Properties of Shape Tuning of Macaque Inferior Temporal Neurons Examined Using Rapid Serial Visual Presentation

Wouter De Baene, Elsie Premereur, and Rufin Vogels
Laboratorium voor Neuro- en Psychofysiologie, K.U. Leuven Medical School, Leuven, Belgium

Submitted 18 July 2006; accepted in final form 20 January 2007

De Baene W, Premereur E, Vogels R. Properties of shape tuning of macaque inferior temporal neurons examined using rapid serial visual presentation. J Neurophysiol 97: 2900–2916, 2007. First published January 24, 2007; doi:10.1152/jn.00741.2006. We used rapid serial visual presentation (RSVP) to examine the tuning of macaque inferior temporal cortical (IT) neurons to five sets of 25 shapes each that varied systematically along predefined shape dimensions. A comparison of the RSVP technique using 100-ms presentations with that using a longer duration showed that shape preference can be determined with RSVP. Using relatively complex shapes that vary along relatively simple shape dimensions, we found that the large majority of neurons preferred extremes of the shape configuration, extending the results of a previous study using simpler shapes and a standard testing paradigm. A population analysis of the neuronal responses demonstrated that, in general, IT neurons can represent the similarities among the shapes at an ordinal level, extending a previous study that used a smaller number of shapes and a categorization task. However, the same analysis showed that IT neurons do not faithfully represent the physical similarities among the shapes. The responses to the two-part shapes could be predicted, virtually perfectly, from the average of the responses to the respective two parts presented in isolation. We also showed that IT neurons adapt to the stimulus distribution statistics. The neural shape discrimination improved when a shape set with a narrower stimulus range was presented, suggesting that the tuning of IT neurons is not static but adapts to the stimulus distribution statistics, at least when stimulated at a high rate with a restricted set of stimuli.

INTRODUCTION

Visual object recognition and categorization are extremely difficult for artificial vision systems, although they seem to be accomplished effortlessly by the brain. In macaques, these processes are thought to depend on the highest stage of the ventral stream: the inferior temporal (IT) cortex (Dean 1976; Logothetis and Sheinberg 1996). Single IT neurons can be strongly selective for object attributes such as shape, texture, and color, while remaining tolerant to some transformations such as object position and scale (for a review, see Logothetis and Sheinberg 1996; Riesenhuber and Poggio 2002; Tanaka 1996).

In contrast with its extensive use in behavioral research (e.g., Chun and Potter 1995; Potter and Levi 1969; Subramaniam et al. 2000), the rapid serial visual presentation (RSVP) paradigm has rarely been used in combination with single-cell recordings in the higher visual cortex. In RSVP, images are presented sequentially and continuously (with no interstimulus interval or ISI) with each image replacing the previous one at the same location on the screen. Keysers et al. (2001) and Földiáik et al. (2004) pioneered the use of the RSVP paradigm to examine the selectivity for complex images of neurons in the superior temporal sulcus (STS). Although an increasing presentation rate resulted in a flattening of the neuronal tuning, the stimulus-coding ability of the population of STS neurons recorded was preserved even at the highest presentation rates (14 ms/image), suggesting that RSVP is a useful technique for studying the stimulus selectivity of STS neurons with a large number of stimuli. However, in that and another study (Kiani et al. 2005) using RSVP, stimuli were highly complex and differenced sharply.

In studying the stimulus selectivity of IT cells, several researchers (e.g., Brincat and Connor 2004; Kayaert et al. 2005a; Op de Beeck et al. 2001; Sigala and Logothetis 2002) opted for the use of parametric shape configurations principally because it allows examining how the responses of IT neurons to complex stimuli are related to the parametric variation built into the stimulus sets. The use of all shapes to search for and test responsive neurons is a prerequisite for obtaining an unbiased measure of the responses and tuning of an IT neuron population to a set of parameterized shapes. However, the total experimental time available is limited when using the conventional presentation techniques, thus limiting the number of stimuli that could be presented in most of these studies. This drawback to use parametric shape sets can be overcome by the application of the RSVP paradigm because it allows the presentation of many stimuli. Thus one aim of the present study was to determine whether the RSVP technique is useful for studying the shape selectivity of IT neurons using parameterized sets of shapes (Kayaert et al. 2005a; Op de Beeck et al. 2001). The validity of the RSVP technique to study shape selectivity for parametric sets is not obvious since the differences among the shapes in such sets are much smaller than the stimulus differences employed in the previous IT studies using RSVP (Földiáik et al. 2004; Keysers et al. 2001; Kiani et al. 2005).

Kayaert et al. (2005a) recorded from IT cells while showing simple shapes (e.g., a rectangle or triangle) parametrically manipulated along simple shape dimensions (e.g., taper or axis curvature). They found systematic response modulation along these simple dimensions with the largest response, on average, to the extreme values along a given dimension. The findings of Kayaert et al. (2005a), which suggest monotonic tuning in IT cortex for simple shapes, raise questions concerning the tuning

Address for reprint requests and other correspondence: R. Vogels, Laboratorium voor Neuro- en Psychofysiologie, K.U. Leuven Medical School, Campus Gasthuisberg, Herestraat 49, bus 1021, Leuven, B-3000, Belgium (E-mail: Rufin.Vogels@med.kuleuven.be).

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.
curves for other, more complex shapes. Op de Beeck et al. (2001) used more complex shapes, but their squared configurations, consisting of merely eight shapes, did not allow disentangling preferences for extreme values on a given dimension from preferences for intermediate values along that dimension. Because of the nature of their configuration, every shape corresponded a priori to an extreme value on a dimension. Hence in a first experiment, our primary aim was to examine the neural representation in IT cortex of complex shapes that vary systematically along shape dimensions using the RSVP paradigm. The application of the RSVP method provided an opportunity to increase the number of shapes per configuration. Thanks to this, we could also examine the extent to which the similarity among complex shapes is also represented on an ordinal and metrical level in IT cells when configurations are used that consist of a larger number of shapes than those employed by Op de Beeck et al. (2001).

As in the Kayaert et al. (2005a) study, the major portion of the neurons recorded in the present study showed a preference for the extreme values of the parametric space, at least within the range of values tested. One important issue, discussed by Kayaert et al. (2005a), is the degree to which the tuning for the extremities of the space is due to the repeated presentation of a restricted set of stimuli, i.e., the statistics of the stimulation history. Recent work has demonstrated that early visual and auditory neurons adapt to recent stimulus statistics so that information transmission is enhanced (e.g., Brenner et al. 2000; Dean et al. 2005; Fairhall et al. 2001; Sharpee et al. 2006; Smirnakis et al. 1997). Similar adaptive mechanisms might be operating at higher levels of the visual system or the effects of such adaptive mechanisms at earlier levels of the visual system could be inherited in IT. It is known that the responses of IT neurons can depend on previous stimulus presentations, e.g., stimulus repetition commonly reduces the responses of IT neurons (Baylis and Rolls 1987; Gross et al. 1967, 1969; Miller et al. 1991a; Riches et al. 1991; Sobotka and Ringo 1993), even with intervening presentations of other stimuli (e.g., Brown et al. 1987; Miller et al. 1991b; Sawamura et al. 2006). Because a high number of stimuli are presented repeatedly in RSVP, this paradigm might be more sensitive to adaptive effects than classical testing paradigms in which one stimulus is presented per trial after acquisition of fixation and the intertrial interval is relatively long. This also implies that RSVP might be a useful technique with which to demonstrate the effect of stimulus statistics on neuronal tuning. Given recent demonstrations of an adaptive rescaling of neuronal responses in the fly visual system (Brenner et al. 2000) and guinea pig inferior colliculus (Dean et al. 2005) when the variance of the stimulus distribution is altered, we used RSVP to examine whether the tuning of IT neurons adapts to the properties of the stimulus distribution other than the mean. Note that the manipulation of stimulus distribution properties is possible only when parameterized sets of stimuli are used. Thus in a second experiment, we measured the responses of IT neurons to a set of shapes that varied along a single dimension. The neurons were tested in two successive blocks using stimuli that differed in stimulus variance and density between blocks. Parts of the results of the present study have been published in abstract form (De Baene and Vogels 2005).

METHODS

Subjects

Two male rhesus monkeys (Macaca mulatta; monkeys J and B) served as subjects. Before conducting the experiments, aseptic surgery under isoflurane anesthesia was performed to attach a fixation post to the skull and to stereotactically implant a plastic recording chamber. The recording chambers were positioned dorsal to IT, allowing a vertical approach, as described by Janssen et al. (2000). During the course of the recordings, we took a structural magnetic resonance (MRI) image, with a copper sulfate filled tube inserted in the grid at one of the recording positions. The recording positions were estimated by comparing this MRI with depth readings of the white and gray matter transitions and of the skull base during the recordings.

All animal care and experimental and surgical procedures followed national and European guidelines and were approved by the K.U. Leuven Ethical Committee for animal experiments.

Stimuli

All shapes (maximum size: 7.1°) were filled with pixel noise and were presented foveally in continuous rapid random sequences at a rate of 100 ms/image on a gray background on a monitor positioned 60 cm from the monkeys (60-Hz frame rate; 1,024 × 768 pixels). A trial started after 250 ms of stable fixation and ended when the monkey broke fixation or when every stimulus had been presented twice (for experiment 1) or 20 times (for experiment 2). Different visual stimuli were used in each experiment (see following text).

EXPERIMENT 1. We generated five parametric sets in which shapes were varied systematically along two dimensions, permuting into 25 combinations of values of the two dimensions per set in a circular configuration (see Fig. 1). The range of the dimensions of the parametric configurations was restricted in two ways. First, within a set, the pixel-based dissimilarities (cf. see following text, Eq. 3) between the shapes along a given dimension had to be largely similar to the pixel-based dissimilarities between the shapes along the other dimension. Second, these pixel-based dissimilarities, as calculated per set, had to be similar across sets. For sets 1 and 2, the dimensions taper and aspect ratio were manipulated. The amplitude of two radial frequency components were varied in sets 3 and 4, whereas the taper of the bottom and the top part of a two-part shape were independently manipulated in set 5. The isolated upper parts of the stimuli on the vertical axis of set 5 as well as the isolated lower parts of the stimuli on the horizontal axis of set 5 constituted a sixth set (Fig. 1). The shapes of set 6 were presented at the same positions as when presented in combination (set 5). The resulting 135 stimuli were presented in a random order.

EXPERIMENT 2. In experiment 2, the parametric dimension showing the maximal response modulation in experiment 1 was chosen for each of the five circular parametric sets of experiment 1 (i.e., without set 6). Based on these five dimensions, we created two one-dimensional configurations of nine shapes each per set (Fig. 2) and labeled these the “narrow-range” and the “wide-range” configurations. The parametric distances between the subsequent stimuli of the wide-range configurations were chosen such that no additional features were introduced for the shapes at the extremes. The narrow-range configurations were obtained by halving parametric distances between subsequent stimuli in the wide ranges. This resulted in a fourfold increase in variance and in an increase in the mean (pixel-based) dissimilarity (cf. following text, Eq. 3) between stimuli in the wide-range sets compared with the narrow-range sets, as well as in a lower density of the wide-range stimulus configurations.

Procedure

In both experiments, eye position was monitored through the pupil position using an infrared eye tracking system (ISCAN, EC-240A) at
a sampling rate of 120 Hz. Monkeys were rewarded with a drop of fruit juice at an increasing pace as long as they kept their gaze within 3° (monkey J) or 1.5° (monkey B) of a black fixation target (0.17° diam) in the center of the display.

EXPERIMENT 1. We searched for responsive neurons by presenting the 135 stimuli in a random order at a continuous rate (no ISI) of 3.3 images/s (27 neurons) or 10 images/s (57 neurons) in trials of maximally 270 stimuli while the monkey was passively fixating. We visualized the responses of the cell in a peristimulus time histogram (PSTH) averaged over trials and all stimuli. If this PSTH showed that the cell responded, we continued presenting the stimuli for ~15 min. If the PSTH indicated that the cell did not respond to any of the stimuli, we abandoned this cell and searched for another. All 27 neurons that were initially tested with the slow presentation rate of 3.3 images/s (stimulus presentation duration = 300 ms) were subsequently tested with the short, standard 100-ms presentation times, allowing a comparison of the responses for the two presentation rates.

EXPERIMENT 2. Search test. We searched for responsive neurons with the nine stimuli of every shape set presented for 300 ms, randomly intermixed, with an ISI of 700 ms. For half of the recorded cells, the wide-range configurations were used in the search test; for the other half, the narrow-range configurations were used. When we found a neuron responsive to ≥1 of these 45 stimuli (based on the visual inspection of the PSTHs), the shape set with the largest neuronal responses was selected for the subsequent test.

RSVP test. The stimuli from the selected shape set with the same range configuration as that in the search task were presented randomly intermixed at a rate of 10 images/s (with no ISI) in trials of no more than 180 stimuli for a total duration of ~15 min. Afterward, the stimuli of the other range configuration of the same shape set were presented in a similar fashion for ~30 min. The second range was presented twice as long as the first range. This enables us to examine the temporal evolution of the adaptation to the stimulus statistics even when this adaptation process was slow. The order of the wide- and narrow-range configurations was counterbalanced across neurons.

Recordings

Standard extracellular recordings were performed with Tungsten microelectrodes, lowered in a guiding tube, into the lower bank of the superior temporal sulcus and lateral convexity of IT during a passive fixation task. The signals of the electrode were amplified and filtered using conventional single-cell recording equipment. Spikes from individual neurons were isolated on-line using Plexon software (Plexon, Dallas, TX). The timing of the single units and the stimuli and behavioral events were stored with 1-ms resolution on a personal computer for later off-line analysis.

Data analysis and tests

In both experiments 1 and 2, the response of the neuron was defined as the mean number of spikes in a 50- to 200-ms analysis window relative to stimulus onset. Alternative analysis windows (50–250 ms and 100–200 ms) showed highly similar results. The first three stimuli of every trial were excluded from all analyses because the responses of the majority of the recorded neurons were characterized by a burst at the start of every trial, lasting for ~300 ms (i.e., the total presentation duration of 3 stimuli at a rate of 100 ms/image). The last stimulus of every trial was also excluded from all analyses. This last stimulus differed from all others in the trial in that it was not followed by any other stimulus, excluding any potential backward masking effect that could have been present for all other stimuli.

EXPERIMENT 1. All analyses were performed on those neurons showing shape selectivity within one or more shape sets. The shape selectivity of the neurons was examined by assessing the statistical significance of the observed variance of neuronal mean responses to stimuli within a shape set by using a permutation test. The range of variances per shape set expected by chance was determined by calculating new variances from the data after permuting the order of the stimuli within each trial while maintaining the actual spike counts. A distribution of 1,000 permuted variances was generated, representing the distribution of variances that would have been expected to occur by a chance association between stimulus and neuronal firing. If the observed variance of the neuronal mean responses to stimuli within a shape set was larger than the 95th percentile of the values in its own variance distribution, that neuron was considered to be shape selective within that shape set (P < 0.05, 1-tailed). The shape selectivity for the 10 shapes of set 6 (the isolated parts of the shapes on the axes of set 5) was also tested with this permutation test.

As a measure of the selectivity for the shapes of a set, we computed the depth of selectivity (DOS) (Rainer and Miller 2000) for every
neuron and shape set. This measure of the degree of tuning of a neuron to a given stimulus set is defined as

\[
\left[ n - \left( \sum R_i / R_{\text{max}} \right) \right] / (n - 1) \tag{1}
\]

where \( n \) = number of shapes of a set, \( R_i \) = mean firing rate to \( i \)th shape and \( R_{\text{max}} = \max \{ R_i \} \). The DOS can vary between 0 (when the neuron responds equally to all shapes) and 1 (when there is a response to only one shape).

To estimate the reliability of our procedures in measuring the degree of selectivity, we employed an odd-even split half method. First we computed for each neuron the DOS indices separately for the odd and even repetitions of the stimuli and correlated these DOS indices across neurons. Because the split-half correlation is based on only half of the data, this was corrected by computing the Spearman-Brown split-half coefficient (Lord and Novick 1968)

\[
r_{\text{sh}} = 2r_{xy} / (1 + r_{xy}) \tag{2}
\]

where \( r_{\text{sh}} \) = corrected split-half reliability coefficient and \( r_{xy} = \) correlation between the DOS-indices for the odd and even repetitions.

To investigate whether the population responses of IT neurons can reveal low-dimensional representations of similarity in the parametrically configured shapes, we compared the within-set configurations obtained from position corrected pixel-based similarities with the neuronal representation space in IT cortex. Pixel-based similarities between two shapes were computed for each of 99 × 99 relative positions. The positions of one shape corresponded to a 99 × 99-square grid (step size = 1 pixel) that was centered on the other shape. As the pixel-based dissimilarity measure, we computed the Euclidean distance between the gray-level values of the pixels of two shapes. This procedure was done for each of the 99 × 99 relative positions, according to the formula

\[
\left[ \left( \sum (G^i - G^j)^2 \right) / n \right]^{1/2} \tag{3}
\]

where \( G^1 \) and \( G^2 \) = gray levels of pixel \( i \) for shape 1 and 2 and \( n = \) number of pixels. The similarity of a stimulus pair was defined as the smallest value of the 99 × 99 positions.

The representation of shape similarities at the neuronal population level was analyzed by computing the Euclidean distance between a pair of stimuli \( i \) and \( j \) in the multidimensional space spanned by the responses of all neurons

\[
\left[ \left( \sum \left| \text{Resp}_{i}(n) - \text{Resp}_{j}(n) \right|^2 / p \right) / n \right]^{1/2} \tag{4}
\]

where \( n = \) cell number and \( p = \) number of recorded cells.

For each stimulus group, the different sets of similarity data were analyzed with nonmetric multidimensional scaling (MDS) using Statistica software.

To determine whether a systematic relationship existed between responses to the nine stimuli on the axes of set 5 (further referred to as the compound stimuli; there are 5 stimuli per axis, but the 2 orthogonal axes have the central stimulus in common, producing 9 distinct stimuli) and responses to the isolated parts of these stimuli (further referred to as the constituent stimuli; set 6), we performed a linear regression analysis per cell with the responses to the compound stimuli as the dependent variable and with the sums of the responses to the respective constituent stimuli as the predictor variable. Additionally, we ran the same regression analysis after pooling the responses of all cells showing shape selectivity for the 10 constituent stimuli to examine the relationship between the responses to the constituent stimuli and to the compound stimuli at a population level. In these regression analyses, a slope of 0.5 would indicate that the responses to the compound stimuli are the averages of the responses to the constituent stimuli, presented in isolation, whereas a slope of 1.0 indicates that the responses to the compound stimuli are the sum of the responses to the constituent stimuli.

We examined the temporal dynamics of both the selectivity among the shapes within one stimulus set at a population level using an information-theoretic approach (cf. Sugase et al. 1999). According to this approach, each predictable piece of information associated with an occurrence of a neuronal response \([H(S); R])\) is quantified as the decrease in entropy of the stimulus occurrence \([H(S)]\)

\[
H(S; R) = H(S) - H(S|R) \tag{5}
\]

where \( S = \) set of stimuli \( s \), \( R = \) set of signals \( r \) (i.e., spike counts), \( p(s|r) = \) conditional probability of stimulus \( s \) given an observed spike.
count \( r \) and \( p(s) \) = a priori probability of stimulus \( s \). The brackets indicate the average of the signal distribution \( p(r) \).

The use of a small number of trials induces an upward bias in the estimation of transmitted information. To correct for this, we subtracted the first-order correction term \( C_f \) from the value calculated using Eq. 5, as \( C_f \) represents almost all the error due to limited sampling (Panzeri and Treves 1996)

\[
C_f = \frac{1}{2N \ln 2} \left( \sum B_s - B - (S - 1) \right)
\]

where \( N \) = total number of stimulus presentations, \( B_s \) = number of nonzero response bins for the presentations of stimulus \( s \), \( B \) = total number of bins, and \( S \) = number of stimuli. Thus the corrected transmitted information, \( I_c \), is defined as follows

\[
I_c = I(S; R) - C_f
\]

To examine the time course of the transmitted information, we computed the neuronal responses using sliding windows of different durations. The results shown in Fig. 9 are based on a window duration of 50 ms and the middle of the window was moved in 5-ms steps beginning 5 ms after the stimulus onset, up to 277 ms. These values are identical to those used by Sugase et al. (1999), facilitating a comparison of our with their results. A shorter time window of 10 ms produced highly similar results. The temporal evolution of the selectivity among shapes from different sets was examined by computing, for each time window, the information transmitted by the neuronal responses to four randomly selected stimuli, each from a different set (for the selection of the 4 stimulus sets, see following text), using Eq. 7. This was repeated 1,000 times. The temporal dynamics of the discrimination among the shapes of one stimulus set was examined by computing, again per time window, the information transmitted by the neuronal responses to four randomly selected stimuli from within that set. This was also repeated 1,000 times per set. This procedure permits excluding a potential bias due to a different number of stimuli or a different number of trials when comparing the transmitted information for stimuli within one set with the transmitted information for stimuli from different sets.

For both selectivity between sets and within sets, the information latency was measured from the stimulus onset to the center of the first of at least three consecutive windows for which the information differed significantly (using a paired \( t \)-test) from the information in the window with the center at 5 ms after stimulus onset. The peak was defined as the center of the window in which the transmitted information, averaged across cells, reached a maximum.

Stimulus sets were chosen for this selectivity time course analysis based on a cluster analysis (Ward’s method; Statistica) of the neural similarity matrix of all neurons and all stimuli (see RESULTS). This clustering algorithm starts from a configuration with as many clusters as stimuli and groups similar stimuli in a series of steps (starting with the most similar ones) until all stimuli are clustered together.

EXPERIMENT 2. All analyses were performed on those neurons showing shape selectivity in the RSVP test for at least one of the two range configurations, as tested by a one-way ANOVA (\( P < 0.05 \)) for each range configuration (of 9 stimuli each).

To compare the neuronal selectivity at the population level for the shapes from the narrow- and wide-range configurations, the stimuli were first ranked based on the difference between the mean response to stimuli A and B, averaged across the two ranges, and the mean response to stimuli D and E of both ranges (see Fig. 2 for definitions of stimuli A, B, etc.). If the former average response was larger than the latter, the stimuli were ranked in ascending order (i.e., A B C D E). If the opposite was true, a descending ranking was used (i.e., E D C B A). The same ranking was used for the nine stimuli. The responses to the nine stimuli of the wide- and narrow-range configurations were fitted with a second-order polynomial least-squares fit.

For all further analyses, we focused on the shapes common to the two ranges, i.e., the A to E stimuli of Fig. 2, to study the effect of stimulus context on the responses and selectivity of our cells. For every cell, the depth of selectivity index (DOS, Eq. 1) was calculated for both the narrow and wide range to quantify the degree of selectivity for the common shapes. To show the evolution of the neuronal selectivity over time, the data were subdivided into blocks of 10 presentations per stimulus and for each range configuration, DOS indices were calculated per block.

To quantify the neuronal ability to discriminate among shapes, we employed receiver operator characteristic (ROC) analysis (e.g., Cohn et al. 1975; Vogels and Orban 1990). For each neuron, ROC curves were generated by computing the distribution of the responses in the different presentations of a stimulus and then computing the proportion of spikes that exceeded a particular response criterion (in steps of 1 spike). The ROC analysis was done using the middle stimulus C and one of the extreme stimuli, either A or E (i.e., the one ranked as having the maximum response for that cell), of experiment 2. The area under the ROC curve generated by the neuronal response distributions for this pair of stimuli yields a score for the neuronal discrimination ability. Perfect discrimination results in an area of 1; random discrimination produces an area of 0.5. Because we were interested in discriminability per se, the lowest value—chance performance—is 0.5. Thus values <0.5 were corrected by subtracting these from 1.

To examine whether the stimulus statistics altered the amount of information carried by the responses of the cells, the information transmitted by the neuronal responses regarding the presentation of shapes A to E (in a 50- to 200-ms time window) was quantified per cell for both the narrow and the wide range using Eq. 7.

RESULTS

Experiment 1

We recorded from 84 neurons (67 from monkey J; 17 from monkey B) using the RSVP procedure with a 100 ms/image presentation rate. Eighty neurons showed shape selectivity within one or more shape sets as measured with a permutation test (see METHODS), resulting in a significant response modulation for a total of 240 shape sets. For each cell, there was an average of 76.52 presentations per stimulus (minimum = 12, maximum = 151).

Across animals, the recording positions were estimated to range from 12 to 16 mm anterior to the external auditory meatus and included the lower bank of the superior temporal sulcus and the cortical convexity lateral to the anterior middle temporal sulcus.

COMPARISON OF SHORT AND LONG PRESENTATION RATES. For 27 of the 80 cells that were selective at a 100-ms presentation rate, stimuli were first presented in an RSVP procedure using 300-ms presentation duration before using the regular 100-ms/image presentation rate. Thus for this sample of neurons, one can correlate the shape selectivity for the fast, 100-ms and slow, 300-ms presentation durations. For both presentation rates, responses were computed using the 50- to 200-ms analysis window. We analyzed the responses to the shapes of those sets (n = 83) for which there was a significant modulation for the standard, fast presentation rate. To qualitatively compare the responses for the two presentation rates, for each neuron, the 25 stimuli within a set were ranked according to the size of their responses in the 300-ms presentation condition. The same ranking was then used to order the stimuli presented.
at 100 ms/image. As shown in Fig. 3A, the mean response in the 100-ms/image presentation rate condition, averaged across neurons, decreased as a function of stimulus rank. A general linear model (GLM) repeated-measures ANOVA with rank as within-neuron factor showed that this modulation for the 100-ms/image sequences was significant \(F(24,1968) = 29.71, P < 0.001\), indicating that the overall ranking was preserved at this fast presentation rate. A similar overall preservation of shape rank was obtained when the stimuli were ranked using the responses of the fast presentation rate \(F(24, 1968) = 30.39, P < 0.001\). To quantify the similarity of the selectivity patterns under both presentation duration conditions, we correlated for each neuron the responses to the 25 shapes in the fast presentation rate with those to the same shapes in the slow presentation rate. The median percent of variance in the response pattern measured at the fast presentation rate that could be explained by the response pattern measured at the slow rate was 0.32 (1st quartile = 0.13; 3rd quartile = 0.56), which corresponds to a correlation coefficient of 0.57. The preceding correlation analysis (as well as the ranking) was performed on the 25 stimuli of a set for which the neuron responded selectively. When the responses to all 135 stimuli were correlated instead, the median correlation between responses to the fast and slow rates was even higher: \(r = 0.75\) (explained variance = 0.57).

The degree of shape selectivity within a set, as measured by the depth of selectivity (DOS; see METHODS), obtained at the fast presentation rate was highly correlated with that for the slow rate: \(r = 0.81\) (explained variance = 0.66). Although highly correlated, the DOS for the slowest rate was significantly higher than for the 100-ms/image rate (averages of 0.45 and 0.34, respectively; Wilcoxon matched pairs test, \(P < 0.001\)). The Spearman-Brown split-half reliability coefficients (see METHODS) for the DOS indices were 0.89 and 0.96 for the slow and fast presentation rate conditions, respectively. The drop in selectivity with increasing presentation pace cannot be attributed to a possible higher noise level for the fast presentation rate condition because the Spearman-Brown split-half reliability coefficient for the fast presentation rate condition was not lower than that for the slow presentation rate condition. This difference in DOS, however, was due to smaller responses to the best shapes and larger responses to the...
nonpreferred shapes in the fast presentation condition. This can be appreciated by comparing the responses for stimulus ranks 1 and 25 for the 300-ms presentation duration in Fig. 3A with those for the 100-ms presentation duration of Fig. 3B (stimulus rank 1: 25 vs. 23 spikes/s; stimulus rank 25: 7 vs. 10 spikes/s). This suggests an underestimation of the degree of stimulus selectivity at the fastest presentation pace.

Note that the ranking curves for the fast and slow presentation of Fig. 3 are more similar for the fast rate reference ranking (Fig. 3B) than when the slow rate is used as a reference (Fig. 3A). This can be explained as follows. Given the imperfect correlation of the responses in the two presentation conditions, the ranking curve for the slow rate will flatten if the shape ranks are based on the fast rate (compare — in Fig. 3, A and B). Also, the ranking curve for the fast rate will flatten when the slow rate is used as the reference (compare ••• in Fig. 3, B and A). Given the higher selectivity for the slow compared with the fast rate, the flattened curve for the slow rate will become somewhat similar to the ranking curve for the fast rate when the latter is used as reference (Fig. 3B), whereas the curves for the two rates become more dissimilar when the slow rate is used as a reference (Fig. 3A).

TUNING WITHIN PARAMETRIC SHAPE SPACES. Most of the neurons recorded showed a preference (i.e., the largest response) for the extreme values of the parametric space (see Fig. 4 for a representative neuron). Instead of showing a uniform distribution across the parametric shape space, the neuronal shape preferences across the population of recorded neurons were concentrated at the extremes of the stimulus dimensions (Fig. 5A). This preference for the extremes was significantly higher than expected by chance: in 215 of the 240 tested shape sets, the maximum response was for an extreme value of the parametric configuration \[ P < 0.001 \] as tested with a Binomial test over all sets; null hypothesis with expected relative frequency \[ 0.64 \] (16/25 extreme stimuli); all \[ P < 0.002 \] for separate Binomial tests per parametric set. We further tested whether the mean normalized responses of the cells showing shape selectivity for that shape set were uniformly distributed using a one-way ANOVA per shape set. For each set, the analyses showed that the mean responses were not uniformly distributed over cells \( (P<0.01 \text{ for sets } 1 \text{ and } 5; \ P<0.001 \text{ for sets } 1, 3, \text{ and } 5) \). Indeed Fig. 5B shows that the neurons responded more strongly to the extreme values of a shape configuration.

![Fig. 4. Schematic representation of the mean response strength (mean number of spikes in hertz counted in a 50- to 200-ms time window after stimulus onset) of an example neuron to all shapes of the shape sets. The ordering of the shapes is the same as in Fig. 1 for sets 1–5. For set 6, the shapes are ordered from top to bottom (i.e., for both the stars and the pentagons, stimulus 1 is the upper shape shown for set 6 in Fig. 1; stimulus 5 is the lower one).](image)

![Fig. 5. Population responses. Every row shows the results for a different shape set (from set 1 to set 6, as labeled in Fig. 1). A: distribution of the relative frequency of the preferred shape (shape producing the maximum response) for each of the sets. Only cells showing a significant modulation for that set are included (number of cells per set is shown). B: schematic representation of the mean normalized responses to all shapes per set, averaged over all cells showing a significant modulation for that set. Responses were normalized to the maximum response per cell. For set 6, the SE is shown for each stimulus.](image)
The tuning for the extremities of the shape configuration was present for each of the five shape spaces and thus not just for the simple shape dimensions (aspect ratio and taper; shape sets 1, 2, and 5) manipulated by Kayaert et al. (2005a). However, it should be noted that the radial frequency components we manipulated resulted in a systematic variation in the degree of indentation along the horizontal and vertical dimensions of sets 3 and 4, respectively. It is apparent from the response profiles that the latter change resulted in the strongest modulation.

We used the responses of the 80 neurons to those shape sets with significant modulation (240 sets) to compute the neuron-based (dis)similarity between each pair of stimuli. To determine how the neurons represent the similarities among the shapes, we analyzed the neural-based Euclidean distances (see METHODS) between the stimuli using MDS and compared the obtained neural-based configurations with the parametric, position corrected pixel-based stimulus configurations (Fig. 6).

Note that the pixel-based configurations preserved the stimulus order of the parametric configurations (Fig. 1) and, in addition, showed that the physical distances among the shapes were similar along the two dimensions in each of the five configurations. Two-dimensional configurations explained most of the variance in the neural similarities for stimulus sets 1–5 [averaged across monkeys: 86% (n = 32 neurons), 91% (n = 52), 88% (n = 59), 87% (n = 61) and 92% (n = 36) for stimulus sets 1–5, respectively].

An inspection of the configurations shown in Fig. 6 clearly suggests that single stimulus dimensions unrelated to shape, such as area, cannot solely explain the observed stimulus selectivity. If, for example, this selectivity could be explained solely by area, for both sets 1 and 2, shapes 1, 5, 9, 17, and 25 (see Fig. 1) should be clustered together, shape 11 should be clustered with shape 15, and shape 19 should be clustered together with shape 23. The configurations of Fig. 6 clearly show that this is not the case.

The stimulus order in the neuron-based two-dimensional (2D) configurations matched the order of the pixel-based configurations for sets 2 and 3. The neuron-based 2D configuration for stimulus set 1 deviated from that of the pixel-based configuration for two stimulus pairs, again demonstrating a good overall fit between physical and neural similarities. Note that for set 3, the neurons represented the horizontal dimension, i.e., “indentation”, more acutely than the vertical one, which fits the distribution of the tuning shown in Fig. 5A. An even more striking difference in sensitivity for the two radial frequency dimensions was present for stimulus set 4. For the latter stimulus set, the sensitivity on the horizontal axis was much weaker than along the vertical indentation dimension, resulting in a highly anisotropic distribution of the stimuli in the two-dimensional space. However, note that along the vertical dimension, the stimulus order is relatively well preserved, indicating that the neurons represent variations along this dimension at the ordinal level.

Stimulus set 5 is a special case because the shapes were compound stimuli consisting of two shapes, each of which was varied systematically along one dimension. Figure 6E shows the 2D configuration for shape set 5 and indicates that there was an overall correspondence between the parametric space and the neural configuration, albeit not as clear as that for sets 1–3. Also this set displayed a strong difference in sensitivity for the two stimulus dimensions: the neurons were more sensitive to variations along the vertical “star” dimension than along the horizontal “pentagon” dimension.

FIG. 6. Multidimensional scaling (MDS-derived 2-dimensional) configurations of stimuli for sets 1–5. The numbers refer to the stimuli as they are labeled in Fig. 1 (set 1). - - - and —, stimuli positioned on the horizontal and vertical meridian, respectively, of the parametric configurations (Fig. 1). Left: configurations based on the position-corrected pixel-based similarity between stimuli. Right: configurations based on the neural-based similarity between stimuli. To aid in a visual comparison of the pixel-based with the neuron-based configurations, the latter were Procrustes transformed (i.e., a combination of translation, scaling, rotation, and reflection). A–E: configurations for sets 1–5, respectively.

COMPARISON OF RESPONSES TO TWO-PART SHAPES AND THEIR SINGLE PARTS. An important question regarding the shapes in set 5 is how the responses to these two-parts shapes relate to the responses to the single parts, i.e., the constituent stimuli. We quantified this relationship for 59 neurons that showed significant shape selectivity for the 10 constituent stimuli (set 6; see METHODS) or for the compound stimuli (set 5) as tested with a permutation test. Figure 7A shows a scatter plot for one of these cells in which the responses to the nine compound stimuli are plotted against the sum of the responses to the constituent stimuli presented in isolation. The responses to the compound stimuli were much smaller than the sum of the responses to the constituent stimuli, indicating a strongly nonadditive relationship between responses to the two-part and single shapes. The
slope of the regression line relating the sum of the responses to the constituent stimuli and the responses to the compound stimuli was 0.51. Thus for this neuron, the responses to the compound stimuli were very close to the average of the responses to the constituent stimuli presented in isolation. As expected from such averaging, the responses to the two-part shape were significantly lower than the responses to the part eliciting the worst response when presented alone \( t(8) = 2.34, P < 0.05 \); Fig. 7B] and significantly higher than the responses to the part eliciting the worst response \( t(8) = -7.14, P < 0.001 \); Fig. 7C] when presented alone.

For the population of 59 neurons, the mean slope of the regressions was 0.47 ± 0.51 (SD), suggesting that this averaging effect was present at the population level. To examine this further, we pooled the data from all 59 neurons and performed the regression analysis on the pooled data. As shown in Fig. 7D, the responses to the compound stimuli were highly correlated with the sum of responses to the constituent stimuli presented alone \( r = 0.94 \). The slope of the regression for the pooled data set was 0.51, which indicates that the responses of a population of IT neurons to the compound stimuli can be predicted, virtually perfectly, as the average of its responses to the constituent stimuli presented in isolation. Again as expected, the responses to the compound stimuli were significantly lower than the responses to the respective best constituent shape \( t(530) = 5.44, P < 0.001 \); Fig. 7E] and significantly higher than the responses to the respective worst constituent shape \( t(530) = -7.26, P < 0.001 \); Fig. 7F].

As discussed in detail by Zoccolan et al. (2005), who reported a similar averaging effect, shifts in attention might explain such an effect. According to this attention hypothesis, the mean response to the compound stimuli will correspond to the average of the responses to the constituent stimuli presented in isolation if the attention of the monkey was directed toward one part (e.g., upper part) for approximately half of the presentations and the other part (e.g., lower part) for the rest of the presentations. This hypothesis assumes that the response distribution across all presentations of each compound stimulus is drawn more or less equally from the distributions of responses to the two constituent shapes. Thus the attention hypothesis predicts that the variance of the distribution of responses to each compound stimulus equals the variance of the distribution that is obtained by combining the response distributions to the constituent shapes when the latter are presented in isolation. Following Zoccolan et al. (2005), we computed the Fano factors, i.e., the ratio of the response variance to the mean response, for the response distributions of each of the compound shapes (observed Fano factor) and compared those to the Fano factors computed for the distributions obtained by combining the responses to the constituent shapes, i.e., the response distributions predicted by the attention hypothesis. In performing this analysis for all 59 neurons, we found that the observed mean Fano factor \( 1.42 \) was significantly smaller than the value predicted by the attention hypothesis \( 1.69 \); paired t-test, \( t(530) = 12.16, P < 0.001 \). When only those compound stimuli were included for which the responses of the neuron to the respective two constituent stimuli differed by a factor of two or more (i.e., where 1 constituent stimulus was much more effective than the other), similar results were found \( t(530) = 12.16, P < 0.001 \). These results indicate that the reported average effect is not likely to be merely the result of shifts in attention.

TIME COURSE OF SHAPE SELECTIVITY. The RSVP paradigm is similar to reverse correlation paradigms that have been used to examine the time course of e.g., orientation and spatial fre-
quency selectivities in earlier visual areas such as V1 (e.g., Bredfeldt and Ringach 2002; Ringach et al. 1997, 2003). As in the latter studies, the present RSVP data can be used to examine the time course of shape selectivity. Because we used different shape sets, we can compare the time course of the selectivity for the shapes belonging to different sets with that of the selectivity for the shapes of a single set. If shapes from different sets are, at least on average, less similar than shapes from the same set, we could use this characteristic to address the question of whether the onset of selectivity depends on shape similarity because we would therefore expect earlier between-set than within-set selectivity.

To assess whether this prerequisite of greater within-set versus between-set similarity had been met, we first performed a hierarchical cluster analysis on the neuron-based similarities that were computed between all possible pairs of stimuli using the responses of all 80 neurons (neural-based Euclidean distances; see METHODS). As shown in Fig. 8, all stimuli from set 1 were assigned to the same cluster before being clustered with stimuli from other shape sets. The stimuli from set 2 were also clustered together first as well as the stimuli from set 5. However, the stimuli of sets 3 and 4 were not cleanly assigned to two separate clusters. Some stimuli from both sets were clustered with one another before being clustered with other stimuli from their respective sets. This is illustrated in more detail in Fig. 8, inset, in which the shapes of the different sets that were clustered are indicated by colored squares. Interestingly, the clustered shapes of the two sets are similar regarding their “blobby” nature. When either set 3 or 4 was excluded from the cluster analysis, the results showed the expected clustering: all stimuli from a set were first assigned to the same cluster before being clustered with stimuli from the other three sets. Therefore, to examine whether the onset of selectivity is earlier for the between-set compared with the within-set selectivity and to meet the prerequisite of larger within-set versus between-set similarity, we excluded shape set 4 from further analyses (exclusion of set 3 instead of set 4 resulted in highly similar findings).

The temporal dynamics of the selectivity within and between shape sets was examined at the population level by evaluating the information about the stimuli that was available in the neuronal responses using a 50-ms sliding window, in steps of 8 ms (highly similar results were found using a 10-ms sliding window and steps of 5 ms). Figure 9 shows the time course of the mean transmitted information measures computed within each of the four sets and between the four sets. The average between-sets transmitted information was, as expected, larger than the within-set transmitted information. Note that at the population level, the average within-set transmitted information differed among sets with sets 1 and 5 showing the least information and sets 2 and 3 the most. This fits with the larger number of neurons showing shape selectivity as assessed by the permutation test (see preceding text) for sets 2 and 3 compared with sets 1 and 5. To compare the within and between sets transmission rates, we analyzed both the latencies and the peaks of the curves. The between-sets information peaked at 149 ms. The information for set 2 peaked at the same time, whereas for the other shape sets, the peak was reached slightly later (at 173 ms, 157 ms and 173 ms for sets 1, 3, and 5, respectively, resulting in a mean within-sets peak of 163 ms). The latency for the between-set transmitted information was 69 ms, whereas the latencies for the within-sets information were 109, 77, 85, and 109 ms for sets 1–3 and 5, respectively (resulting in a mean within-set latency of 95 ms).

Similar results were obtained if the within-sets transmission rates were computed on only those cells showing significant response modulation (instead of on all 80 recorded cells). The latencies for the within-sets transmitted information were 101, 77, 85, and 109 ms (mean within-sets latency = 93 ms), whereas the peaks were reached after 173, 149, 157, and 165 ms for sets 1–3 and 5, respectively (mean within-sets peak =
When we compare these values to the between set latency and peak of 69 and 149 ms, respectively, then it is clear that although overall the within-set selectivity takes somewhat longer to develop than the between-set selectivity, this difference can be rather small [minimum latency difference of 8 ms (set 2) and minimum peak difference of 0 ms (set 2)].

Experiment 2

We recorded from 46 neurons (26 from monkey J; 20 from monkey B), 40 of which showed shape-selectivity within at least one of the two range configurations. For 22 neurons, the wide-range configuration was presented first. For the remaining 18 cells, the narrow-range configuration was presented first. For the stimuli of the first-presented configurations, an average of 808.31 presentations per stimulus was obtained per cell (minimum = 595, maximum = 961). For the stimuli of the configurations presented second, 1,513.48 presentations were shown on average per cell (minimum = 590, maximum = 4,027).

Across animals, the recording positions were estimated to range from 12 to 16 mm anterior to the external auditory meatus. All neurons, except one that was recorded from the lower bank of the STS, were from the cortical convexity lateral to the anterior middle temporal sulcus or from the lip of the STS (area TEM) (Seltzer and Pandya 1978).

Comparison of RSVP and Slow, Intermittent Stimulus Presentation. As in experiment 1, we wanted to check whether stimulus selectivity is preserved at the fast RSVP rate of 100 ms/image. Each neuron was tested in the search test using the same stimuli as in the subsequent RSVP condition. In that search test, stimuli were presented for 300 ms with an ISI of ≥700 ms. These are standard procedures for testing the responses of visual neurons. As for experiment 1, we used a ranking procedure to qualitatively compare the selectivity in the two testing procedures. It is important to note that in the search test, a minimum number of trials were presented per stimulus, just enough to check which shape set elicited the best responses. We compared the responses in the two testing procedures for those neurons (n = 39) for which in the search test, the stimuli were presented at least twice (median number of presentations = 3; 1st quartile = 2; 3rd quartile = 4). For both testing procedures, the responses were quantified using an identical 50- to 200-ms analysis window. We therefore ranked the responses to the stimuli of a given shape set in the RSVP test according to the responses in the search test. As shown in Fig. 10, the responses to the stimuli presented in the RSVP test monotonically decreased as a function of the stimulus rank determined in the search test. This effect of stimulus rank was significant [repeated-measures ANOVA; \( F(8,304) = 10.01, P < 0.001 \)], demonstrating that, at the population level, the stimulus preference at the 100-ms/image RSVP rate was overall similar to that obtained when using the intermittent presentations. Note that the average response level was considerably lower for the RSVP than for the intermittent presentations. The stronger forward and backward masking and stronger repetition suppression (see following text) in the RSVP compared with the slow intermittent presentations are the most likely factors contributing to this difference in the overall responsiveness.

To quantify the similarity of the selectivity patterns under both presentation duration conditions, we correlated for each neuron the responses to the nine shapes in the search test with those to the same shapes in the RSVP test. The median correlation coefficient was 0.52 which corresponds to a median explained variance of 0.27 (1st quartile = 0.03; 3rd quartile = 0.70).

The DOS indices obtained in the search test correlated significantly with those obtained in the RSVP test \( r = 0.35 \) \((P < 0.05; n = 39)\). The Spearman-Brown split-half reliability coefficients for the DOS-indices were 0.87 and 0.97 for the search test and RSVP test, respectively. Thus the low correlation between the DOS of the search and RSVP tests is not due to a low reliability of the DOS measure in the tests but reflects a genuine change in the degree of selectivity. The mean DOS for the search test was significantly higher than for the RSVP test (0.32 and 0.18, respectively; Wilcoxon matched pairs test, \( P < 0.001 \)).
EFFECT OF STIMULUS STATISTICS ON STIMULUS SELECTIVITY. We performed analyses on the stimuli common to the two range configurations (stimuli A–E in Fig. 2) to study the effect of the stimulus distribution statistics, i.e., stimulus context, on the responses and selectivity of all tested cells. First, we performed a repeated-measures ANOVA on the neuronal responses for these common shapes with order of sets as a between-neurons variable (wide range first or narrow range first) and stimulus rank and range (wide or narrow) as within-neurons variables. Stimulus rank was determined as described in METHODS. As expected from the stimulus ranking procedure, the main effect of rank was significant ($F(4,152) = 24.77, P < 0.001$). There was no main effect of range ($F < 1$), but more importantly, the interaction range × rank was significant ($F(4,152) = 7.53, P < 0.001$), indicating a difference in selectivity between the sets with different ranges.

To elaborate on this result, we compared the slopes of the polynomial fits to the population responses ($n = 40$ neurons) for the nine stimuli (Fig. 11A): the slope of the polynomial fit for the narrow range was steeper than the slope of the polynomial fit for the wide range (narrow range: $y = 0.29x^2 - 3.87x + 32.03$; wide range: $y = 0.10x^2 - 1.82x + 27.35$), indicating that the narrower range increased the selectivity of the same shapes. The polynomial fits for the two set orders are shown separately in Fig. 11A, insets. For both orders, the slope of the polynomial fit for the narrow range was steeper than the slope of the polynomial fit for the wide range (narrow range first: narrow range: $y = 0.21x^2 - 2.89x + 28.67$; wide range: $y = 0.04x^2 - 1.15x + 25.93$; wide range first: narrow range: $y = 0.38x^2 - 4.86x + 35.39$; wide range: $y = 0.16x^2 - 2.50x + 28.77$). Additional repeated-measures ANOVAs on the neuronal responses for the common shapes for both orders separately with stimulus rank and range (wide or narrow) as within-neurons variables confirmed that, for both orders, the main effect of rank and the interaction range × rank were significant (wide range first: rank: $F(4,84) = 19.25, P < 0.001$; range × rank: $F(4,84) = 7.88, P < 0.002$; narrow range first: rank: $F(4,68) = 8.87, P < 0.001$; range × rank: $F(4,68) = 3.31, P < 0.02$), whereas the main effect of range did not reach significance ($F < 1$ and $F(1,17) = 2.42, P > 0.10$ for, respectively, the wide range and narrow range first). Note that the overall response level appears to depend on the order (compare Fig. 11A, insets a and b), but statistical testing showed that this interaction between range and order did not reach significance (range × order: $F(1,38) = 2.77, P > 0.10$). In fact, there was no significant main effect of order of sets ($F < 1$), and no interaction-effects with order were significant (rank × order: $F < 1$; range × rank × order: $F < 1$).

This difference between selectivities for the two range configurations could also be demonstrated by the significantly smaller (Wilcoxon matched pairs test, $P < 0.001$) DOS index for the wide-range configuration as a function of the overall response level (maximum information transmission rate possible, $2.32$ bits). Same conventions as in B.

**FIG. 11.** Shape selectivity compared for 2 stimulus ranges (experiment 2) at the population level. A: comparison of the ranked responses to stimuli from the 2 range configurations (narrow range: ●; wide range: ▲; both with SEs). — and --, 2nd-order polynomial fit for the narrow ($R^2 = 0.998$) and wide range ($R^2 = 0.993$), respectively. Second-order polynomial fits for the split data, based on whether the narrow range or the wide range was tested first, are shown in the insets a and b, respectively (narrow range first: $R^2$ narrow range = 0.995, $R^2$ wide range = 0.998; wide range first: $R^2$ narrow range = 0.994, $R^2$ wide range = 0.993). B: depth of selectivity (DOS) indices for the wide-range configuration as a function of the DOS indices for the narrow-range configuration. Each point represents 1 neuron. ●, neurons for which the narrow-range configuration was presented before the wide-range configuration; ▲, neurons for which the opposite stimulation order was used. —, equal DOS indices for both range configurations. C: ROC values for the wide-range configuration are plotted against the ROC values for the narrow-range configurations. For both ranges, ROC analyses were performed using I of the extreme stimuli (either A or E, depending on which 1 was ranked as having the maximum response) and the middle stimulus (C). Same conventions as in B. D: in formation transmission rates ($I'$'s) for the wide-range configuration are plotted against the $I'$'s for the narrow-range configurations (maximum information transmission rate possible, 2.32 bits). Same conventions as in B.
To show that not only is selectivity increased for the narrower range configuration but that the discriminability of the shapes is also improved, we employed ROC analysis. The goal of the ROC analysis was to measure the ability of the neurons to discriminate between the middle stimulus C and one of the extreme stimuli A or E (Fig. 2), depending on which of both stimuli was ranked as having the best response. The ROC analysis takes into account not only mean differences in firing rate—as the DOS index does—but also considers the variability of the spike counts for the presentations of a stimulus. For each range configuration, this analysis was applied to the responses of each cell to the stimuli C and A or E of Fig. 2. The ROC values obtained for the narrow-range configuration were significantly larger than for the wide-range configuration (Wilcoxon matched pairs test, $P < 0.001$), indicating that the discriminability of the same shapes was larger for the narrow-range stimulus distribution than for the wide-range stimulus distribution (Fig. 11C). A separate analysis for the two orders excluded the possibility that the difference in the ROC values of the two range configurations is merely a presentation order effect: for each presentation order, the mean ROC value for the wide-range configuration was significantly smaller compared with the narrow range (Wilcoxon matched pairs test, both orders: $P < 0.01$; Fig. 11C). If discriminability of the shapes is improved for the narrow-range configuration over the wide range, one would expect the neurons to transmit more information about the stimuli in the narrow-range configuration as well. For both range configurations, we calculated the information available in the neuronal responses to the five stimuli common to the two configurations in a 50- to 200-ms time window. A Wilcoxon matched pairs test showed that significantly more information was indeed transmitted by the neurons in the narrow-range compared with the wide-range configuration ($P < 0.02$; mean $I_{c}$ narrow range $= 0.025$ bits; mean $I_{c}$ wide range $= 0.018$ bits; Fig. 11D).

If the increase in selectivity for the narrow compared with the wide range results from an adaptation to the stimulus statistics, one would expect this difference to evolve during stimulus exposure. Thus we computed DOS indices in succeeding blocks of 10 presentations per stimulus and did this for the first 30 blocks. These analyses, examining the evolution of the DOS index over time, were performed on the stimuli received by 10.220.32.247 on September 29, 2016 http://jn.physiology.org/ Downloaded from configuration. Population data for the narrow- and wide-range configurations (shapes A–E as labeled in Fig. 2). The computations were performed for the first 300 presentations/stimulus in succeeding blocks of 10 presentations/stimulus for the common shapes of the 2 range configurations (shapes A–E as labeled in Fig. 2). Population data for the narrow- and wide-range configuration are depicted by gray and black lines, respectively. The evolution of the response strength (in the range depicted on the left y axis) is represented by dotted lines. The evolution of the DOS index (in the range depicted on the right y axis) is represented by solid lines (with SEs).

**DISCUSSION**

The present study used RSVP to investigate the tuning of macaque IT cortical neurons to shapes that varied systematically along predefined shape dimensions. A comparison of the RSVP technique using 100-ms presentations with that using longer presentation durations and an ISI showed that the shape preference can indeed be determined using RSVP. Using relatively complex shapes that vary along relatively simple shape dimensions, we found that the large majority of neurons preferred the extremes of the shape configuration, extending the results of a previous study using very simple shapes and

![Fig. 12. Temporal evolution of the DOS index and the response strength for the narrow- and wide-range configurations (experiment 2). The computations were performed for the first 300 presentations/stimulus in succeeding blocks of 10 presentations/stimulus for the common shapes of the 2 range configurations (shapes A–E as labeled in Fig. 2). Population data for the narrow- and wide-range configuration are depicted by gray and black lines, respectively. The evolution of the response strength (in the range depicted on the left y axis) is represented by dotted lines. The evolution of the DOS index (in the range depicted on the right y axis) is represented by solid lines (with SEs).]
standard stimulation procedures (Kayaert et al. 2005a). In addition, a population analysis using MDS showed that for some stimulus sets, the neurons can represent the similarities among the shapes at an ordinal level, extending a previous study that used a much smaller number of shapes and a categorization task (Op de Beeck et al. 2001). However, the same analysis demonstrated that the IT neurons do not faithfully represent the physical similarities among the shapes: the sensitivity of IT neurons can depend strongly on the particular stimulus dimension. We also showed that the IT neurons adapt to properties of the stimulus distribution other than the mean. The degree of shape discrimination by a population of neurons depended on the stimulus distribution statistics. Although the latter effect was small, it was significant and suggests that the tuning of IT neurons is not fixed but adapts to the stimulus statistics at least when stimulated at a high rate with a restricted set of stimuli. The implications of these and other findings of the present study will be discussed in more detail in the following text.

**RSVP versus traditional testing**

RSVP-methods have been used in the retina, LGN, and early visual areas with success (e.g., Berry et al. 1997; Jones and Palmer 1987; Reid and Shapley 2002; Ringach et al. 1997). In these studies, stimuli were shown for short durations (<50 ms) without ISIs, and regression methods (reverse correlation) were used to obtain receptive field maps and stimulus selectivities. Fewer studies have employed this method in higher visual cortex, especially in the ventral visual stream. Studies using backward masking in IT (Kovács et al. 1995; Rolls and Tovee 1994) showed that although the response level declines with decreasing stimulus onset asynchrony, the stimulus preference is preserved. This suggests that fast RSVP can be used to show the stimulus preferences of IT neurons and indeed, subsequent studies by Keysers and colleagues (2001) demonstrated that the stimulus preference of STS neurons is relatively well preserved for fast RSVP sequences. We show in the present study that even for stimuli that are much more similar than those used by Keysers et al. (2001), the shape preference at fast RSVP sequences correlate with those obtained using longer presentation durations (experiment 1) or standard testing procedures (experiment 2). This suggests that RSVP is a useful method for studying the shape tuning of IT neurons.

However, four caveats should be taken into account when using RSVP to study the shape selectivity of IT neurons. First, pilot work using presentation durations shorter than 100 ms produced less reliable results and weaker responses, but this might have been specific to the sort of stimuli used. Thus one should perform control studies with longer stimulus durations when using presentation times <100 ms in IT. Second, we found that the degree of stimulus selectivity was on average somewhat smaller with shorter presentations; this agrees with the finding of Keysers et al. (2001) that increasing the presentation rate results in a flattening of the neuronal tuning. These results thus imply that RSVP might underestimate the degree of shape selectivity. Third, RSVP implies the occurrence of forward and backward masking effects as well as effects of repetition. Each of these three factors will result in decreased responses, and the suppression arising from repetition is likely to be stronger when stimulus similarity within the stimulus sets is larger. The latter is a likely explanation for the large difference in response levels between the standard and RSVP paradigms that we observed for the highly similar shapes of experiment 2. Such repetition-based suppression effects can be reduced by using interleaved presentations of highly different stimulus sets as we did in experiment 1. Fourth, adaptation to stimulus statistics might be more prominent when RSVP is used, given the high number of repeated presentations of a given stimulus set within a short time span. Indeed, we could demonstrate effects of the stimulus distribution on the measured neuronal selectivity (experiment 2) although the significant effects were relatively small and restricted to the degree of shape selectivity and did not extend to shape preference. Thus in general, the present data suggest that although the RSVP method is well suited to studying stimulus preferences (stimulus ranking), estimates of the degree of selectivity (tuning bandwidth) should be made with some caution because these appear to depend on the specifics of the measurement procedure.

**Tuning for extremities of shape configurations**

The very large majority of neurons responded most strongly to a shape located at an extremity of the explored shape spaces. This sort of tuning for extremities was observed for each of the five shape sets in experiment 1. Tuning for the extremities of parametric shape spaces has been described before by Kayaert et al. (2005a), but in that study, the stimuli were much simpler (variations of a triangle and rectangle) than in the present investigation. The present study shows that tuning for extremities is also present for complex shapes when dimensions such as taper, aspect ratio, and “indentation” (amplitude of Fourier descriptor) are varied, at least within the stimulus range explored in the present study.

Kayaert et al. (2005a) suggested that IT neurons show dimension-specific shape tuning and tend to prefer the extremes of these dimensions, i.e., monotonic tuning. Similar monotonic tunings have been reported in IT neurons for faces varying along dimensions such as caricaturization, aspect ratio, and inter-eye distance (Freiwald et al. 2005; Leopold et al. 2006). As yet, the significance of tunings for extremities is unclear (for further discussion, see Kayaert et al. 2005a; Leopold et al. 2006). One might conjecture that IT neurons represent shapes in terms of their distance with respect to extremities along specific shape dimensions. One issue to consider regarding the interpretation of the observed stronger responses for extreme stimuli is that the employed stimuli are likely to be suboptimal for the tested IT neurons. The critical question here is why the extreme stimuli are less suboptimal than the other stimuli given the likely high-dimensional space in which IT neurons are tuned. A satisfactory answer to this important question will require a full description of the nature of the tuning functions of IT neurons as well as knowledge about the relative position and range of the stimulus set with respect to these tuning functions.

The possibility cannot be excluded that IT neurons learn the stimulus statistics of the parametric shape spaces and thus that the observed tunings depend on the stimulation history and the specific stimulus spaces. Experiment 2 demonstrated that the responses of IT neurons can indeed be modified by changes in input statistics. These effects were small in comparison to the
degree of monotonic tuning, but stimulus statistics might exert a more profound effect with more extensive daily repetition of the same stimulus spaces as is the common practice in single-cell recording experiments.

The MDS results clearly show that IT neurons are more sensitive for some stimulus variations (e.g., indentation; stimulus sets 3 and 4) than for others. This is in agreement with previous studies using calibrated sets of shapes (Kayaert et al. 2005a,b). This sort of differential sensitivity for stimulus dimensions is difficult to explain by stimulation-history-dependent mechanisms because the stimulation frequency was equal for the different dimensions. This does not mean that IT tunings are not modifiable by experience: indeed, several studies show that the degree of selectivity of IT neurons is greater for extensively trained than for novel stimuli (Baker et al. 2002; Freedman et al. 2006; Kobatake et al. 1998).

Responses to two-part shapes and their single parts compared

The shapes of set 5 consisted of two parts connected by a thin line, similar to the stimuli used by Baker et al. (2002). Using these so-called “baton stimuli”, Baker et al. (2002) found nonlinear interactions between the constituent parts of the compound stimuli in animals that were trained to discriminate between the compound shapes. In our study, the responses to the two-part shapes could be predicted surprisingly well from the responses to the constituent shapes (set 6) presented at the same visual field locations. In fact, the response to the two-parts shapes was the average of the individual responses to the constituent shapes, which agrees extremely well with Zoccolan et al. (2005). In both studies, the constituent shapes were relatively small (~2°), both shapes were presented close to each other and short presentation durations (100 ms) were employed (although with a 100-ms ISI in the Zoccolan et al. study) while the monkeys were passively fixating. So it appears that under the behavioral conditions and for the small stimulus variations used in our and the Zoccolan et al. study (2005), the shape interactions do not depend on shape identity but depend mainly on response. However, Mis-sal et al. (1999) found that in some IT neurons, shape interactions can depend on the shape. In that study though, the interacting shapes differed more than in the present study, were much larger and were presented at greater eccentricities. The role of these stimulus factors needs to be addressed in further studies. The animals in ours and the Zoccolan et al. study were not trained to discriminate between the compound stimuli. This might be of importance because Baker et al. (2002) demonstrated that the response and selectivity for compound shapes can be affected by the training history of the subjects.

Time course of shape selectivity and shape similarity

Sugase et al. (1999) and Matsumoto et al. (2004) investigated how well single IT responses could categorize stimuli consisting of geometric shapes and monkey or human faces with various expressions. At the population level, global categorization (i.e., categorizing the stimuli as shapes, monkey faces, or human faces) started in the earliest part of the responses, whereas fine categorization (i.e., categorization within a global category, e.g., based on facial expression) occurred, on average, 51 ms later. Tsao et al. (2006) obtained similar results by comparing face categorization (faces vs. objects) and face identification. We examined whether this difference in the time of onset of stimulus selectivity is also present when the stimuli differ in only shape but not in other features and are not biologically relevant categories (such as faces). If the difference in the onset of stimulus selectivity in between-versus within-categories is merely related to average differences in similarity, then one should observe a similar effect for groups of shapes, although in the latter case, all stimuli belong to the global biologically irrelevant category of geometrical shapes. We defined the shape categories using cluster analysis of the neural responses. We found that shape categories thus defined were differentiated slightly faster than the individual shapes of a single set. Because the procedure that we used to compute the latencies differed from the one used by Sugase et al. (1999), we can compare only the relative difference in the onset of stimulus selectivity in between-versus within-categories between both studies. The average between-versus within-category latency difference (26 ms, minimum: 8 ms, maximum: 40 ms) found in the present study was only half the mean latency difference of 51 ms between the global and fine categorization reported by Sugase et al. (1999).

It cannot be excluded that this discrepancy between ours and Sugase et al.’s result might be due to a possible smaller difference in the average similarity between stimuli from the same versus a different category in our study. This conjecture is difficult to verify given the large dissimilarities between the stimulus sets of the two studies.

The shape selectivity peaked at about 150 ms, which fits the peak of the shape discrimination of IT neurons reported by Rollenhagen and Olson (2000) and obtained with a classical testing paradigm. However, it lies between the two values obtained by Tsao et al. (2006) for face categorization (133 ms) and identification (192 ms) and is later than the human versus animals face classification obtained in IT using a similar RSVP technique (Kiani et al. 2005). Thus it is possible that biologically relevant stimuli are discriminated faster by IT neurons than abstract, meaningless shapes, but this should be examined with stimuli calibrated for contrast, luminance and other low-level confounding variables.

Adaptation to stimulus distribution statistics

Experiment 2 is to our knowledge the first demonstration that the measured tuning of IT neurons is influenced by the stimulus distribution statistics: extensive stimulation using very similar stimuli increased the selectivity of IT neurons for these stimuli compared with stimulation with less similar stimuli. This suggests that the selectivity of IT neurons is not fixed but can be dynamically adjusted based on the input statistics. Note that this effect occurred for highly familiar stimuli and thus differs from changes in selectivity observed with the introduction of novel stimuli (Rolls et al. 1989).

Adaptation to input statistics is seen in the visual system of widely different animal species, such as the fly (Brenner et al. 2000) or the guinea pig (Dean et al. 2005). The time course as well as the size of our effect differs, however, from that observed in the fly visual system (Fairhall et al. 2001). The
latter adapts much faster (over a time scale of seconds) and to much a larger degree than that seen in IT, suggesting different underlying mechanisms. The end-result, however, is qualitatively similar: an adaptive rescaling of the input with respect to the stimulus range. Further studies are needed to determine whether IT neurons adapt to other measures of the stimulus distribution such as the mean or higher-order statistics, and whether similar adaptive effects are observed using other stimulation protocols (e.g., with ISIs).

Our experimental design does not allow disentangling the effect of variance from that of the density of the stimuli in a block because we kept the number of stimuli constant in the two ranges. In the narrow-range condition, shapes are packed at higher density and thus are more similar than in the wide-range condition. Repetition suppression for a particular stimulus may depend on the mean similarity of that stimulus to other recently presented stimuli and because the mean similarity was larger for the narrow compared with the wide range, stronger repetition suppression in the narrow condition might have caused the increased selectivity in the latter condition. However, as shown in Fig. 12, there was no clear difference in either the time course or size of the overall response decrease between the two ranges, which does not fit with stronger repetition suppression in the narrow compared with the wide range. One could argue, however, that the change in selectivity results from a similarity-based repetition suppression mechanism that depends on the last one to five or so presentations: by nature of the design, these will have a greater average similarity for the narrow compared with the wide range. Such an explanation runs counter to the observation, shown in Fig. 12, that the difference in selectivity between the two ranges needs ≥50 stimulus presentations to develop. However, it cannot be excluded that similarity-based repetition suppression on a longer time scale produces the observed differences in selectivity between the two ranges. Indeed, whatever the underlying mechanisms, the present data suggest that the observed change in selectivity takes quite a number of presentations to develop, implying a relatively slow adaptation to a change in the input statistics.

Although the effect of stimulus statistics on selectivity was significant, it was rather small. This is a comforting observation since any large effect of stimulus set statistics on neural measures would produce serious methodological problems in assessing neuronal selectivity. Nevertheless, one should be aware of such issues, especially when using high-rate stimulations of frequently reoccurring stimuli.

ACKNOWLEDGMENTS

The technical assistance of P. Kayenbergh, G. Meulemans, and M. De Paep is gratefully acknowledged. We thank S. Raiguel for comments on the manuscript, R. Peeters for help in taking structural magnetic resonance images from our monkeys, and P. De Mazière for help in running some of our analyses on the KULeuven HPC Cluster.

GRANTS

This work was supported by the following grants: IDO/02/004, GOA/2005/18, Genoeseckunde Stichting Koningin Elizabeth. Human Frontier Science Program RGP18/2004 and Detection and Identification of Rare Audiospatial Cues FP6-IST_027787.

REFERENCES


