cue presentation (from 80 ms to 480 ms after the cue onset) were used. Spontaneous discharge rates were not subtracted.

Discussion

These results suggest that AIT neurones are subject to tuning processes and acquire a preference for learned forms. The property of this tuning mechanism is further clarified by the following explanation: the parameter space of FDs, in each axis corresponds to one FD parameter x_i (e.g. amplitude A_4 in Fig. 2B). Each learned original pattern in the screening set is specified by a set of parameters ($x_1^\circ, x_2^\circ, \ldots, x_n^\circ$), which corresponds to a single point chosen randomly and sparsely in this parameter space. The parametric transformation corresponds to a local perturbation ($x_1^\circ, \ldots, x_{k-1}^\circ, x_k^\circ + \Delta x_k, x_{k+1}^\circ, \ldots, x_n^\circ$) around the point of an original pattern, where $|\Delta x_k|$ is less than half the minimum separation of the sparsely distributed 24 points of original patterns. We assume a continuous mapping from each point in the parameter space to a single-cell discharge rate: $f(x_1, x_2, \ldots, x_n)$. As noted in the Introduction, there is no guarantee that the point of an original pattern, which elicited the strongest response among the screening set, gives a local maximum of $f(x_1, x_2, \ldots, x_n)$ (Fig. 4). Nevertheless, the parametric transformation always caused a decrease in neuronal response, which indicates that any tested points of original patterns give local maxima in responsiveness. Accordingly, neuronal responses should be optimally tuned to the original patterns. This neuronal tuning cannot be explained by learning-independent innate selectivity, since the original patterns in the screening set were selected randomly in the form space. If the extent of the parametric transformation was too large, it would be impossible to study neuronal tuning when there are several local maxima in the form space. Therefore, it is crucial to utilize the local parametric transformation to examine the fine form selectivity.

In primates, form information is subsequently pro-

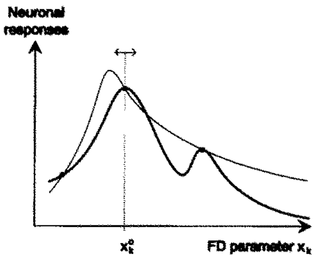

FIG. 4. Neuronal tuning as a local maximum of responses. Schematic single-cell responses $f(x_1^\circ,\ldots,x_{k-1}^\circ,x_k,x_{k+1}^\circ,\ldots,x_n^\circ)$ (see text for definition) to various forms are shown as the function of a FD parameter x_k, where $(x_1^\circ,\ldots,x_k^\circ,\ldots,x_n^\circ)$ is the point of an original pattern that elicited the strongest response among the 24 original FDs in the screening set. These 24 patterns provide only discrete points (three are shown), from which the graph of $f(x_1,x_2,\ldots,x_n)$ is estimated; two exemplars of interpolated graphs are shown as thin and thick curved lines. Here we assume that the middle and the points most to the right give strong responses, whereas the point most to the left meets the spontaneous discharge level. The parametric transformation corresponds to a local perturbation around x_k°, as denoted by ↔ symbol. The local change of $f(x_1,x_2,\ldots,x_n)$ in the region of ↔ symbol may be a monotonic decrease as shown by a thin curved line, where half of the time the parametric transformation would cause an increase in neuronal responses. The experimental results indicate that any tested points of original patterns give local maxima of $f(x_1, x_2,\ldots,x_n)$ as shown by a thick curved line, in which x_k° gives a local maximum. Accordingly, neuronal responses should be optimally tuned to the original patterns.

cessed in a hierarchy of visual areas that include V1, V2, V4, and the AIT cortex.[4] In accordance with the intermediate role of V4, some neurones in V4 show selective responses to concentric gratings, which are elementary components of FD patterns.[11] It has been reported that extraretinal factors such as attention modulate responses of V4 neurones.[12,13] Although AIT neurones receive outputs from V4 neurones with these characteristics, the weaker neuronal response caused by the parametric transformation is not due to extraretinal factors. The geometric transformation did not produce a comparable reduction in neuronal responses notwithstanding the same procedure to impose new association with a transformed pattern. This significant difference should be ascribed to the parametric manipulation of forms. Consistent with the finding that the responses to geometrically transformed patterns were on average equal to the original stimuli, inferotemporal neurones show invariant responses over changes in stimulus size, contrast, orientation and retinal location.[6,14,15]

We previously reported pair-coding neurones and pair-recall neurones that can participate respectively in the coding and recall processes of visual long-term memory.[7] The pair-recall neurone represents a typical example of interaction between sustained activation and association among the neuronal mechanisms that subserve the formation or expression of memory traces.[16] Evidence for picture-selective sustained activation has been documented in the inferotemporal[17,18] and the prefrontal[19] cortices. The fact that each pair-coding neurone responds to both paired associates suggests an important role of interaction between tuning and associative mechanisms. The functional significance of the tuning mechanism is supported by clinical observations that bilateral anterior temporal lesions virtually preclude any visual recognition at the level of unique identity or fine subordinate categorization.[20] One key factor for this across-category deficit is the similarity of physical structures among exemplars. Furthermore, the contribution of the medial temporal lobe to the associative mechanism has been clearly demonstrated in recent lesion studies by Murray *et al.*[8]

References

1. Mishkin M. *Phil Trans R Soc Lond* **B298**, 85–95 (1982).
2. Squire LR and Zola-Morgan S. *Science* **253**, 1380–1386 (1991).
3. Sakai K and Miyashita Y. *Curr Opin Neurobiol* **3**, 166–170 (1993).
4. Van Essen DC, Anderson CH and Felleman DJ. *Science* **255**, 419–423 (1992).
5. Tanaka K. *Science* **262**, 685–688 (1993).
6. Schwartz EL, Desimone R, Albright TD *et al. Proc Natl Acad Sci USA* **80**, 5776–5778 (1983).
7. Sakai K and Miyashita Y. *Nature* **354**, 152–155 (1991).
8. Murray EA, Gaffan D and Mishkin M. *J Neurosci* **13**, 4549–4561 (1993).
9. Zahn CT and Roskies RZ. *IEEE Trans Comput* **c-21**, 269–281 (1972).
10. Snedecor GW and Cochran WG. *Statistical Methods*, 8th edition. Ames, IA: Iowa State University Press, 1989: 97.
11. Gallant JL, Braun J and Van Essen DC. *Science* **259**, 100–103 (1993).
12. Haenny PE, Maunsell JHR and Schiller PH. *Exp Brain Res* **69**, 245–259 (1988).
13. Spitzer H, Desimone R and Moran J. *Science* **240**, 338–340 (1988).
14. Miyashita Y and Chang HS. *Nature* **331**, 68–70 (1988).
15. Gross CG. *Phil Trans R Soc Lond* **B335**, 3–10 (1992).
16. Desimone R. *Science* **258**, 245–246 (1992).
17. Miyashita Y. *Nature* **335**, 817–820 (1988).
18. Fuster JM. *J Neurophysiol* **64**, 681–697 (1990).
19. Wilson FAW, Scalaidhe SPÓ and Goldman-Rakic PS. *Science* **260**, 1955–1958 (1993).
20. Damasio AR, Damasio H and Tranel D. In: Edelman GM, Gall WE and Cowan WM, eds. *Signal and Sense: Local and Global Order in Perceptual Maps*. New York: Wiley, 1990: 451–473.

ACKNOWLEDGEMENT: This work was supported by a Grant for Scientific Research on Priority Areas from the Japanese Ministry of Education, Science and Culture (No. 02102008).

**Received 3 December 1993;
accepted 9 January 1994**