Vision Res. Vol. 25, No. 8, pp. 1021–1035, 1985
Printed in Great Britain.

0042-6989/85 $3.00 + 0.00

ROLE OF LOW AND HIGH SPATIAL FREQUENCIES IN THE FACE-SELECTIVE RESPONSES OF NEURONS IN THE CORTEX IN THE SUPERIOR TEMPORAL SULCUS IN THE MONKEY

E. T. ROLLS, G. C. BAYLIS and C. M. LEONARD*
University of Oxford, Department of Experimental Psychology, South Parks Road, Oxford OX1 3UD, England

(*Received* 2 *July* 1984; *in revised form* 10 *December* 1984)

Abstract—There are neurons in the cortex in the anterior part of the superior temporal sulcus of the macaque monkey with visual responses selective for faces. One aim of the present study was to analyze further the information which leads them to respond, by measuring their responses to parametrically filtered stimuli. The responses of 32 such single neurons were measured to faces which were digitized, lowpass filtered at spatial frequencies of 2, 4, 8, ... 128 cycles/face, highpass filtered at frequencies of 4, 8, ... 64 cycles/face, and presented in random sequence using a video framestore. It was found that many of the neurons could respond to blurred images of faces, with a mean frequency at half-maximum amplitude of the neuronal response to the series of lowpass filtered images of faces of 3.3 cycles/face. Almost all the neurons had lowpass cutoff frequencies defined in this way below 8 cycles/face. Many of the neurons could also respond to images of faces in which the only information remaining was a limited amount of high spatial frequency edge information. The mean frequency at half-maximum amplitude of the neuronal response to the series of highpass filtered images of faces was 29.7 cycles/face. Almost all the neurons had highpass cutoff frequencies above 8 cycles/face. Thus, many of the neurons could respond to a lowpass and a highpass filtered image of a face even when these had no spatial frequencies in common. The mean separation between the lowpass and highpass cutoff frequencies was 3.2 octaves. For comparison, face recognition in man can be performed with images which contain only information up to 8 cycles/face, or with highpass filtered images which contain only information down to 8 cycles/face. The response of the neurons was not always a smooth function of frequency, but could decrease as higher frequencies were included in the lowpass filtered images of faces, or as low frequencies were included in the highpass filtered images of faces. This indicates that information in certain frequency bands was able to inhibit these neurons. This was particularly likely to occur for the non-optimal face stimulus for a given neuron, indicating that the selectivity of these neurons to different faces was a combination of the excitation produced by some information in faces and inhibition produced by other.

Face Recognition Face recognition Spatial frequency Inferior temporal cortex Monkey Superior temporal sulcus

INTRODUCTION

A small proportion of the neurons in the temporal lobe visual cortex which respond to visual stimuli have responses which occur primarily to faces (Gross *et al.*, 1972; Bruce *et al.*, 1981; Perrett *et al.*, 1982; Desimone *et al.*, 1984). We have described one such population (48 of 497 cells recorded) in the cortex in the superior temporal sulcus (Perrett *et al.*, 1982). The responses of these neurons to faces (real or projected, human or rhesus monkey) were selective in that they were 2–10 times as large to faces as to gratings, simple geometrical stimuli or complex 3-D objects. The responses to faces were excitatory, sustained and were time-locked to the stimulus presentation with a latency of between 80 and 160 msec. The cells were unresponsive to auditory or tactile stimuli and to the sight of arousing or aversive non-face stimuli. The magnitude of the responses of the cells was relatively constant despite transformations such as rotation so that the face was inverted or horizontal, and alterations of color, size or distance. Rotation to profile substantially reduced the responses of 21 cells. Masking out or presenting parts of the face (i.e. eyes, mouth, or hair) in isolation revealed that different cells responded to different features or subsets of features. For several cells, responses to the normal organization of cut-out or line-drawn facial features were significantly larger than to jumbled controls. These findings indicated that explanations in terms of arousal, emotional or motor reactions, or simple visual receptive fields are insufficient to account for the selective responses to faces and face features observed in this population of STS neurons. It was suggested that these neurons are part of a system specialised to code for faces or features present in faces. It is possible (see Rolls, 1981a,b, 1984) that damage to this system is related to prosopagnosia, or difficulty in face recognition, in man (Meadows, 1974; Whiteley and Warrington, 1977; Benton, 1980; Damasio *et al.*, 1982) and to the tameness and social disturbances which follow temporal lobe damage and are part of the Kluver–Bucy syndrome in the monkey (Kluver and Bucy, 1939; Kling and Steklis, 1976).

*Present address: Department of Neuroscience, University of Florida College of Medicine, Gainesville, Florida, U.S.A.

Consistent with these possibilities, it has been found that at least some of these neurons respond differently to different faces, so that they could provide information useful for recognition or for eliciting different emotional and social responses to different individuals (Baylis *et al.*, 1985).

The aim of the study described here was to analyze further the information which leads these neurons to respond by measuring their responses to faces in which defined amounts of information had been parametrically varied by spatial frequency filtering. For example, responses to a series of low-pass frequency filtered faces with different cut-off frequencies (i.e. faces with different degrees of blurring) showed the extent to which different neurons required fine detail in order to respond, as opposed to a blurred face in which the general configuration of features might be preserved but in which no high frequency information remained. Responses were also measured to a series of high-pass frequency filtered faces to determine the extent to which the edge information present in faces might represent sufficient information for the responses of these neurons to occur. In addition to allowing analysis of the information on the basis of which these neurons responded, these experiments allowed comparisons to be made between the spatial frequencies in faces to which these neurons responded and the spatial frequencies necessary for face recognition by humans. Thus Harmon and Julesz (1973; Harmon, 1973) have shown that frequencies below 10 cycles/face image are sufficient for humans to recognize faces to at least some extent. Fiorentini *et al.*, (1984) have extended this by showing that spatial frequencies above 8 cycles/face, as well as those below 8 cycles/face, are adequate for recognizing different individuals' faces. It was thus of interest to compare the spatial frequencies important for the responses of these neurons with those important for face recognition.

In order to provide a completely standardized method of presenting a set of precisely spatial frequency filtered images repeatedly in a random sequence, faces, together with a large range of non-face images, were digitized, spatial frequency filtered on a PDP11 computer, stored on the computer disk, and loaded in random sequence into a video framestore ready for presentation on a video monitor on each trial. This method of stimulus presentation also allowed the responses of the neurons to be measured to sine wave gratings and to boundary curvature descriptors (see Schwartz *et al.*, 1983), in order to provide evidence on whether these neurons could be activated by such stimuli varied systematically over a wide range of parameters.

METHODS

Recording techniques

The activity of single neurons was recorded with glass-insulated tungsten microelectrodes (after Merrill and Ainsworth, 1972, but without the platinum plating) in 3 alert macaque monkeys (Macaca mulatta) (weight 1.7–6.5 kg) seated in a primate chair using techniques that have been described previously (Rolls *et al.*, 1976). The action potentials of single cells were amplified using techniques described previously (Rolls *et al.*, 1979), were converted into digital pulses using the trigger circuit of an oscilloscope, and were analysed on-line using a PDP11 computer. The computer collected peristimulus rastergrams of neuronal activity for each trial and displayed, printed and stored each trial, as well as computing the peristimulus time histogram by summing trials of a given type. To facilitate latency measurements, the cumulative sum distribution was calculated from the sum peristimulus time histogram. For each trial the number of action potentials occurring in a 500 msec period (and a 250 msec period) starting 100 msec after the stimulus onset was printed. This period was chosen because the neurons studied responded to visual stimuli with latencies just greater than 100 msec, and the monkeys consistently fixated the stimuli for this period. Fixation of the stimuli was confirmed using permanently implanted silver/silver chloride electrodes for electro-oculogram recording. The EOG recordings provided eye position with an accuracy of 1–2 deg, and were sampled by the computer every 10 msec and saved with the action potentials for each trial. Data from trials during which the monkey was not already fixating the screen when the stimulus was switched on or during which eye movements of more than 3 deg occurred in the first 600 msec (while the firing rate was being measured) were rejected.

X-radiographs were used to locate the position of the microelectrode on each recording track relative to permanently implanted reference electrodes and bony landmarks. The position of cells was reconstructed from the X-ray co-ordinates taken together with serial 50 μ histological sections which showed the reference electrodes and micro-lesions made at the end of some of the microelectrode tracks.

The monkey's behavioral reactions to the sight of the visual stimuli were monitored through a small hole in the side of the chair.

Stimulus presentation

The visual stimuli were presented in one of two ways. First, stimuli were stored in digital form on a computer disk, and displayed on a video monitor (Microvitec) using a video framestore (Matrox QRGB 256). The resolution of these images was 256 wide by 256 high with 256 gray levels. The monitor provided maximum and minimum luminances of 8.0 and 0.13 ft-L respectively, and was adjusted internally for linearity to within 3% using a photometer. The computer randomized the order of presentation of these stimuli, switched the stimuli on and off for each trial, and synchronized its data collection so that the stimulus was turned on at the start of the 21st bin

of the peristimulus time histogram. This method allowed completely standardized and randomized presentation of quantitatively specified stimuli as diverse as sine wave gratings and faces, and allowed image processing techniques such as spatial frequency filtering and subregion extraction to be applied to the stimuli presented. Second, the stimuli were presented by the opening of a fast rise time (less than 15 msec), large aperture shutter (Compur Electronic 5FM, 6.4 cm aperture) which opened for 1.5 sec after a 0.5 sec signal tone (400 Hz) provided to allow the monkey to fixate before the shutter opened. The stimuli were presented against a uniform background (a large white screen). This method allowed the presentation of three-dimensional stimuli such as real faces and 3-D objects which differed along a wide range of parameters such as size, shape, and color, and also allowed 2-D stimuli such as photographs of a wide range of faces to be presented.

The monkeys performed a visual discrimination task during the testing to ensure that they looked at the stimuli. If a circle, the positive discriminative stimulus (S+), appeared the monkeys could lick to obtain a fruit juice reward, and if a square of the same area and luminance, the negative discriminative stimulus (S−), appeared the monkey had to withhold licking in order to avoid aversive hypertonic saline. A 0.5 sec signal tone (400 Hz) preceded the presentation of the stimulus, and if the monkey was fixating correctly before the stimulus appeared, he had sufficient time to perform the discrimination and obtain multiple licks of the fruit juice tube in the short (1.0 or 1.5 sec) period in which the stimulus was on. This procedure was designed to ensure fixation of the stimuli (Rolls *et al.*, 1979). If any other stimulus appeared (such as a grating, a 3-D object, or a face), then if the monkey licked he obtained fruit juice (i.e. all stimuli except the square were treated as S+). The order of presentation of the stimuli was randomized. The EOG recordings confirmed that this procedure resulted in consistent fixation of the stimuli.

When digitized visual stimuli were being presented on the video monitor, one set of 4–12 visual stimuli were used at a time. Each set of stimuli was designed to provide neuronal response data relevant to one or several hypotheses. For example, one set included 5 different faces, to test whether the neuron responded differently to different faces, and some non-face stimuli such as a sine wave grating, a boundary curvature descriptor (see below), and a complex visual image (see Baylis *et al.*, 1985, Fig. 1), to provide an indication of whether the neuron responded differently to face and to non-face stimuli. Another set consisted of a series of low-pass and high-pass spatial frequency filtered images of one face. Another series consisted of sine wave gratings with different spatial frequencies, and another of boundary curvature descriptors (see below). The computer randomized the sequence in which the members of the set were presented, and after it had presented the sequence once, it restarted the set with another random sequence. The computer was allowed to repeat the set 4–10 times in order to provide sufficient data for an analysis of variance in order to determine whether the neuron responded differently to the different stimuli within the set. After data had been collected on one set, the experimenter then started a different set. Within each set, S− trials appeared with a probability which was usually specified as 0.25, but could be reduced.

Visual stimuli

Face stimuli. Photographs were prepared of macaque monkey faces (looking directly at the camera), and of human faces. The photographic negatives were digitized using a Scandig 3 (Joyce-Loebl Ltd, Gateshead, U.K.) scanning digitizer of photographs, and stored in an image file with a resolution of 256 × 256 × 8 bits, ready for presentation on the Matrox QRGB 256 framestore. The image processing was performed with the Semper image processing software package (Saxton and Koch, 1982) on a PDP11. All computations were performed with data in floating point format, to ensure accuracy. The spatial filtering was performed by a digital Fourier transform convolution with filters circularly symmetric in the spatial frequency domain. In the frequency domain, the profile of the low-pass and high-pass filters was flat with a gaussian taper (see Fig. 1). The cutoff frequencies of the low-pass filters were 2, 4, 8, 16, 32 and 64 cyles/image (see Fig. 1). The cutoff frequencies of the high-pass filters were 4, 8, 16, 32 and 64 cycles/image (see Fig. 1). For the high-pass filters the DC level was added to the filtered image in order to match the mean luminance to that of the original unfiltered image. As the face was adjusted to occupy the width of the image, these frequencies represent cycles/face. No adjustment was made to the contrast of the filtered images, as we

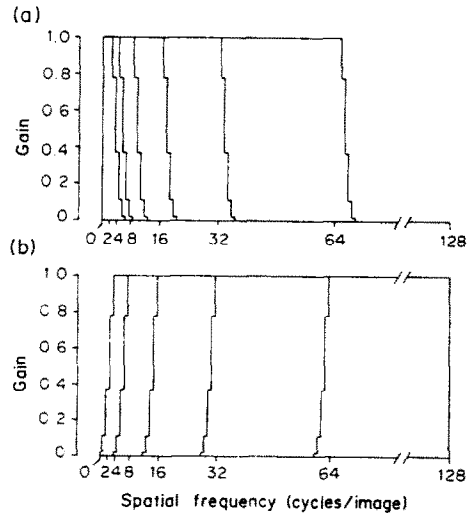


Fig. 1. The shapes of the six low-pass (a) and five high-pass (b) spatial frequency filters.

wished to determine the contribution of the different frequencies present in a normal face to the neuronal response to that face. Examples of the computed filtered images used in this study are shown in Fig. 2. Each set of images included an unfiltered image of the face (the response to which is plotted at 0 cycles/image for the high-pass point and at 128 cycles/face for the low-pass point on graphs since the image resolution was 256 by 256, and thus the highest frequency represented in the digitized original was 128 cycles/image). (In the case of the images with two infants, the spatial frequencies at which points are plotted allows for two faces across the width of the image.)

To determine the extent to which responses to parts of the face might be related to the spatial frequency tuning of a cell, parts of the face were presented in isolation or were blanked out. For each of several faces, series of such parts of faces were stored on disk, and the parts of one face were shown in random series as video images by the computer. The responses to these series of parts of a face were then compared to the responses of the neurons obtained in other series to low-pass and high-pass filtered images of the same face.

Non-face stimuli. The responses of the cells were tested to a wide range of non-face stimuli, as follows.

Sine-wave gratings: a set of sine wave gratings with spatial frequencies of 1–64 cycles/image and with orientations spaced $\pi/4$ radians apart was presented on the video monitor in random sequence. Each image subtended 12 deg at the retina. *Boundary curvature descriptors*: A set of boundary curvature descriptors with frequencies of 0–15 cycles, with amplitudes which ranged from 0.5 to 2.0, and with 4 different phases was presented (Schwartz *et al.*, 1983). *Three-dimensional objects*: Over one thousand three-dimensional junk objects were collected, and 6–30 of these, chosen randomly, were used to test whether there was any indication that a neuron responded to a complex non-face stimulus. If there was any indication of a response, much more extensive testing with non-face stimuli was performed. The objects were chosen to differ from one another in size, shape, color, surface pattern and texture, but for convenience of storage the objects were less than 20 cm long. Since these junk objects varied along different visual dimensions, testing neuronal responses to several of them could potentially reveal selectivity for particular visual characteristics. Objects were positioned between 2 cm and 1 m behind the shutter.

Non-face video images. Over 500 non-face video images of complex scenes were available on the system, and were used in recognition memory tasks (Baylis and Rolls, 1983). Neurons which responded to these stimuli were not classified as face-selective, and were not studied here. One complex non-face image (stimulus H of Baylis *et al.*, 1985) was spatial frequency filtered with the same method used for the face stimuli, and the series of images generated was used to determine whether even for filtered images, the responses of the neurons described here were selective for faces.

Procedure

As tracks were made into the cortex in the superior temporal sulcus, the responses of each neuron were measured to a standard digitized set of stimuli of different faces and of non-face stimuli (Baylis *et al.*, 1985). If a neuron responded to one or more of the faces, but to none of the non-face stimuli in the set, then a wide range of digitized and real 3-D non-face stimuli were shown, to determine whether the response of the neuron was selective for faces. The criterion was that the response to the optimal face stimulus should be more than twice as large as to the optimal non-face stimulus. (In fact, the majority of the neurons in the cortex in the superior temporal sulcus classified as showing responses selective for faces responded much more specifically than this. For half these neurons, their responses to the most effective face was more than five times as large as to the most effective non-face stimulus, and for 25% of these neurons, the ratio was greater than 10:1. These ratios show that while responding preferentially to faces, these neurons do not have absolute specificity for faces. Further information on and discussion of the extent to which these neurons have selective responses is given by Baylis *et al.*, 1985. The non-face stimuli from which the optimal was chosen included sine wave gratings, boundary curvature descriptors, complex 2D stimuli, and complex 3D junk objects, as described in the Method.) If the neuron satisfied the criterion, then the series which contained the low-pass and high-pass filtered images of the optimal face stimulus was run repeatedly until each image in the series had been shown 4–10 times (to allow statistical analysis of the results). Then other series, such as parts of the same monkey's face, or a spatial frequency filtered series of a non-optimal face for comparison, were run.

Treatment of results

For each cell measures of responses were calculated from the total number of action potentials occurring on each trial in the period 100–600 msec following stimulus onset. This period was chosen because the cells studied typically responded to visual stimuli with latencies just greater than 100 msec. Recordings of fixation usually confirmed that the monkeys fixated during this period of firing rate measurement, but trials with poor fixation were rejected from the analysis.

Analyses of variance were then performed on the responses of each cell to the different stimuli. If a significant difference between the responses to the different stimuli was indicated, then subsequent multiple t, Tukey, and Newman–Keuls' analyses (see Bruning and Kintz, 1977) were performed to determine how the different stimuli differed in their

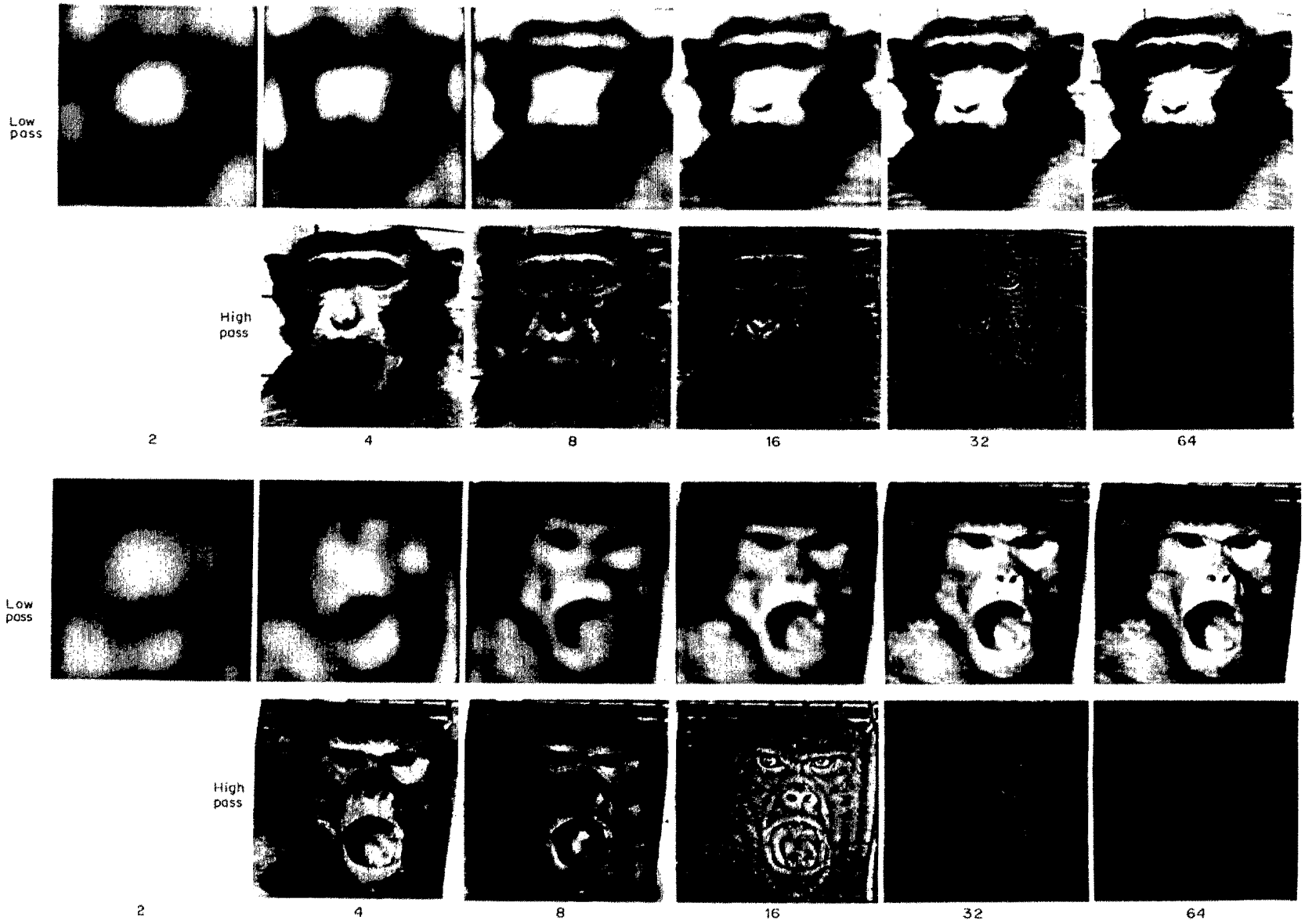


Fig. 2. Examples of the spatial frequency filtered face stimuli.

Fig. 9. Examples of images of parts of faces.

efficacy. One analysis of variance was performed over the responses to a wide range of non-face and face stimuli, to determine whether a neuron responded differently to the face as compared to the non-face stimuli. Other analyses of variance were performed to analyse the differences of response to stimuli within one set, such as different low-pass and high-pass filtered images of the same face. In the figures, the mean firing rate and its standard error to each stimulus based on typically 4–10 presentations of the stimulus are shown. Pearson's product-moment correlations are cited as values of r in the text.

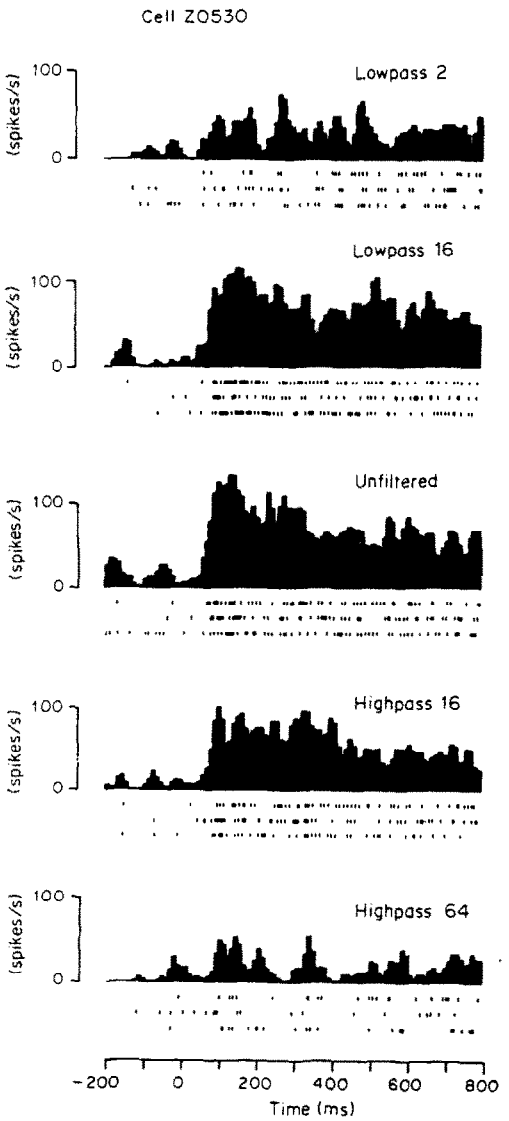


Fig. 3. Examples of the responses of one of the neurons to some of the stimuli used. Three peristimulus rastergrams, and histograms based on 6 presentations of the stimulus, are shown for each stimulus. The onset of the visual stimulus was at time 0. The bin width was 10 msec. Low-pass 2 corresponds to the face low-pass filtered at a frequency of 2 cycles/face.

RESULTS

It was possible to measure the responses to low-pass and high-pass filtered images of faces of 32 neurons in the cortex in the superior temporal sulcus. The recording sites of the neurons from which this subpopulation was drawn are shown elsewhere (Baylis *et al.*, 1985), and were in many cases in region TPO (see Seltzer and Pandya, 1978). Peristimulus firing rate histograms and rastergrams illustrating how one cell (Z0530) responded to some of the stimuli used are shown in Fig. 3. An example of one of the response functions (that of neuron Z0530) is shown in Fig. 4, with standard errors for each experimental point, and the response functions of the other neurons are shown in Fig. 5 in more compressed form. The responses are shown as the change of firing rate from the spontaneous value produced by each of the filtered images. Each point represents the mean of 4–10 observations. For the low-pass filtered images (see solid circles in Figs 4 and 5), the response to the most blurred image (2 and below cycles per stimulus) is frequently small, and as the low-pass filter includes more and more high frequencies as the curve is followed to the right, the response usually increases. The unfiltered image is plotted at the 128 cycles/stimulus point on the low-pass graph, as this represented all frequencies in the 256 by 256 pixel image. For the high-pass filtered

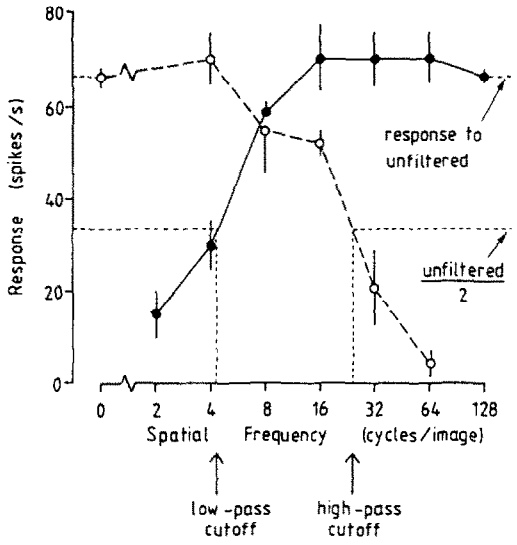
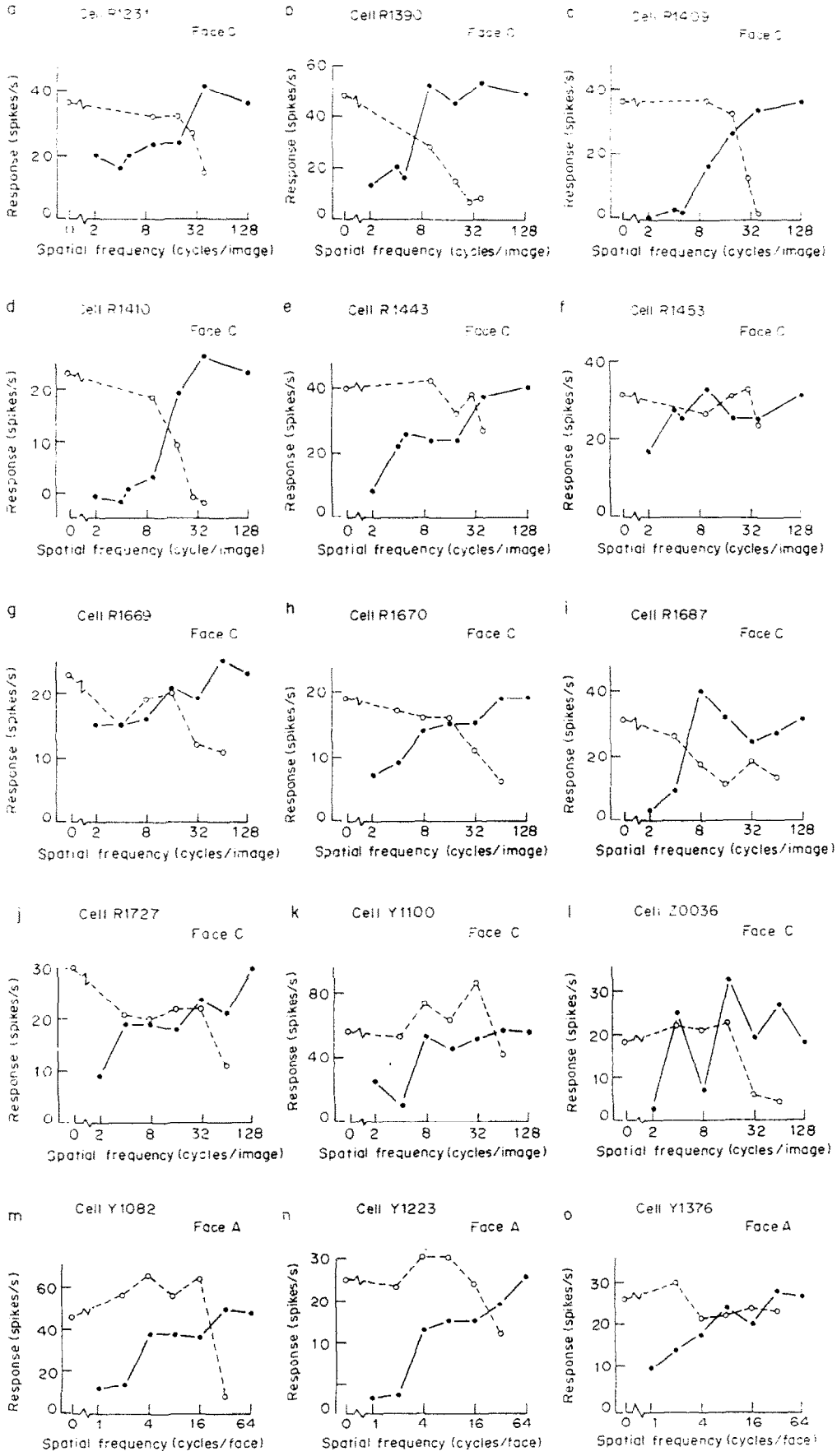


Fig. 4. An example for one cell (Z0530) of the response to low-pass (solid circles, continuous curve) and high-pass (open circles, dashed curve) filtered stimuli. Each point represents the mean of 4–7 presentations, and vertical bars show the standard error of the mean. The figure also illustrates the measurement of the low-pass and high-pass cutoff spatial frequencies. For example, the high-pass cutoff frequency, indicated by the right-hand vertical dashed line, is the frequency at which the firing rate of the cell was half that (see horizontal dashed line "unfiltered/2") to the unfiltered face (see horizontal dashed line "response to unfiltered").


a
Face C
Response (spikes/s)
Spatial frequency (cycles/image)
b
Cell R1390
Face C
Response (spikes/s)
Spatial frequency (cycles/image)
c
Face C
Response (spikes/s)
Spatial frequency (cycles/image)
d
Face C
Response (spikes/s)
Spatial frequency (cycle/image)
e
Cell R1443
Face C
Response (spikes/s)
Spatial frequency (cycles/image)
f
Cell R1453
Face C
Response (spikes/s)
Spatial frequency (cycles/image)
g
Cell R1669
Face C
Response (spikes/s)
Spatial frequency (cycles/image)
h
Cell R1670
Face C
Response (spikes/s)
Spatial frequency (cycles/image)
i
Cell R1687
Face C
Response (spikes/s)
Spatial frequency (cycles/image)
j
Cell R1727
Face C
Response (spikes/s)
Spatial frequency (cycles/image)
k
Cell Y1100
Face C
Response (spikes/s)
Spatial frequency (cycles/image)
l
Face C
Response (spikes/s)
Spatial frequency (cycles/image)
m
Cell Y1082
Face A
Response (spikes/s)
Spatial frequency (cycles/face)
n
Cell Y1223
Face A
Response (spikes/s)
Spatial frequency (cycles/face)
o
Face A
Response (spikes/s)
Spatial frequency (cycles/face)

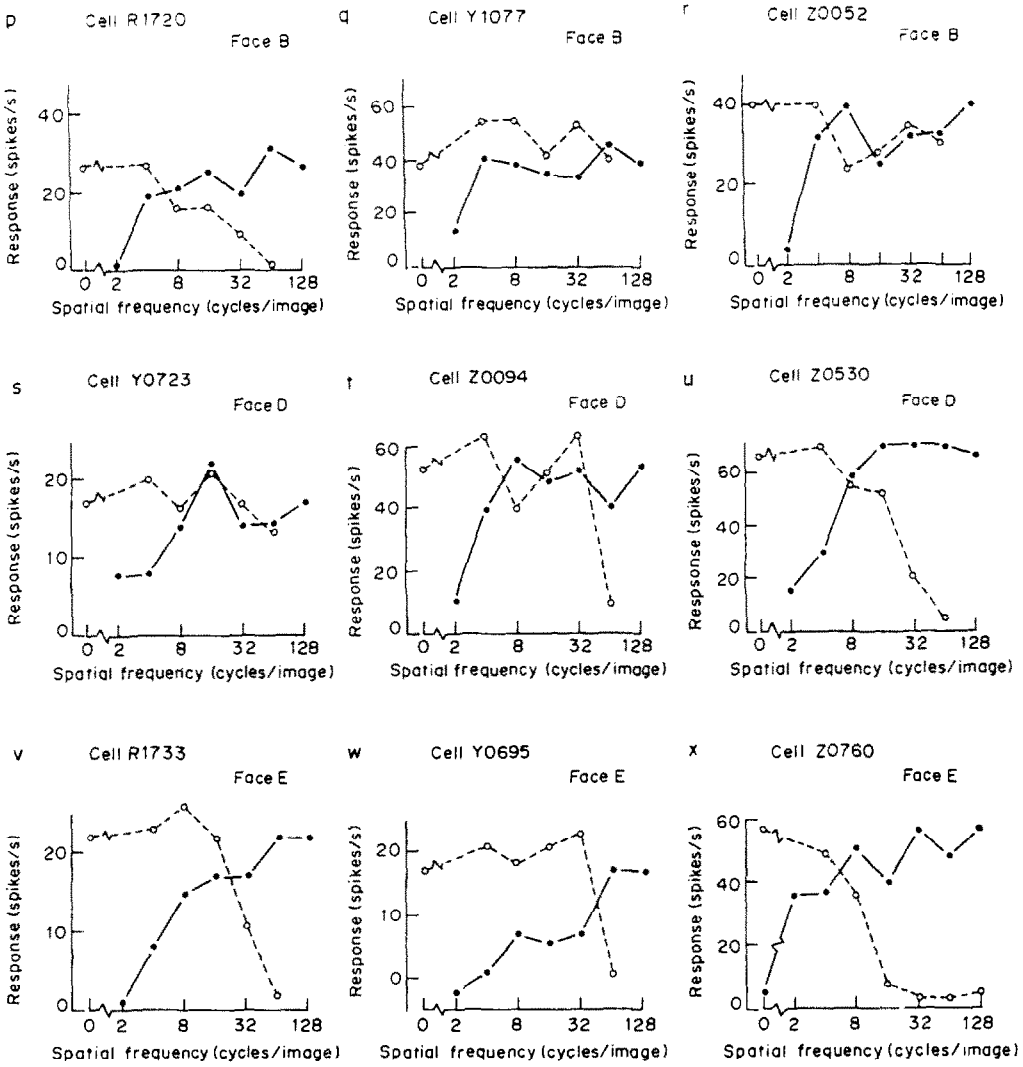


Fig. 5. Responses to low-pass and high-pass filtered stimuli of faces of the 24 neurons for which measurements were made with one series of filtered images, namely the series made from the optimal face stimulus. For clarity standard errors are not shown.

images (see open circles in Figs 4 and 5), the response to the image with only the highest frequencies included (64 and above cycles per image) is frequently small, and increases as the high-pass filter includes more and more low frequencies as the curve is followed to the left. The unfiltered image is plotted at the 0 cycles per stimulus point on the low-pass graph, as this represented all frequencies in the 256 by 256 pixel image. The significance of the effect of different frequencies on the response magnitude was confirmed for every neuron with analyses of variance. To enable the response functions of the different neurons to be compared, high-pass and low-pass cutoff frequencies were obtained from the curves as follows. The cutoff frequencies were measured from the curves as the frequencies at which the response had decreased to half its value to the unfiltered stimulus. Interpolation was used along the logarithmic frequency axis. Thus for the neuron illustrated in Fig. 4, the low-pass cutoff frequency was $2^{2.1}$ cycles/face (i.e. 4.3 cycles/face), and the high-pass cutoff frequency was $2^{4.6}$ cycles/face (i.e. 24.3 cycles/face). Although the curve is not very steep, and is more like a rolloff, use of this measure is conventional (Kulikowski *et al.*, 1982), and does allow the tuning of the different neurons to be compared. The difference between these cutoff frequencies was also calculated to give an estimate of the bandwidth of the neuron, and for the example shown in Fig. 4 was 4.6–2.1 (i.e. 2.5) octaves.

It is shown in Figs 4 and 5 that as the width of the low-pass filter (solid circles) is increased from low values upwards (i.e. as the image becomes less blurred), the majority of the neurons started to respond more vigorously, with quite sharp increases in response frequently becoming evident as frequencies up to 4 or 8 cycles/stimulus for the different

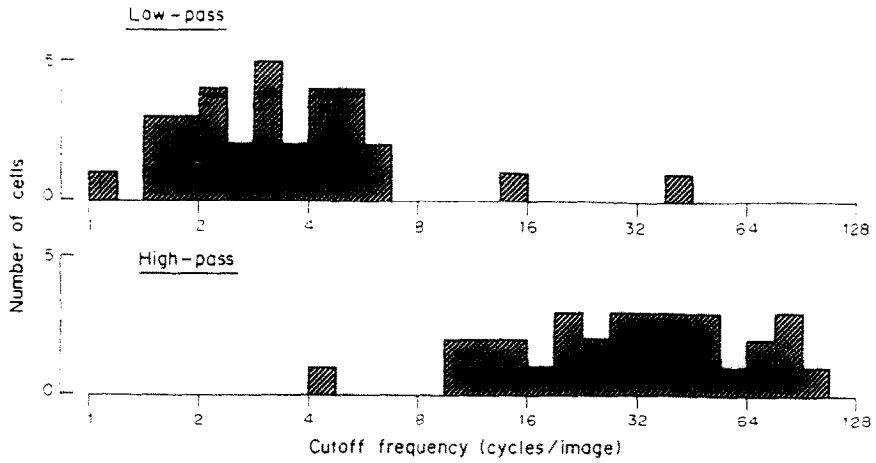


Fig. 6. Histograms showing the distribution of the low-pass (a) and high-pass (b) cutoff frequencies of all 32 neurons analyzed, when the optimal face stimulus was shown.

neurons were included in the stimulus. The mean position on these curves at which the response amplitude was half that to an unfiltered face was 3.3 cycles/stimulus. (This was calculated using the frequency obtained for each neuron by interpolation along the logarithmic frequency axis). The median value was 3.1 cycles/image. The distribution of these cutoff frequencies for the low-pass filtered stimuli is shown for the different neurons in Fig. 6. It is clear from Figs 4, 5 and 6 that the low-pass cutoff frequencies and the pattern of the responses to the low-pass filtered stimuli were different for the different cells. For every cell studied, the ANOVA showed a significant effect of spatial frequency.

As the width of the high-pass filter was increased from high values downwards (i.e. as the image included more than just the sharpest edges), the majority of the neurons started to respond more vigorously, with quite sharp increases in response frequently becoming evident as frequencies (from 128 cycles/image) down to 32, 16 or 8 cycles/stimulus for the different neurons were included in the stimulus (see Figs 4 and 5, solid circles). The mean position on these curves at which the response amplitude was half that to an unfiltered face was 29.7 cycles/stimulus. (This was calculated using the frequency obtained for each neuron by interpolation along the logarithmic frequency axis.) The median value was 29.3 cycles/image. The distribution of these cutoff frequencies for the high-pass filtered stimuli is shown for the different neurons in Fig. 6. It is clear from Figs 4, 5 and 6 that the high-pass cutoff frequencies and the pattern of the responses to the high-pass filtered stimuli were different for the different cells.

It was found that many of the neurons could respond to a low-pass and a high-pass filtered image of a face even when the filtered images had no frequencies in common. This is shown in Figs 4 and 5 in which the frequencies at which the response amplitudes were half those to an unfiltered face were separated by 1–3 octaves in many cases. The mean separation between these frequencies was 3.2 octaves, calculated over the different neurons. Thus many of these neurons could respond to a blurred image of a face, and a face with primarily high spatial frequency edge information in it, when these images were so different that their nearest common spatial frequencies were 3 octaves apart. The number of octaves which separate the low-pass and the high-pass cutoff frequencies for the different neurons is shown in Fig. 7. Interestingly, there was no correlation between the low-pass cutoff frequency and the high-pass

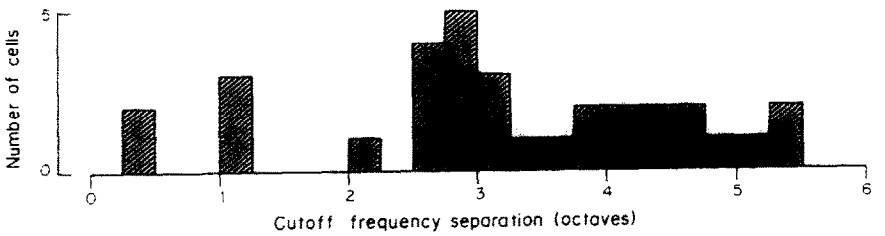


Fig. 7. Histogram showing the separation of the low-pass and high-pass cutoff frequencies of all 32 neurons. Each point represents data from one neuron.

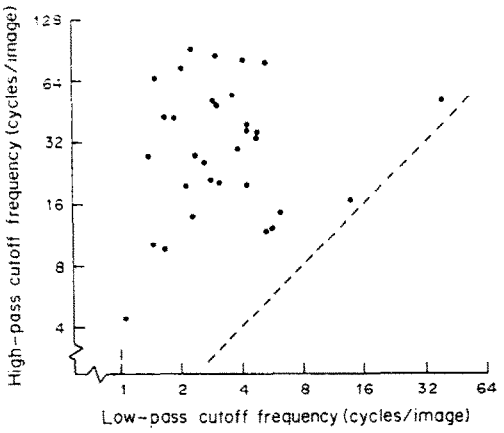


Fig. 8. Each point shows for one neuron the measured high-pass and low-pass cutoff frequencies. There is little correlation between these cutoff frequencies, and it is notable that the low-pass cutoff frequency is never higher than the high-pass cutoff frequency. The dashed line is the $y = x$ line.

cutoff frequency (defined as above) of each neuron ($r = 0.14$, d.f. = 30) (see Fig. 8).

It was of interest to analyse whether the response of the neurons to these low-pass and high-pass filtered images of faces correlated with their response to different parts of the face. For example, it was possible that neurons which could respond to a part of the face such as an eye presented alone might do so on the basis of relatively high spatial frequencies contained in such parts of a face. The parts of the face which were presented in random series included eyes alone, mouth alone, nose alone, and the face without these parts (see examples in Fig. 9). When the proportion of these parts to which the neuron responded with a response greater than half that to a whole face was compared with the high-pass filter cutoff frequency, a highly significant correlation was found ($r = 0.69$, d.f. = 20, $P < 0.001$). Thus neurons which tended to respond well to individual parts of faces could respond to a high-pass filtered image of a face with only relatively high frequencies in it (see Fig. 10). The correlations between the fraction of the response to a whole face which could be obtained with one of the parts eyes, nose and mouth with the high-pass cutoff frequencies were also in some cases significant (with eyes, $r = 0.42$, d.f. = 20, $P < 0.05$; with nose, $r = 0.74$, d.f. = 10, $P < 0.01$; with mouth, $r = 0.51$, d.f. = 10, $P < 0.1$; with nose and mouth, $r = 0.53$, d.f. = 8). There was no correlation between the low-pass cutoff frequencies and the responses to these parts of the face. Because of the correlation between the high-pass cutoff frequency and the responses to parts of the face, there was some correlation between the octave separation of these neurons and whether they responded to parts of the face.

It was also of interest to analyse whether the "octave separation", which provided a measure of the extent to which a neuron could utilize widely separated spatial frequencies (see Fig. 7), correlated with other response properties of the neurons. One possibility for example was that neurons with a wide octave separation might respond to a relatively large proportion of different faces. The generalization index was defined as the proportion of faces in the standard digitized set of 5 different faces to which the neuron responded with greater than half its response to its optimal face stimulus. It was found that there was no correlation between the octave separation and the generalization index ($r = 0.01$, d.f. = 24). (Nor was there any correlation between the generalization indices of the neurons and their low-pass and high-pass cutoff frequencies.) There was also no correlation between the discriminability between the most and least effective faces in the standard set of faces (i.e. d', see Baylis, Rolls and Leonard, 1985) or the breadth of tuning to the set of faces (i.e. H, see Baylis *et al.*, 1985) and the spatial frequency separation index and cutoff frequencies.

The responses to low-pass and high-pass filtered faces described above were to the optimal face stimulus in the digitized series. Different types of response were obtained if the neuronal response was measured to low-pass and high-pass images from a face which was not optimal for that neuron. Several examples of the differences are shown in Fig. 11. For example, neuron R1644 [Fig. 11(a)] simply responded less to most of the filtered stimuli of the non-optimal face as compared to the optimal face. Neurons Z0060 and Z0117 [Fig. 11(c) and (f)] did not respond to the filtered images of the non-optimal face. On the other hand, neurons Z0527, Z0522 and R1731 [Fig. 11(b–d)] responded to some but not to other of the filtered stimuli of the non-optimal face stimulus. This is consistent with the possibility that information

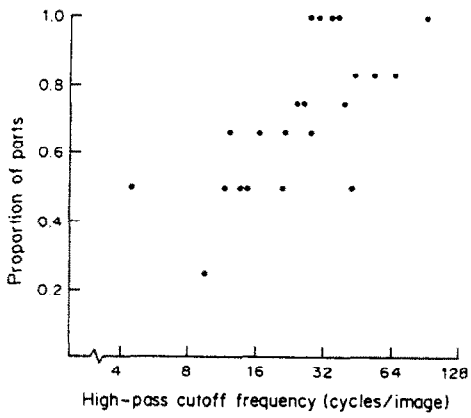


Fig. 10. The proportion of the set of the series of parts of faces to which each neuron responded plotted as a function of the high-pass cutoff frequency of that neuron. Neurons which responded to many of the individual parts of faces tended to have high high-pass cutoff frequencies (see text).

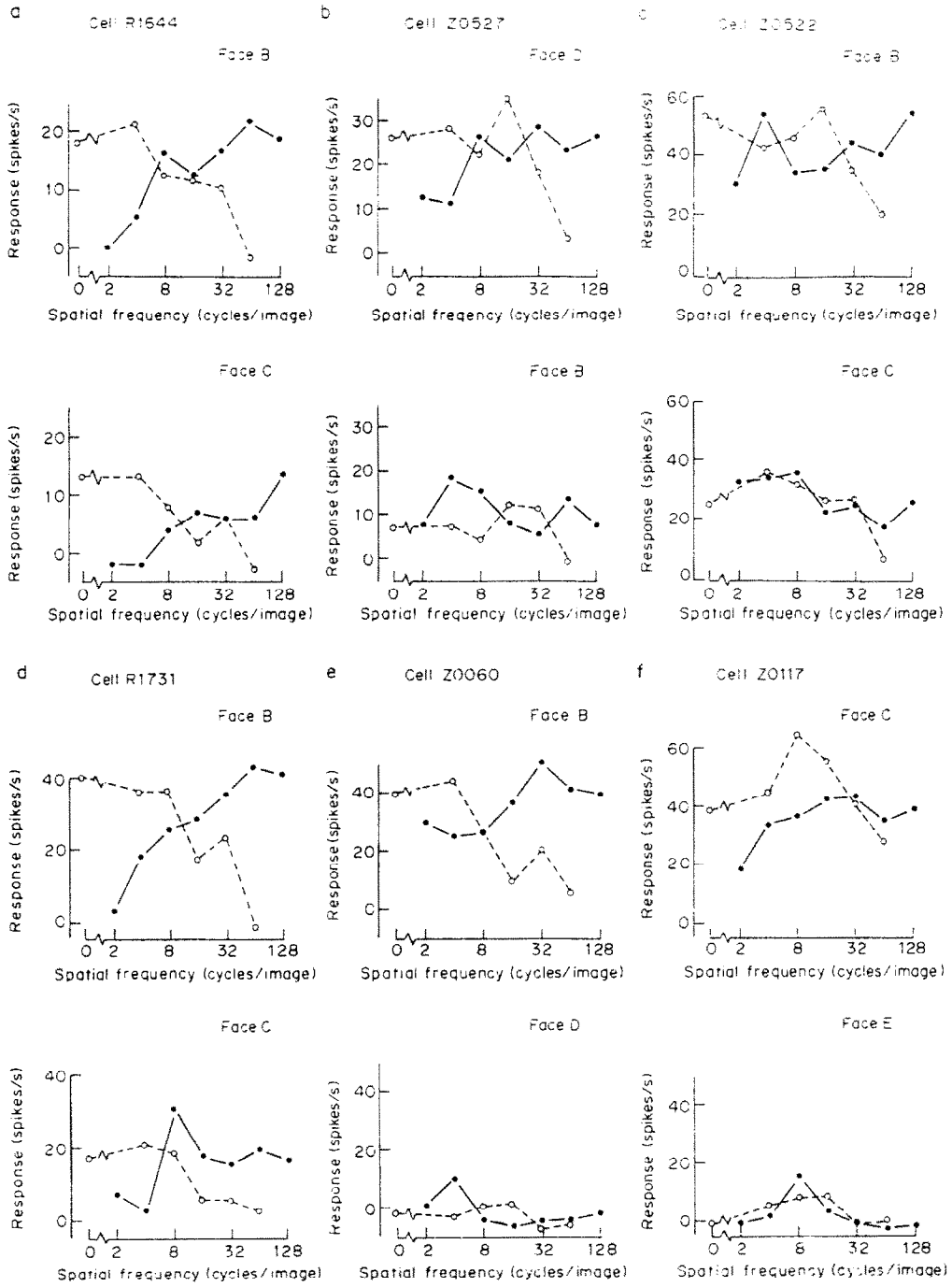


Fig. 11. Responses to low-pass and to high-pass filtered images of the optimal face stimulus (above in each case) and a non-optimal face stimulus (below), for 6 neurons for which responses to filtered stimuli of both optimal and non-optimal face stimuli were measured.

relevant to some aspects of the neuron's criteria for responsiveness were met by some of the filtered stimuli, but that introducing further frequencies from the non-optimal face introduced information which produced inhibition and was normally responsible for the lack of response of the neuron to the non-optimal face.

It was found that the images in the non-face spatial frequency filtered series were not effective in eliciting responses from nineteen of these neurons tested. (Typically these non-face filtered images produced no significant change of firing rate, and in the few cases where there was a small change of firing rate, this was not greater than one fifth of the response to an

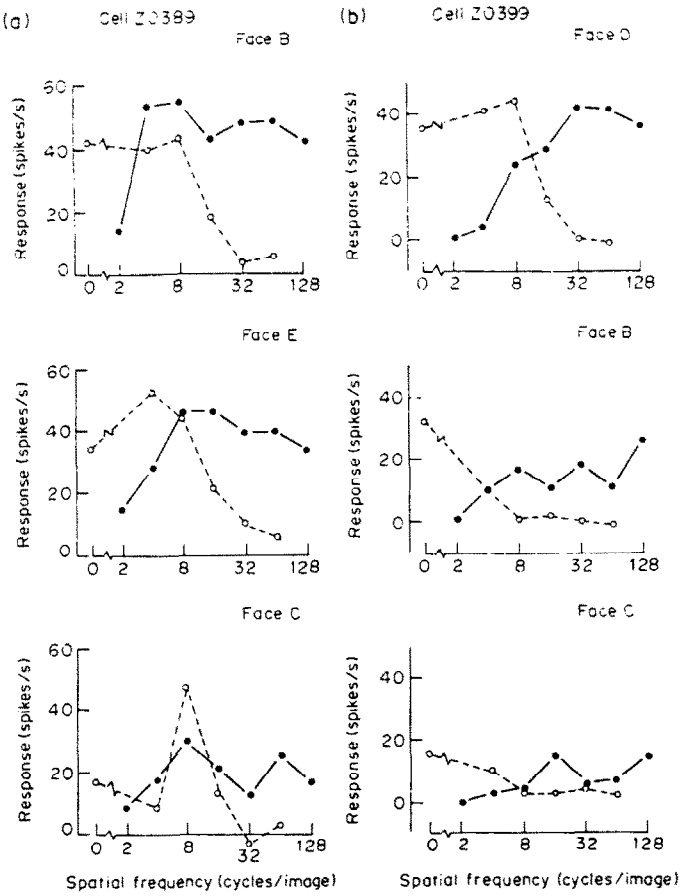


Fig. 12. Responses to low-pass and high-pass filtered images of the optimal, median, and least effective face stimuli for two neurons in which it was possible to measure the responses to filtered stimuli of all three types.

effective face stimulus.) Thus their responses to the spatial frequency filtered images of faces represented responses which occurred primarily to information contained in an image of a face, even when this information was reduced by spatial frequency filtering.

DISCUSSION

The responses of these neurons to low-pass and high-pass filtered faces show that these neurons can in many cases respond to simplified images of faces, that is to faces in which considerable information has been removed. The responses to these filtered images also indicate the types of information necessary to activate these neurons. For example, many of the neurons responded to low-pass filtered faces which included frequencies up to 8 cycles/face (see Figs 4, 5 and 6). These images are very blurred (see Fig. 2) and contain relatively little information. This opens the way for further analyses of the minimal stimulus configuration necessary to activate these neurons. For example, it will be of interest to determine how rearrangement of different parts of these low-pass filtered images of faces influences the responses of these neurons, to determine how sensitive these neurons are to the configuration of the features present in these images. Many of the neurons also responded to high-pass filtered images of faces which included only frequencies above 16 cycles/face (see Figs 4, 5 and 6). Thus low frequencies are not essential for the activation of these neurons, and edge information is sufficient. It is of interest that in many cases a neuron could be activated by either low frequencies or high frequencies present in a face (see Figs 4, 5, 6 and 7), even when these frequencies were separated by 3 octaves. This could be useful in a system in which responses to a face should occur irrespective of whether the face was blurred or was presented as edges, as in a line drawing. It may be noted here that it is a property of human perception that objects including faces can be recognised even when an object is blurred, and even when only a line drawing of an object containing only edge information is presented.

The tuning of the neurons to the different low-pass and high-pass filtered faces was different for the different neurons, as shown in Fig. 5 and by the cutoff frequencies shown in Fig. 6, so that an explanation

for the tuning in terms of a single function such as the overall modulation transfer function of the visual system is not sufficient to account for the results. Indeed, the fact that the neurons responded differently to the different frequencies is further evidence that these neurons could act as filters useful in conveying information useful in distinguishing between faces (Baylis *et al.*, 1985; Leonard *et al.*, 1985). The finding that the response of the neurons was not always a smooth function of frequency, but could for example decrease as higher frequencies were included in the low-pass filtered images of faces or as low frequencies were included in the high-pass filtered images of faces (e.g. neuron R1731) indicates that information in certain frequency bands was able to inhibit these neurons. (These inhibitory effects were statistically significant in many cases, as shown by the Newman–Keuls' analyses which followed the analyses of variance.) This was particularly likely to occur for the non-optimal face stimulus for a given neuron, indicating that the selectivity of these neurons to different faces was a combination of the excitation produced by some aspects of faces and inhibition produced by others. Such inhibition was particularly likely to occur for high spatial frequencies, implying that these frequencies might be particularly important in distinguishing between individual faces. Sometimes some inhibition at some frequencies was apparent in the responses of a neuron to the optimal face in the set. This is not surprising in terms of the present hypothesis, in that it is not likely that the face most effective in activating a neuron will happen to have been included in the set of 5 faces for which series of spatial frequency filtered images were available.

Harmon and Julesz (1973; Harmon, 1973) have shown that in man low spatial frequencies, below 10 cycles/face, are sufficient for the recognition of faces of different individuals. Fiorentini, Maffei and Sandini (1984) have extended this work, and shown that the spatial frequencies above 8 cycles/face that are present in a face can also provide the basis for face recognition. It is of interest that the great majority of the cells described here had low-pass cutoff frequencies below 8 cycles/face, that is they responded well to faces containing only frequencies below 8 cycles/face (see Fig. 6). Similarly, the great majority of the cells described here had high-pass cutoff frequencies above 8 cycles/face (and even above 16 cycles/face), that is they responded well to faces containing only frequencies above 8 cycles/face (see Fig. 6). Thus almost all of the cells described here responded well to both the low frequencies shown by Harmon and Julesz (1973) to be sufficient for face recognition, and to the high spatial frequencies shown by Fiorentini *et al.* (1984) to be sufficient for face recognition. The responsiveness of these neurons to both low and to high spatial frequencies present in faces could thus provide a basis for these psychophysical findings.

One series of investigations which would provide useful further evidence on this neural system would be to determine to what extent these neurons can distinguish between different faces when these are low-pass or high-pass filtered with cutoff frequencies of 8 cycles face. Some evidence on this is shown in Fig. 10, from which it appears that even when frequencies below 8 cycles/face are included in the stimulus, the neurons still respond differently to the different faces. Although more experiments on this are needed, this preliminary evidence suggests a very close parallel between the ability of the human observer to distinguish between the faces of different individuals when only for example low spatial frequencies are present, and a corresponding property of the neurons described here. A further interesting series of related investigations would be to compare the responses of these neurons, and of human face recognition, to bandpass filtered images of faces (e.g. 2–4, and 4–8 cycles/face), to obtain further evidence on the function of different parts of the spatial frequency spectrum in face recognition and its neural basis.

The relatively wide range of frequencies within which information can activate these neurons is of interest in relation to the hypotheses that there are tuned spatial frequency channels at earlier stages of the visual system. The width of these channels has been estimated as 1.5 octaves (Campbell, 1983; Kulikowski *et al.*, 1982), and simple cells in area 17 of the cat have estimated mean bandwidths at half maximal amplitude of 1.0 ± 0.2 (Maffei and Fiorentini, 1973), 1.3 ± 0.3 (Andrews and Pollen, 1979) or 1.4 octaves (Kulikowski and Bishop, 1981). Values consistent with these have been found in the monkey (Schiller *et al.*, 1976; DeValois *et al.*, 1982). Insofar as the neurons described here can respond to information from a wider frequency range than this, if there are spatial frequency channels earlier in the visual system, then information from these channels would need to be recombined (in a particular way) in order to account for the width of tuning of the neurons described here. It is also worth noting here that although the neurons described here are tuned to accept information in a certain range of spatial frequencies, this does not mean that any power in this frequency band will activate these neurons. Indeed, the contrary is the case, in that these neurons were activated by only some of the face stimuli (see the values for d' and the generalization index in Figs 4 and 5 of Baylis *et al.*, 1985), and not by non-face stimuli such as gratings or complex stimuli, so that to the extent that their response properties depend on the spatial frequency of the stimulus, an additional constraint is that the phase information within this frequency band must also be appropriate.

It was found that these neurons were more likely to respond well to isolated parts of a face if they had a high cutoff frequency for high-pass filtered stimuli (see Fig. 10). This is understandable, in that if a

neuron can respond to only a small range of high spatial frequencies present in a face, then it is consistent that it can also respond to a part of that face, in which of course the high frequencies are present, but the very lowest spatial frequencies are absent. The observation that the high-pass cutoff frequency, the low-pass cutoff frequency, and the octave separation, did not correlate with the extent to which the neuron responded differently to the different faces, as reflected by the values for d', breadth of tuning (H) and the generalization indices, could be because in the majority of cases the spatial frequency measurements were for the optimal face stimulus. It will be of interest to determine whether there is any relation between these spatial frequency measures taken on the non-optimal face stimulus, and the selectivity of the neurons for different faces.

Acknowledgement—This work was supported by the Medical Research Council. C.M.L. held a NIH senior postdoctoral fellowship.

REFERENCES

Andrews B. W. and Pollen D. A. (1979) Relationship between spatial frequency selectivity and receptive field profile of simple cells. *J. Physiol.* **287**, 163–176.

Baylis G. C. and Rolls E. T. (1983) Responses of neurons in the inferior temporal visual cortex in long and short term memory tasks. *Soc. Neurosci. Abstr.* **9**, 21.

Baylis G. C., Rolls E. T. and Leonard C. M. (1985) Selectivity between faces in the responses of a population of neurons in the cortex in the superior temporal sulcus of the monkey. *Brain Res.* In press.

Benton A. (1980) The neuropsychology of facial recognition. *Am. Psychol.* **35**, 176–186.

Bruce C., Desimone R. and Gross C. G. (1981) Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *J. Neurophysiol.* **46**, 369–384.

Bruning J. L. and Kintz B. L. (1977) *Computational Handbook of Statistics* 2nd edn. Scott, Foresman, Glenview, IL.

Campbell F. W. (1983) Why do we measure contrast sensitivity? *Behav. Brain Res.* **10**, 87–97.

Damasio A. R., Damasio H. and Van Hoesen G. W. (1982) Prosopagnosia: anatomic basis and behavioral mechanisms. *Neurology* **32**, 331–341.

Desimone R., Albright T. D., Gross C. G. and Bruce C. (1984) Stimulus-selective responses of inferior temporal neurons in the macaque. *J. Neurosci.* **4**, 2051–2062.

DeValois R. L., Albrecht D. G. and Thorell L. G. (1982) Spatial frequency selectivity of cells in macaque visual cortex. *Vision Res.* **22**, 545–559.

Fiorentini A., Maffei L. and Sandini G. (1984) Role of high spatial frequencies in face perception. *Perception* **12**, 195–201.

Gross C. G., Rocha-Miranda C. E. and Bender D. B. (1972) Visual properties of neurons in inferotemporal cortex of the macaque. *J. Neurophysiol.* **35**, 96–111.

Harmon L. D. (1973) The recognition of faces. *Scient. Am.* **229**, 70–82.

Harmon L. D. and Julesz B. (1973) Masking in visual recognition: effects of two-dimensional filtered noise. *Science* **180**, 1194–1197.

Kling A. and Steklis H. D. (1976) A neural substrate for affiliative behavior in nonhuman primates. *Brain Behav. Evol.* **13**, 216–238.

Kluver H. and Bucy P. C. (1939) Preliminary analysis of functions of the temporal lobes in monkeys. *Archs Neurol. Psychiat.* **42**, 979–1000.

Kulikowski J. J. and Bishop P. O. (1981) Linear analysis of the responses of simple cells in the cat visual cortex. *Expl Brain Res.* **44**, 386–400.

Kulikowski J. J., Marcelja S. and Bishop P. O. (1982) Theory of spatial position and spatial frequency relations in the receptive fields of simple cells in the visual cortex. *Biol. Cybernet.* **43**, 187–198.

Leonard C. M., Rolls E. T., Wilson F. A. W. and Baylis G. C. (1985) Neurons in the amygdala of the monkey with responses selective for faces. *Behav. Brain Res.* **15**, 159–176.

Maffei L. and Fiorentini A. (1973) The visual cortex as a spatial frequency analyzer. *Vision Res.* **13**, 1255–1267.

Meadows J. C. (1974) The anatomical basis of prosopagnosia. *J. neurol. neurosurg. Psychiat.* **37**, 489–501.

Merrill E. G. and Ainsworth A. (1972) Glass-coated platinum-plated tungsten microelectrodes. *Med. Biol. Engng* **10**, 662–672.

Perrett D. I., Rolls E. T. and Caan W. (1982) Visual neurons responsive to faces in the monkey temporal cortex. *Expl Brain Res.* **47**, 329–342.

Rolls E. T. (1981a) Processing beyond the inferior temporal visual cortex related to feeding, memory, and striatal function. In *Brain Mechanisms of Sensation* (edited by Katsuki Y., Norgren R. and Sato M.). Chap. 16, pp. 241–269. Wiley, New York.

Rolls E. T. (1981b) Responses of amygdaloid neurons in the primate. In *The Amygdaloid Complex* (edited by Ben-Ari Y.), pp. 383–393. Elsevier, Amsterdam.

Rolls E. T. (1985) Connections, functions and dysfunctions of limbic structures, the prefrontal cortex, and hypothalamus. In *The Scientific Basis of Clinical Neurology* (edited by Swash M. and Kennard C.). Chap. 14, pp. 201–213. Churchill Livingstone, London.

Rolls E. T., Burton M. J. and Mora F. (1976) Hypothalamic neuronal responses associated with the sight of food. *Brain Res.* **111**, 53–66.

Rolls E. T., Sanghera M. K. and Roper-Hall A. (1979) The latency of activation of neurones in the lateral hypothalamus and substantia innominata during feeding in the monkey. *Brain Res.* **164**, 121–135.

Rolls E. T., Thorpe S. J., Boytim M., Szabo I. and Perrett D. I. (1984) Responses of striatal neurons in the behaving monkey. 3. Effects of iontophoretically applied dopamine on normal responsiveness. *Neuroscience* **12**, 1201–1212.

Saxton W. O. and Koch T. L. (1982) Interactive image processing with an off-line minicomputer: organization, performance and applications. *J. Microscopy* **127**, 69–83.

Schiller P. H., Finlay B. L. and Volman S. F. (1976) Quantitative studies of single-cell properties in monkey striate cortex. III. Spatial frequency. *J. Neurophysiol.* **39**, 1334–1351.

Schwartz E. L., Desimone R., Albright T. D. and Gross C. G. (1983) Shape recognition and inferior temporal neurons. *Proc. natn. Acad. Sci.* **80**, 5776–5778.

Seltzer B. and Pandya D. N. (1978) Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey. *Brain Res.* **149**, 1–24.

Whiteley A. M. and Warrington E. K. (1977) Prosopagnosia: a clinical, psychological, and anatomical study in three patients. *J. neurol. neurosurg. Psychiat.* **40**, 394–430.