

Differences in Onset Latency of Macaque Inferotemporal Neural Responses to Primate and Non-Primate Faces

Roozbeh Kiani,¹ Hossein Esteky,^{1,2} and Keiji Tanaka³

¹Research Group for Brain and Cognitive Sciences, School of Medicine, Shaheed Beheshti University, Tehran, Iran; ²School of Cognitive Sciences, Institute for Studies in Theoretical Physics and Mathematics, Niavaran, Tehran, Iran; and ³Cognitive Brain Mapping Laboratory, RIKEN Brain Science Institute, Saitama, Japan

Submitted 21 May 2004; accepted in final form 10 April 2005

Kiani, Roozbeh, Hossein Esteky, and Keiji Tanaka. Differences in onset latency of macaque inferotemporal neural responses to primate and non-primate faces. *J Neurophysiol* 94: 1587–1596, 2005; doi:10.1152/jn.00540.2004. Neurons in the visual system respond to different visual stimuli with different onset latencies. However, it has remained unknown which stimulus features, aside from stimulus contrast, determine the onset latencies of responses. To examine the possibility that response onset latencies carry information about complex object images, we recorded single-cell responses in the inferior temporal cortex of alert monkeys, while they viewed >1,000 object stimuli. Many cells responded to human and non-primate animal faces with comparable magnitudes but responded significantly more quickly to human faces than to non-primate animal faces. Differences in onset latency may be used to increase the coding capacity or enhance or suppress information about particular object groups by time-dependent modulation.

INTRODUCTION

Responses of neurons in the visual system vary in their time courses depending on the stimuli, even within single cells (Richmond et al. 1987). One of the most prominent modulations in the time course of responses exists in their onset latencies (Gawne 2000; Gawne et al. 1996; Oram et al. 2002; Reich et al. 2001; Tamura and Tanaka 2001; Tovee et al. 1993). Latencies of responses in cells of the primary visual cortex (V1) and inferior temporal cortex depend on the stimulus contrast, while the orientation of bar stimuli only changes the magnitudes of responses (Gawne 2000; Gawne et al. 1996; Oram et al. 2002; Reich et al. 2001). However, aside from stimulus contrast, it is not known what kinds of features in stimulus images determine the onset latency of responses. Cells in the inferior temporal cortex are selectively activated by complex visual stimuli (Desimone et al. 1984; Gross 1973; Logothetis and Sheinberg 1996; Perrett et al. 1982; Tanaka 1996). Differences in the onset latency of responses may be used to code complex object features in the inferior temporal cortex. Eifuku et al. (2004) found changes in onset latency of responses to faces with changes in the viewing angle, whereas consistent onset latencies were reported for responses to different face views in another study (Oram and Perrett 1992). To examine the possibility that response onset latencies carry information about complex object images, we examined responses of neurons in the inferior temporal cortex to >1,000 object images in alert fixating monkeys.

Address for reprint requests and other correspondence: H. Esteky, Research Group for Brain and Cognitive Sciences, School of Medicine, Shaheed Beheshti University, P.O. Box 19835-181, Tehran, Iran (E-mail: esteky@ipm.ir).

METHODS

Two adult male macaque monkeys (*Macaca mulatta*) weighing 5.8 and 6.5 kg were used. All experimental procedures conformed to the guidelines on the care and use of laboratory animals of the Iranian Society for Physiology and Pharmacology and the National Institutes of Health.

Recordings and stimuli

In an aseptic surgery, a recording chamber was stereotaxically placed on the dorsal surface of the skull, on the left side for one monkey and on the right side for the other monkey. After recovery, extracellular single-cell recordings were made from the anterior inferior temporal cortex with tungsten electrodes (FHC) while the monkey performed a fixation task. The electrode was advanced with an Evart-type manipulator (Narishige, Japan) from the dorsal surface of the brain through a stainless steel guide tube inserted into the brain down to 10–15 mm above the recording sites. Recordings were made on an evenly spaced grid, regardless of the stimulus selectivity of cells, with 1-mm intervals between penetrations over a wide region, which spanned from the medial lip of the anterior middle temporal sulcus to the fundus of the superior temporal sulcus, and from anterior 15–20 in one monkey and anterior 13–20 in the other monkey. The recording positions were determined stereotaxically referring to the magnetic resonance images acquired before the surgery, and the gray and white matter transitions determined during electrode advancement (Tamura and Tanaka 2001). The action potentials from a single neuron were isolated in real time by a template matching algorithm (Worgotter et al. 1986). A conventional amplitude-based window discriminator was used in parallel to avoid contamination of activities from other neurons.

The monkey had to fixate the eyes, with a precision of $\pm 2^\circ$, on a 0.5° circular fixation spot presented at the center of the display. The eye position was measured by an infra-red system (i_rec, <http://staff.aist.go.jp/k.matsuda/eye/>), which allowed a precision of 1° in measurement of eye position. Limiting the analysis to trials in which the eye position stayed within $\pm 1^\circ$ from the center of the fixation spot did not change the results. The presentation of the stimulus sequence started when the monkey had maintained eye fixation for 300 ms. Drops of juice were provided to the monkey every 1.5–2 s during the fixation. The stimulus presentation was terminated after a continuous fixation for 6.7 s or when the monkey broke eye fixation.

Stimuli, which were mainly full-color photographs of objects, were presented in sequence with a 105-ms presentation time per image without an interstimulus interval in pseudo-random order. Neurons in inferior temporal cortex preserve their stimulus selectivity in such rapid serial presentations, even when presentation times are as short as 14–28 ms (Keysers et al. 2001). In some cells ($n = 27$), stimuli were

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

presented in both an ordinary way (245-ms presentation time with a 245-ms blank period between successive stimuli) and by the rapid-serial-presentation method without blank periods. Nine of the cells showed selective responses to faces, and they showed comparable latency differences between responses to human and animal faces with the two presentation methods [10.1 ± 10.0 vs. 7.3 ± 8.5 (SD) ms with and without blank periods, respectively; $P = 0.4$, paired Wilcoxon]. Stimuli were presented on a gray background of 15 cd/m^2 with their centers located at the center of the display.

To determine the luminosity contrast and total luminance of stimuli, we first determined the gamma functions (Brainard et al. 2002) of the three color channels of the monitor (red, green, and blue) by nonlinearly fitting the intensity of light photometrically measured at 16 evenly spaced values covering the whole range of intensity. The validity of the gamma functions and the assumption of channel independence were then confirmed by an extensive photometrical measurement for each of the three channels and their combinations. The luminance of each pixel of a stimulus was then calculated by introducing red, green, and blue values of the pixel into the respective gamma functions and summing the three output values. The average luminance of the stimulus was determined by averaging the pixel luminance over the stimulus region. The luminance contrast was defined by the Michaelson formula of $(L_1 - L_2)/(L_1 + L_2)$. For the figure/ground contrast, L_1 and L_2 were the average luminance of the stimulus and the luminance of background, respectively. For the within-figure contrast, L_1 and L_2 were the maximum and minimum pixel luminance, respectively, within the stimulus. In the stimulus set, human and animal faces had statistically comparable figure/ground and within-figure contrasts ($P > 0.3$). The total luminance (total energy) of each stimulus was also calculated by integrating the pixel luminance over the stimulus. The total luminance of the human faces was significantly larger than that of the animal faces ($P = 0.02$), but this difference in total luminance could not explain a main part of the latency difference (Fig. 4B) as will be described in RESULTS. We didn't examine the power spectrum of spatial frequency components of stimuli because there is no simple way to test statistical differences between two groups of stimuli.

The size of the larger (vertical or horizontal) dimension of each stimulus was equal to 7° . The synchronous signal of the display was recorded, and the onset of stimulus presentation was determined by the time when the electron beam crossed the center of the display. We presented a set of $>1,000$ stimuli for each cell. Each of the stimuli was presented 5–14 times (9.0 ± 1.8 , mean \pm SD). Most of the stimuli were common across cells.

The stimulus set was introduced to the monkeys a week before the recording started. All non-primate animal faces were novel to the monkeys at the time of introduction. All human faces but four were also novel to the monkeys. The four familiar face images were taken from the experimenters and care persons. There were no significant differences in response latency between the familiar and other faces (the mean onset latency averaged across the 136 face cells was 104 ± 23.6 ms for the familiar faces and 103.3 ± 13.0 ms for the other faces; $P = 0.33$, paired Wilcoxon). Images of the rhesus monkey faces were taken from the monkeys being kept in the same room as the experimental monkeys. Neuronal activities were recorded on 120 days in the first monkey and on 71 days in the second monkey. The difference in response latency between human and animal faces did not change during the course of recordings [Spearman correlation coefficient was 0.13 ($P = 0.23$) in the 1st monkey and 0.15 ($P = 0.32$) in the 2nd monkey].

Data analyses

The spontaneous activity was measured in the 200-ms period immediately before the sequential stimulus presentation was initiated. Presentations in which fixation breaks occurred within 300 ms after stimulus onset were excluded. The last two presentations in completed

sequence presentations were also excluded. To exclude the contamination of lasting excitation from the previous presentation, we excluded presentations in which neural activity in the first 50 ms immediately after the stimulus onset was >2 SD above the mean of the spontaneous activity. The SD of spontaneous activity was calculated for trial-by-trial variations of spontaneous firing. This last condition excluded on average 15% of stimulus presentations. It has been shown that the backward masking has minimal effect on the initial part of neuronal responses with stimulus onset intervals >60 or 80 ms (Kovács et al. 1995; Rolls and Tovee 1994).

The significance of responses to individual stimuli was determined by comparing the magnitudes of responses, measured by the averaged firing rate within a 140-ms window starting at 71 ms after the stimulus onset, with those of the spontaneous firing rate ($P < 0.05$, Kolmogorov-Smirnov test). This relatively large window covered the variation in response latency and durations across stimuli and across cells. The latency-adjusted magnitude of responses was calculated by the averaged firing rate within a 100-ms window starting at the onset of response determined as described in the following text. The onset latency of responses was determined for responses of individual cells to individual stimuli. The exact size of the window was not critical for the main results in this paper. Changes in the window size within a range from 60 to 140 ms did not change the main results.

Cells that showed a significant tendency to respond to a larger proportion of face stimuli than non-face stimuli were selected for further analyses. This tendency was quantified by the odds ratio (OR) (Agresti 1996; Somes and O'Brien 1988). The ratio of the number of effective stimulus members that evoked significant responses ($P < 0.05$, Kolmogorov-Smirnov test) to that of ineffective members (odds) was calculated for faces and for non-face objects, and the OR was defined by the ratio of the odds for faces to those for non-face objects. Thus OR is a ratio between ratios. An OR significantly larger than 1 means that the odds of effective stimuli were significantly larger in faces as compared with non-face objects. The significance level was set at $P = 0.01$.

The response onset was determined from a peristimulus time histogram (PSTH), which was constructed for each stimulus with 1-ms bins and smoothed with a Gaussian kernel (SD, 3 ms). The first 1-ms bin of 20 consecutive bins that was 2.58 SD above the mean of the spontaneous activity defined the response onset. The SD of the spontaneous activity, used here, was calculated across bins in a PSTH made for spontaneous activity in the 200-ms period and smoothed by the same Gaussian kernel.

To separate the dependence of response latency on face group (human vs. animal faces) from the dependence on more primitive features of stimulus images, we used a linear model. The model is expressed by a regression equation: $\text{dev}(S) = w_0 + w_1 I(S) + w_2 G(S) + g$, where $\text{dev}(S)$ is the deviation of stimulus response latency from the mean face response latency of the cell, $I(S)$ is the stimulus luminance or contrast depending on the hypothesis, and $G(S)$ is the face group (0 for human faces and 1 for animal faces). w_i are the coefficients, which were determined by the least square fitting. The noise term g is a Gaussian distribution with a mean of 0. To test the significance of dependence on the stimulus group, we evaluated the null hypothesis $H_0: w_2 = 0$ (F test) (Draper and Smith 1998). The value of w_2 in this model also gives us an estimation of latency difference between human and animal faces independent from effects of differences in the contrast or total luminance between human and animal faces.

The reliability of categorical discrimination of stimuli based on the firing of a neuron can be quantified by Receiver Operating Characteristics (ROC) analysis (Hilgers 1991; Metz 1986; Thompson and Zucchini 1989). To examine the time course of the reliability, we applied ROC analysis to activities of individual cells within a 25-ms window and slid the window in 1-ms steps. For the discrimination between human and animal faces, responses to all individual presentations of all human and animal faces were included in the analysis.

Here the horizontal axis of ROC plots the proportion of responses larger than threshold within all the responses to an animal face presentation, and the vertical axis plots the proportion within all the responses to a human face presentation. The threshold was swept over the whole range of response magnitudes to draw an ROC curve, and the area below the curve gave an estimate of the discrimination reliability (ROC value). For discrimination between faces and non-face objects, responses to all individual presentations of all faces and non-face objects were included. The ROC value depends on the size of the time window in which spikes are counted. The window should be big enough to reduce intertrial variance, while not so big so as to smooth away the differences in the time courses of responses to the two groups of stimuli. We selected 25 ms because we found it a good compromise between these two factors in most cells. Values in the paper show mean \pm SD, unless otherwise mentioned.

RESULTS

We analyzed activity of 554 neurons in the inferior temporal cortex with a set of >1,000 stimulus images (1126 ± 71) including 48 ± 2 (median, 49) human faces, 33 ± 13 (median, 41) non-primate animal faces (noted as animal faces hereafter), and more than 900 non-face objects. The human and animal face images commonly used for a majority of the cells are shown in Fig. 1. On average, 16% ($16 \pm 13\%$, median, 13%) of the stimuli evoked significant responses ($P < 0.05$, Kolmogorov-Smirnov test) in each cell. As in previous studies, we found considerable variation of onset latency in responses of each cell. Different stimuli evoked responses with different onset latencies. A systematic latency difference was found between responses to human faces and those to animal faces, and in this paper, we focused on the cells that were selectively responsive to faces. We first selected 189 cells that responded to a significantly larger proportion of faces than non-face

objects [$P < 0.01$, OR analysis (see METHODS)]. Fifty-three cells that showed significant responses to <20% of human faces (13 cells), animal faces (23 cells), or both of them (17 cells) were then excluded to make the determination of mean response latencies for human and animal faces more reliable. For these cells, the small number of effective stimuli in either stimulus group also made it difficult to statistically compare onset latency between the two groups of stimuli within a single cell. The remaining 136 cells showed significant responses to $63 \pm 22\%$ of the human faces, $54 \pm 22\%$ of the animal faces, and $15 \pm 13\%$ of non-face stimuli.

Latency difference between responses to human and animal faces

These 136 cells tended to respond with shorter onset latencies to human faces than to animal faces. Figures 2 and 3 show the results from one example cell and the whole population, respectively. The cell exemplified in Fig. 2 showed significant responses to 94% of human faces and 83% of animal faces. The magnitudes of the responses to human and animal faces largely overlapped with each other (Fig. 2*B*, right; $P = 0.1$, Wilcoxon), whereas the onset latencies of the responses showed a clear difference (Fig. 2*B*, left; $P < 10^{-7}$, Wilcoxon). The difference between the mean latencies was 24.8 ms.

The same was true for the population of cells as a whole. Figure 3*A* plots the mean onset latencies of individual cells for human faces against their mean onset latencies for animal faces. The onset latency was first determined in responses to individual stimuli, and then averaged across all effective stimuli within the stimulus group to obtain the mean. A point represents a cell, and there are 136 points corresponding to the 136 cells. The points are largely distributed below the diagonal



FIG. 1. The 49 human face images and 43 animal face images commonly presented for a majority of cells. The size of the frame is 7° in visual angle. The frame was invisible during the presentation in experiments.

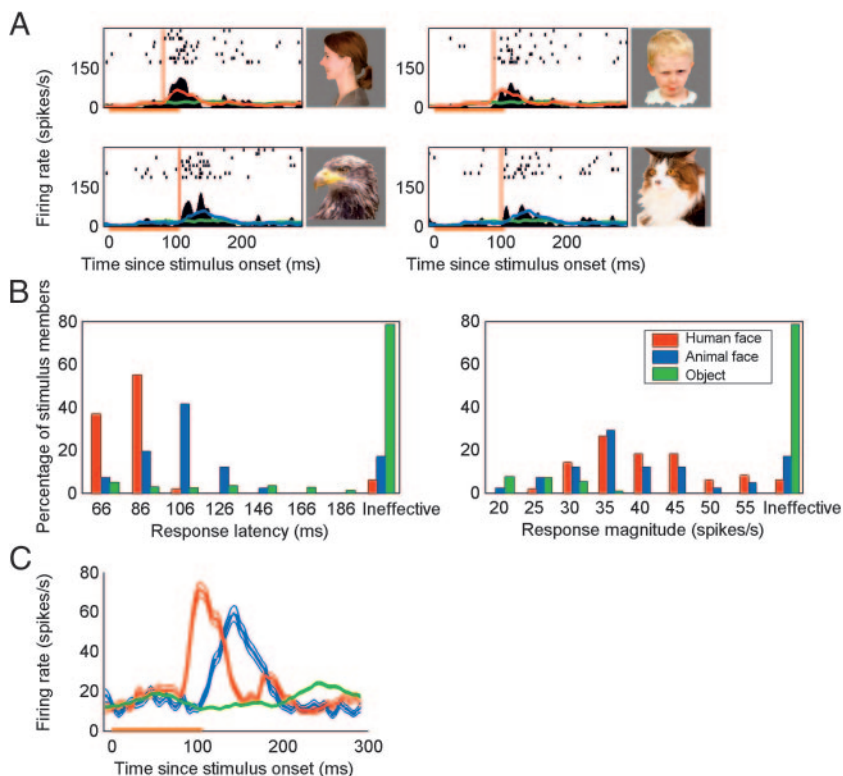


FIG. 2. Different onset latencies of responses to human and animal faces in 1 inferior temporal cell. *A*: peristimulus time histograms (PSTHs) and rastergrams for 2 human and 2 animal faces. Horizontal orange lines show the stimulus presentation period. Vertical red lines indicate the onset latencies determined for the particular response. *B*: distribution of onset latencies and latency-adjusted magnitudes of responses to human faces, animal faces, and non-face objects in this cell. "Ineffective" indicates the stimuli that elicited no significant responses. *C*: averaged responses to human faces (red), animal faces (blue), and non-face objects (green), obtained by averaging responses to individual stimuli across all members belonging to each stimulus group. They are also shown in *A* so that the time courses of individual responses can be compared with those of the averaged responses.

line, and a paired Wilcoxon test showed a highly significant difference ($P < 10^{-18}$) between human and animal means in the cell population. Differences between the two means in individual cells were 14.8 ± 13.3 (SD) ms (median, 14.6 ms),

whereas the mean onset latencies were 103.4 ± 13.1 ms (median, 102.8 ms) for human faces and 118.2 ± 13.2 ms (median, 117.6 ms) for animal faces. A statistically comparable latency difference (10.3 ± 24.6 ms; median, 13.7 ms; $P = 0.3$,

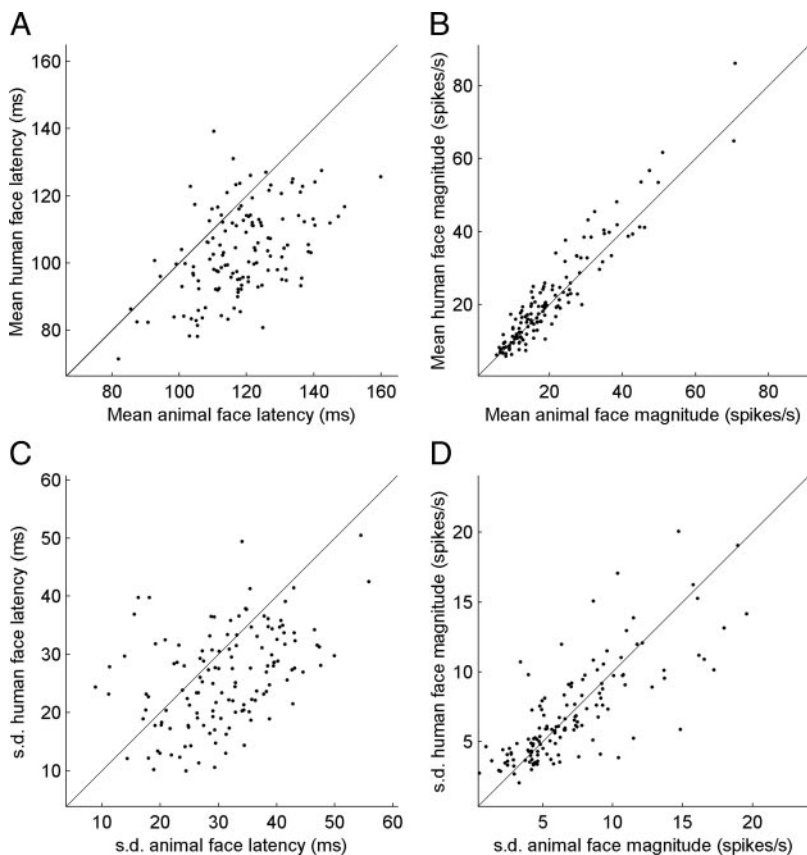


FIG. 3. Scatter plots of mean and standard deviation of onset latency and magnitude for human vs. animal faces. The mean and SD were calculated across human faces or across animal faces for individual cells. Each dot represents 1 of the 136 cells.

Wilcoxon) was observed in the 53 cells that were excluded from the main analyses (see preceding text).

Figure 3B shows a corresponding plot for the means of response magnitudes. The points are distributed along the diagonal line, although the paired Wilcoxon test showed slightly but significantly larger responses to human versus animal faces in the cell population ($P < 0.01$; differences were 1.5 ± 4.5 spikes/s with median of 0.8 spikes/s). The onset latency of individual cells' responses tended to vary more among animal faces than among human faces ($P < 10^{-8}$, paired Wilcoxon; Fig. 3C). There were no such differences between variances among human and animal faces for the magnitudes of individual cells' responses ($P = 0.9$, paired Wilcoxon; Fig. 3D).

A slight negative statistical correlation was previously observed between the magnitudes and onset latencies of responses in inferior temporal cells (Tamura and Tanaka 2001). The differences in latency means and those in magnitude means between responses to human and animal faces were also negatively correlated (Pearson r , -0.2 , $P < 0.05$). To test the possibility that the differences in onset latencies (Fig. 3A) originated in the differences in magnitudes of responses (Fig. 3B), we performed two analyses. First, we excluded 40 cells in which human faces evoked significantly larger responses than animal faces ($P < 0.05$, 1-tailed t -test) and compared the mean onset latencies between responses to human and animal faces for the remaining 96 cells. There was a significant difference in this cell population ($P < 10^{-12}$, paired Wilcoxon; differences were 13.5 ± 12.9 ms with median of 13.3 ms). Moreover, 52 cells that showed numerically larger mean responses to animal faces than to human faces (cells plotted below the diagonal line in Fig. 3B) showed significantly shorter onset latencies to human faces than to animal faces ($P < 10^{-7}$, differences were 11.6 ± 11.1 ms with median of 10.0 ms). The latency differences in the 83 cells that showed statistically comparable magnitudes of responses to human and animal faces were 14.4 ± 13.3 ms, with median of 13.8 ms, which may be a good estimate for the genuine latency difference free from the secondary effects originating in differences in response magnitude.

Second, we used only the responses having overlapping magnitudes between human and animal faces in individual cells to calculate mean latencies, and we compared the mean latencies between human and animal faces in the 136 cells. By removing the nonoverlapping responses, the statistical difference between the magnitudes of mean responses to human faces and those of mean responses to animal faces disappeared

($P > 0.05$, paired Wilcoxon). Nonetheless, means of response onset latencies for human faces were significantly shorter than means for animal faces ($P < 10^{-17}$; differences, 15.1 ± 13.9 ms with median of 15.7 ms). These results showed that the presence of shorter latencies for responses to human faces compared with animal faces was largely independent from the difference in the magnitudes of responses.

One may suspect that the systematic difference in mean onset latencies for human and animal faces was due to greater heterogeneity of animal faces. Faces of some animals evoked early-onset responses, whereas others evoked late responses. The larger variance in onset latencies of individual cells among animal faces compared with that among human faces (Fig. 3C) may support this notion. However, even when we focused on the 53 cells (39% of the 136 cells) that showed no difference in variance of onset latencies between human and animal faces ($P > 0.3$, variance ratio test), the mean onset latencies for human faces were significantly shorter than those for animal faces ($P < 10^{-7}$, paired Wilcoxon, differences were 13.0 ± 12.0 ms). Thus the difference in variance could not be the only cause of the systematic difference in mean onset latencies between human and animal faces.

The luminance contrast of face stimuli could not have contributed to the latency difference, because the figure/ground and within-figure contrasts were comparable across the human and animal face stimuli ($P > 0.3$, t -test). Nevertheless, to further examine the relation between stimulus contrast and response latency, we calculated the mean relative latency, averaged across cells, for each of the face stimuli. For each face stimulus, the onset latency of a significant response was subtracted by the mean latency of the cell to face stimuli and then averaged across cells that showed significant responses to the stimulus. Figure 4A plots this averaged relative latency against the figure/ground contrast of the stimulus. Each dot represents a stimulus. It is obvious that human faces (black circles) were largely separated from animal faces (gray triangles) in the vertical direction at any contrast range. The regression analysis (see METHODS) revealed that the difference in face group (human vs. animal faces) resulted in a 14.4-ms latency difference (significantly larger than 0, $P < 10^{-8}$), whereas there was no significant effect of stimulus contrast ($P = 0.1$). When the within-figure contrast was used as a factor, instead of the figure-ground contrast, very similar results were obtained. The difference in face group manifested a 15.1-ms (significantly larger than 0, $P < 10^{-8}$) difference in mean latency, whereas there was no significant effect of the within-figure contrast ($P = 0.4$).

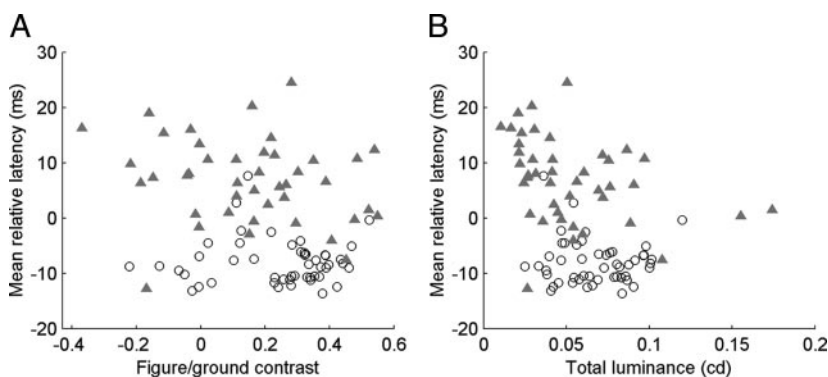


FIG. 4. Effects of figure/ground contrast and total luminance of stimuli on onset latency of cell responses. A black circle represents a human face and a gray triangle an animal face. For each face stimulus, the latency of a significant response was subtracted by the mean onset latency of the cell to face stimuli, and averaged across cells. This averaged relative latency is plotted against the figure/ground contrast (A) or the total luminance (B) of the stimulus. Human faces are largely separated from animal faces in the vertical direction at nearly any contrast or luminance range.

The total luminance (total energy) of the human faces was significantly greater than that of the animal faces ($P = 0.02$). However, as can be seen in Fig. 4B, the averaged relative latencies of responses to human faces were larger than those of responses to animal faces even when the comparison was made between stimuli with comparable total luminance. The regression analysis showed a significant effect of face group, which accounted for a 14.2-ms latency difference (significantly larger than 0, $P < 10^{-8}$). Thus although the total luminance had a marginally significant negative effect on response latency ($P = 0.05$), the majority of the latency difference between responses to human and animal faces was independent from the effect of total luminance.

When the regression analysis was applied to the latency of responses in individual cells, 43% (59 cells) of the 136 face cells showed significant ($P < 0.05$) effects of face group on onset latency, whereas only 7% (9 cells) and 6% (8 cells) showed significant effects of the figure-ground and within-figure contrast, respectively. Only 8% (11 cells) showed significant effects of the total luminance. These latter numbers of cells deviated from the number expected by chance from the P value (6.8) with no (9, $P = 0.15$; 8, $P = 0.24$) or marginal significance (11, $P = 0.05$). These results show that onset latency was much more dependent on the face group than on the contrast or total luminance for single cells. However, we might have observed larger effects of contrast and total luminance if the stimuli covered wider ranges of contrast and total luminance (Oram et al. 2002).

Because there were more profile faces in the animal face group (~26/43) than in the human face group (~6/49), we calculated mean latencies of responses to human profile faces for each cell. Means of onset latencies of responses to human profile faces were significantly ($P < 10^{-8}$, paired Wilcoxon) shorter than the means to animal faces (differences were 10.0 ± 18.4 ms). Thus it is not likely that the differences in onset latencies between human and animal faces were due to the difference in the proportion of profile faces. The size of the stimuli, measured by the largest dimension of the head, was very similar between human and animal faces (139 ± 25 vs. 136 ± 30 pixels, respectively, $P = 0.8$). Moreover, there was no significant correlation between means of onset latencies for individual human faces averaged across cells and the size of the human faces quantified by the distance between the upper lip and the nasal root (the midpoint between the 2 eyes; $P = 0.7$, Spearman) or between averaged latencies for individual animal faces and the size of the animal faces determined by the largest dimension ($P = 0.12$, Spearman).

ROC analyses

Onset latencies of responses to effective human faces were significantly shorter ($P < 0.01$, 1-tailed t -test) than those to effective animal faces in 38% of individual cells (52/136), whereas no cell showed significantly shorter latencies to animal faces. A similar trend was found for most of the remaining cells: 79% (66/84) of cells without significant statistical difference had mean human face latencies shorter than mean animal face latencies. To examine how reliably the instantaneous firing of a single cell could discriminate human faces from animal faces, we applied an ROC analysis (see METHODS)

to the activity of cells within a 25-ms window and moved the window in 1-ms steps. This analysis compared activities in individual presentations of all the stimulus members belonging to the two groups and provides information about how reliably an ideal observer, who is only observing the instantaneous firing rate in one trial but knowing distributions of firing rates in all presentations, can determine whether the presented stimulus belongs to one group or the other. ROC values ~0.5 indicate chance performance, and values of 1 and 0 indicate error-free classification (1 indicates human faces and 0 animal faces here). Most of the cells (110/136 cells) showed peaks with ROC values significantly ($P < 0.05$) larger than 0.5 for human versus animal faces (Fig. 5A). The mean time when the ROC value first exceeded the significance level was 98.7 ± 28.9 ms after stimulus onset. This value corresponds well to the mean onset latencies to human faces, suggesting that the initial spikes to human face presentations were already sufficiently informative for the categorical discrimination of human faces from animal faces. The mean time of the peak in the ROC time courses was 118.4 ± 26.1 ms after stimulus onset. Two-thirds of the cells (84/136 cells) showed troughs significantly smaller than 0.5, and most of them (70 cells) showed both significant peaks and significant troughs. For all 70 cells, the troughs followed the peaks. The mean time of the bottoms in troughs was 168.8 ± 26.6 ms after stimulus onset. The biphasic shapes can also be seen in the averaged ROC time courses averaged over the 136 cells (Fig. 5B, pink). Because the stimulus presentation period was fixed in the present study, we could not determine whether a longer stimulus presentation would result in a change in the time of the trough in human versus animal face ROC time courses. However, responses previously obtained with longer stimulus presentations usually consisted of initial peaks followed by plateaus (though not entirely flat) of reduced magnitudes (Sugase et al. 1999; Tamura and Tanaka 2001). Therefore it is likely that the relative magnitudes of firing rates to human and animal faces reverse, which results in a change of the ROC value from a peak to a trough (crossing the chance ROC value of 0.5), in the initial parts of responses regardless of the stimulus presentation period. The ROC time courses calculated for faces as a whole versus non-face objects (Fig. 5, A, left, and B, green) showed slower time courses with a peak at the time roughly corresponding to the middle between the peak and the bottom of the human versus animal face ROC curve.

The advantage of time-dependent coding in discriminating human versus animal faces was confirmed by applying the ROC analysis to activity in longer time windows. We increased the window width by 1-ms steps with the beginning fixed at the stimulus onset. Figure 6 plots the ROC value against the window width for two cells. The ROC value departs from 0.5 as soon as the window includes the earliest spikes elicited by human faces and reaches a maximum rapidly. Further widening of the window caused a drop to either the chance level (Fig. 6A) or to values smaller than the maximum but still above chance (Fig. 6B). Of 94 cells that showed the maximum ROC value at some window width <200 ms, about half (44 cells) returned to the chance level at 200 ms. For the remaining cells, the difference between the ROC value and 0.5 decreased by $31 \pm 16\%$ (mean \pm SD). As the window became longer to cover whole responses, the difference in distribution of spike

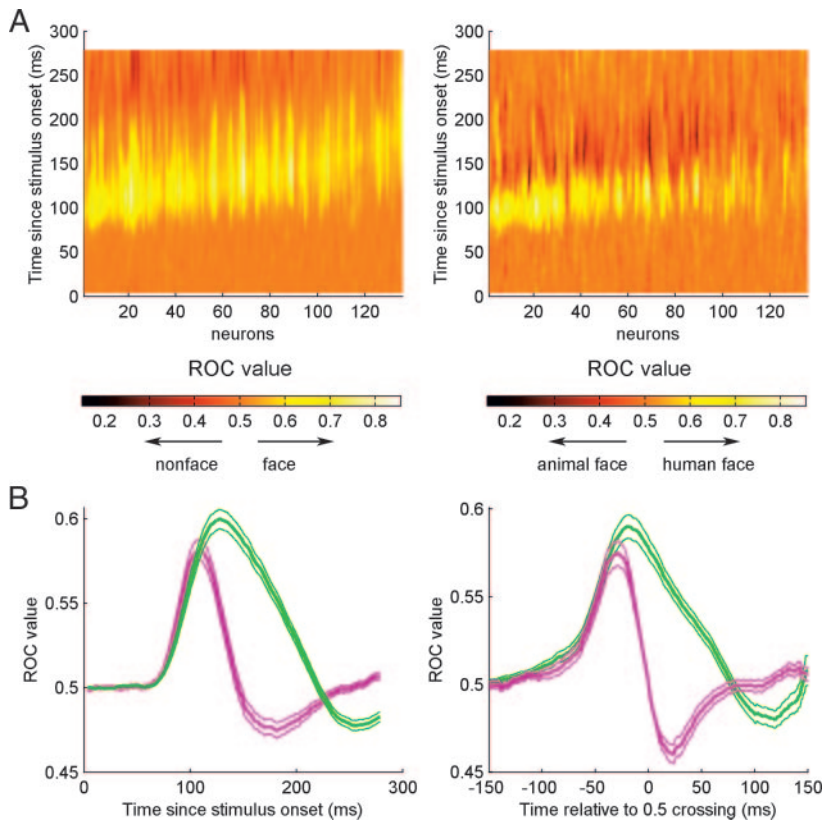


FIG. 5. Time course of categorical discrimination reliability in the 136 face cells. *A*: time courses of the Receiver Operating Characteristics (ROC) value for face vs. non-face discrimination (*left*) and human vs. animal face discrimination (*right*). Vertical axis shows time since the stimulus onset. Horizontal axis shows cells sorted by the latency of the peak in the face vs. non-face ROC time course. *B*: average ROC time courses (mean \pm SE) for face vs. non-face discrimination (green) and human vs. animal face (pink) averaged across the 136 cells. The horizontal axis indicates the time since stimulus onset (*left*) or the time relative to the time when the curve of an individual cell's ROC value crosses 0.5 (*right*).

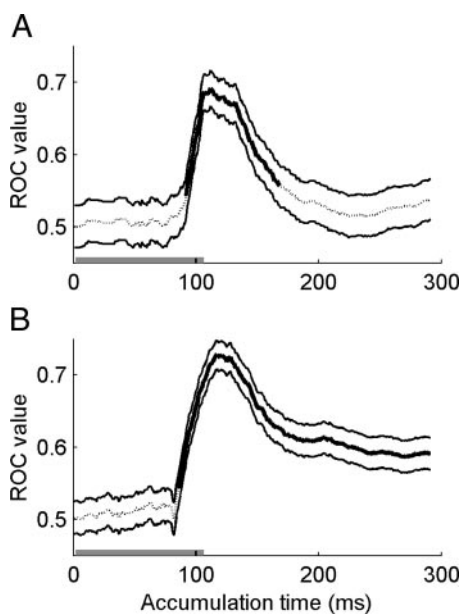


FIG. 6. Variation in discrimination reliability of human and animal faces as a function of accumulation time of neural activity. The window for calculation of ROC values started at image onset and expanded in 1-ms steps. ROC values \pm SE are shown for 2 cells. Solid thick lines depict ROC values significantly different from 0.5 ($P < 0.05$), dotted lines depict values comparable to 0.5 and thin lines show one SE. After a rapid initial increase, ROC values usually return back to chance level (*A*) or drop to a value still above chance but less than the early peak (*B*). Horizontal gray lines show the stimulus presentation period.

counts between human and animal faces decreased, which made the ROC value smaller. This observation shows that an accumulation of spikes over a large window without considering their temporal patterns is a totally inefficient discrimination strategy for nearly half of the cells and a suboptimal strategy for the rest of them.

Latency of responses to monkey faces

In later recordings, we added 16 macaque monkey faces to the image set. In the 63 cells that responded to a significantly larger proportion of faces (including human, monkey, and non-primate animal faces) than non-face objects, means of onset latencies of responses to the monkey faces were significantly shorter than the means for non-primate animal faces ($P = 0.01$, paired Wilcoxon) but not different from the means for human faces ($P > 0.3$).

Latency of responses to faces in non-face cells and that of responses to non-face stimuli in face cells

Distinction of inferior temporal neurons into face cells and non-face cells was a matter of degree rather than categorical distinction. Face cells responded to some non-face stimuli and non-face cells were driven by some faces. Also, the OR for the discrimination of faces versus non-face objects was continuously distributed over cells with a unimodal distribution. The criterion used in this study (OR significantly larger than 1) provides a lenient selection of the cells with relative preference to faces. To examine whether the latency difference between responses to human and animal faces observed in these selected cells was common to occasional responses to human and

animal faces in non-face cells, we accumulated the relative onset latency of significant responses to individual human and animal faces for the two groups of cells. Because non-face cells usually showed significant responses only to a few human faces and a few animal faces, a paired comparison between the two mean latencies of individual cells could not be applied to this group of cells. The relative latency was obtained by subtracting the mean onset latency of the cell averaged over all the effective stimuli (face and non-face objects).

Figure 7 shows the distribution of relative onset latency for the 136 face cells (*A* and *B*) and 365 non-face cells (*C* and *D*). In the group of face cells, distributions of relative latencies for human and animal faces had mean values of -7.5 and $+7.5$ ms, respectively, which are consistent with the main results. Although the mean relative latency in the non-face cell group was still negative (-3.8 ms) for human faces and positive ($+2.8$ ms) for animal faces, the difference between the two means was smaller. In addition to the attenuation of mean latency difference, the variances of relative latency were significantly larger in the non-face cell group than those in the face cell group ($P < 10^{-8}$, variance ratio test). Therefore while the latency difference between face classes was distributed over cells in area TE, face cells were more efficient in discriminating human from animal faces based on onset latency of responses.

To examine whether there was any systematic relation between stimulus properties of non-face stimuli and the onset latency of responses evoked by these stimuli in face cells, we calculated the relative onset latency of responses evoked by non-face stimuli in the face cells whenever the response was significant. The latency was accumulated across cells for each of the 823 non-face stimuli consistently presented to cells. For most (88%) of the non-face stimuli, the distribution of relative latency did not significantly deviate from 0 ($P > 0.05$), which means that these stimuli did not elicit responses with shorter or longer latencies, consistently across cells. The relative laten-

cies were significantly different from 0 for the remaining non-face stimuli (smaller than 0 for 54 stimuli and larger than 0 for 42 stimuli, $P < 0.05$). Although the proportion of these stimuli (12%) was significantly larger than the arbitrary threshold (5%), we could not find any higher-level property (e.g., object category) common to the 54 stimuli or to the 42 stimuli. The two groups of stimuli were also not significantly different in contrast or total luminance.

DISCUSSION

We found that most of the face cells in the inferior temporal cortex of monkeys responded to both human and non-primate animal faces but more quickly to human faces than to animal faces. We added faces of macaque monkeys to the stimulus set in later recordings and found that face cells responded with similar onset latencies to macaque faces and to human faces. Thus in the inferior temporal cortex, faces are discriminated from non-face objects by activation of different groups of cells, whereas primate faces can be discriminated from non-primate animal faces by differences in onset latencies of responses in the same group of cells.

We here refer to only human and macaque faces by “primate faces” because faces of other primates were not included in the present study. A difference of a similar magnitude has previously been found in peak latencies of evoked potentials in human subjects. Human faces evoked potentials 10–14 ms earlier than animal faces (Carmel and Bentin 2002; McCarthy et al. 1999). Slightly smaller but significant difference (6 ms) was also found for faces in scenes (Rousselet et al. 2004).

The latency difference between face groups could not be caused by differences in the exact eye position on the stimuli. Due to the short presentation time (105 ms) and the lack of blank intervals in the present experiments, any saccade triggered by a presented stimulus had to land on the subsequent stimulus. Because the order of stimulus presentation was ran-

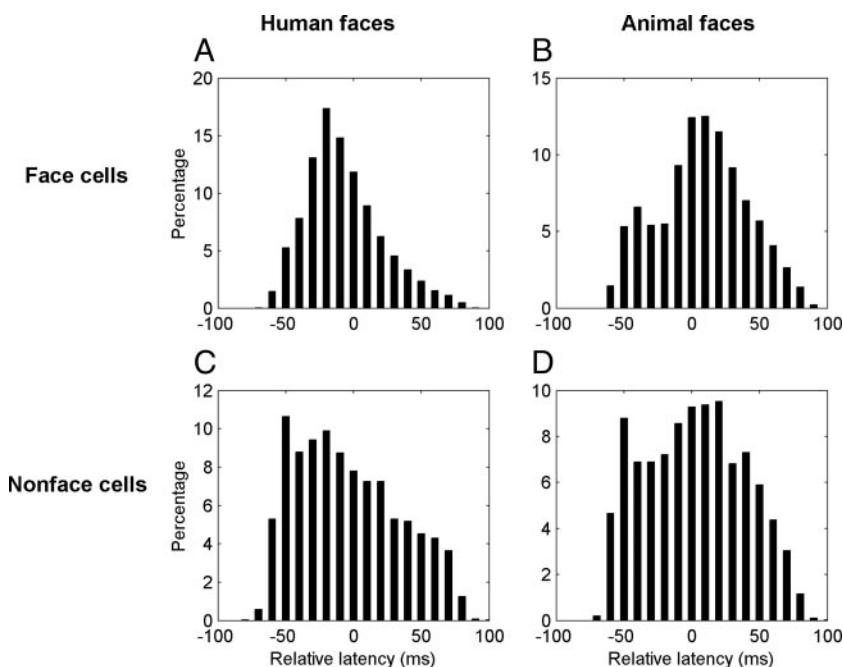


FIG. 7. Distributions of relative onset latency of individual responses to faces in 136 face cells (*A* and *B*) and 365 non-face cells (*C* and *D*). The mean onset latency of each cell was subtracted from the original onset latency to obtain the relative onset latency, which was then pooled across the cells.

domized, such saccades, even if present, could not result in systematic differences in the exact position of eye fixation between stimuli.

Our finding that human faces evoked responses as rapidly as did monkey faces in the inferior temporal cortex of monkeys may seem inconsistent with the known species advantage. Humans can discriminate the identity of human faces better than that of monkey faces, and the reverse is true for monkeys (Pascalis and Bachevalier 1998). We should note here that the neuronal activity was recorded in laboratory monkeys in the present study for which human faces are as salient as monkey faces. It is reasonable to expect that human faces are treated in a similar way as monkey faces in the visual system of laboratory monkeys, for which human faces are as important as monkey faces.

In model neurons, as the stimulus becomes more effective, the magnitude of response increases and its latency decreases, while the change in magnitude saturates sooner than that in latency (reviewed in Koch 1999). Actually, as the contrast of a bar stimulus increases, V1 cells decrease the onset latency of their responses after the magnitude of the response saturates (Gawne et al. 1996; Reich et al. 2001). Such dissociation between changes in magnitude and latency may also occur as stimuli become more effective in other aspects, although it cannot be a general principle because changes occur only in the magnitude and not latency in V1 cells when the orientation of bars changes (Gawne et al. 1996). It is possible that non-primate faces, although weaker stimuli than primate faces for monkey-face cells, may still have been sufficiently effective so as to fall on the saturated portion of the response magnitude curve, and thus produce almost the same firing rates as primate faces. At the same time, those non-primate faces, being weaker stimuli, produced longer latencies.

One possible way the brain might use these latency differences is to enhance or suppress a particular kind of information based on time-dependent modulation. For example, primate faces contain information useful for social interactions, whereas non-primate animal faces may have to be processed in terms of predator versus prey relationships rather than in the social domain. Given the latency differences, each of these different kinds of information may be selectively enhanced or suppressed by time-dependent modulation. Here we need to assume that there are modulatory, or gating, inputs that fire during particular time windows and that different modulatory inputs are activated in different behavioral contexts. Different kinds of information coded in different time windows may also be sorted to different downstream sites by using similar mechanisms. A time-dependent gating has been shown in the signal flow from area MT to preculomotor circuitry, although the size of time gate was larger than that considered here (Seidemann et al. 1998). Latency differences may also be used to increase the capacity of coding as has been repeatedly discussed for differences in the time pattern of responses in general and for latency differences in particular (Eckhorn and Popel 1974; Optican and Richmond 1987; Rieke et al. 1997; Thorpe 1990). Both the enhancing/suppressing and coding require knowing the time that has passed since the stimulus onset. Neural oscillations phase-locked to stimulus onsets, discharges of stimulus-nonselective cells, and neuronal discharges related with previous saccades may be used as refer-

ence points in time (Reich et al. 2001; VanRullen and Thorpe 2002).

ACKNOWLEDGMENTS

We thank M. Noorbakhsh for technical assistance and S. Lehky for a critical reading of the manuscript.

Present address of R. Kiani: Program in Neurobiology and Behavior, University of Washington, Seattle, WA 98195-7270.

GRANTS

The work was supported by the RIKEN BSI Collaborative Research Fund and by IPM School of Cognitive Sciences.

REFERENCES

- Agresti A.** *An Introduction to Categorical Data Analysis*. New York: Wiley, 1996.
- Brainard DH, Pelli DG, and Robson T.** Display characterization. In: *Encyclopedia of Imaging Science and Technology*, edited by Hornak J. New York: Wiley, 2002, p. 172–188.
- Carmel D and Bentin S.** Domain specificity versus expertise: factors influencing distinct processing of faces. *Cognition* 83: 1–29, 2002.
- 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.
- Drapier NR and Smith H.** *Applied Regression Analysis*. New York: Wiley, 1998.
- Eckhorn R and Popel B.** Rigorous and extended application of information theory to the afferent visual system of the cat. I. Basic concepts. *Kybernetik* 16: 191–200, 1974.
- Eifuku S, De Souza WC, Tamura R, Nishijo H, and Ono T.** Neuronal correlates of face identification in the monkey anterior temporal cortical areas. *J Neurophysiol* 91: 358–371, 2004.
- Gawne TJ.** The simultaneous coding of orientation and contrast in the responses of V1 complex cells. *Exp Brain Res* 133: 293–302, 2000.
- Gawne TJ, Kjaer TW, and Richmond BJ.** Latency: another potential code for feature binding in striate cortex. *J Neurophysiol* 76: 1356–1360, 1996.
- Gross CG.** Visual functions of inferotemporal cortex. In: *Handbook of Sensory Physiology*, edited by Autrum H, Jung R, Lowenstein W, Mckay D, and Teuber HL. Berlin: Springer, 1973, vol VII/3B, p. 451–482.
- Hilgers RA.** Distribution-free confidence bounds for ROC curves. *Methods Inf Med* 30: 96–101, 1991.
- Keysers C, Xiao DK, Foldiak P, and Perrett DI.** The speed of sight. *J Cogn Neurosci* 13: 90–101, 2001.
- Koch C.** *Biophysics of Computation: Information Processing in Single Neurons*. New York: Oxford, 1999.
- Kovács G, Vogels R, and Orban GA.** Cortical correlate of pattern backward masking. *Proc Natl Acad Sci USA* 92: 5587–5591, 1995.
- Logothetis NK and Sheinberg DL.** Visual object recognition. *Annu Rev Neurosci* 19: 577–621, 1996.
- McCarthy G, Puce A, Belger A, and Allison T.** Electrophysiological studies of human face perception. II. Response properties of face-specific potentials generated in occipitotemporal cortex. *Cereb Cortex* 9: 431–444, 1999.
- Metz CE.** Statistical analysis of ROC data in evaluating the diagnostic performance. In: *Multiple Regression Analysis: Applications in The Health Sciences*, edited by Herbert DE and Myers RH. New York: Am. Inst. of Physics, 1986, p. 365–384.
- Optican LM and Richmond BJ.** Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. III. Information theoretic analysis. *J Neurophysiol* 57: 162–178, 1987.
- Oram MW and Perrett DI.** Time course of neural responses discriminating different view of the face and head. *J Neurophysiol* 68: 70–84, 1992.
- Oram MW, Xiao D, Dritchell B, and Payne KR.** The temporal resolution of neural codes: does response latency have a unique role? *Philos Trans R Soc Lond B Biol Sci* 357: 987–1001, 2002.
- Pascalis O and Bachevalier J.** Face recognition in primates: a cross-species study. *Behav Process* 43: 87–96, 1998.
- Perrett DT, Rolls ET, and Caan W.** Visual neurons responsive to faces in the monkey temporal cortex. *Exp Brain Res* 47: 329–342, 1982.
- Reich DS, Mechler F, and Victor JD.** Temporal coding of contrast in primary visual cortex: when, what, and why. *J Neurophysiol* 85: 1039–1050, 2001.

- Richmond BJ, Optican LM, Podell M, and Spitzer H.** Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. I. Response characteristics. *J Neurophysiol* 57: 132–146, 1987.
- Rieke F, Warland D, de Ruyter van Steveninck RR, and Bialek W.** *Spikes: Exploring the Neural Code*. Cambridge, MA: MIT Press, 1997.
- Rolls ET and Tovee MJ.** Processing speed in the cerebral cortex and the neurophysiology of visual masking. *Proc R Soc Lond B Biol Sci* 257: 9–15, 1994.
- Rousselet G, Macé MJ, and Fabre-Thorpe M.** Animal and human faces in natural scenes: how specific to human faces is the N170 ERP component? *J Vis* 4: 13–21, 2004.
- Seidemann E, Zohary E, and Newsome WT.** Temporal gating of neural signals during performance of a visual discrimination task. *Nature* 394: 72–75, 1998.
- Somes GW and O'Brien KF.** Odds ratio estimators. In: *Encyclopedia of Statistical Sciences*, edited by Kotz L and Johnson NL. New York: Wiley, 1988, vol. 6, p. 407–410.
- Sugase Y, Yamane S, Ueno S, and Kawano K.** Global and fine information coded by single neurons in the temporal visual cortex. *Nature* 400: 869–873, 1999.
- Tamura H and Tanaka K.** Visual response properties of cells in the ventral and dorsal parts of the macaque inferotemporal cortex. *Cereb Cortex* 11: 384–399, 2001.
- Tanaka K.** Inferotemporal cortex and object vision. *Annu Rev Neurosci* 19: 109–139, 1996.
- Thompson ML and Zucchini W.** On the statistical analysis of ROC curves. *Stat Med* 8: 1277–1290, 1989.
- Thorpe SJ.** Spike arrival times: a highly efficient coding scheme for neural networks. In: *Parallel Processing in Neural Systems*, edited by Eckmiller R, Hartman G, and Hauske G. Amsterdam: Elsevier, 1990, pp. 91–94.
- Tovee MJ, Rolls ET, Treves A, and Bellis RP.** Information encoding and the responses of single neurons in the primate temporal visual cortex. *J Neurophysiol* 70: 640–654, 1993.
- VanRullen R and Thorpe SJ.** Surfing a spike wave down the ventral stream. *Vision Res* 42: 2593–2615, 2002.
- Worgotter F, Daunicht WJ, and Eckmiller R.** An on-line spike form discriminator for extracellular recordings based on an analog correlation technique. *J Neurosci Methods* 17: 141–151, 1986.