

PART A

Biochemistry and Molecular Biology

Retinal Physiology, Cell Biology, Neurotransmitters, Morphology

Central Nervous System Physiology and Morphology

THE RESPONSES OF NEURONS IN THE CORTEX IN THE SUPERIOR TEMPORAL SULCUS OF THE MONKEY TO BAND-PASS SPATIAL FREQUENCY FILTERED FACES

EDMUND T. ROLLS, GORDON C. BAYLIS and MICHAEL E. HASSELMO
University of Oxford, Department of Experimental Psychology, South Parks Road,
Oxford OX1 3UD, England

(Received 24 January 1986; in revised form 18 August 1986)

Abstract—There are neurons in the cortex in the anterior part of the superior temporal sulcus of the macaque monkey with visual responses which would be useful for face recognition (Rolls, 1984; Baylis *et al.*, 1985). To analyze further the information which leads them to respond, their responses were measured to parametrically filtered stimuli. The responses of 48 such single neurons were measured to faces which were digitized and were bandpass spatial frequency filtered. The octave width bands were 2-4, 4-8, 8-16, 16-32, 32-64 and 64-128 cycles per image. It was found that the neurons could respond well to single octaves of the spatial frequencies normally present in faces, that the most effective bands were 4-8, 8-16 and 16-32 cycles per face (cpf), and that the bands 2-4 and 32-64 cpf were partly effective. In investigations of whether the responses of the neurons to an unfiltered face, and to low-pass and high-pass filtered images could be predicted by linear addition of their responses to each of the octave bands shown separately, it was found that the majority of the neurons were non-linear, and responded much less than predicted. It was also shown that this occurred even when the contrast was reduced to 0.25 of that normally present in a face, so that the result was not due just to a ceiling effect of the maximum firing rate. These results help to define parametrically the aspects of the information normally present in a face which are sufficient to produce responses of these neurons to them, and show that linear operations cannot account for information processing in this part of the visual system.

Face recognition Primate temporal lobe Fourier analysis Spatial frequency Inferior temporal cortex 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; Rolls, 1984; Desimone *et al.*, 1984; Baylis *et al.*, 1985). The responses of these neurons to faces (real or projected, human or rhesus monkey) are selective in that they are 2-10 times as large to faces as to gratings, simple geometrical stimuli or complex 3-D objects. The responses to faces are excitatory, sustained and are 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 (Perrett *et al.*, 1982; Rolls, 1984; Baylis *et al.*, 1985). 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; Whitely 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). 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).

In order to analyze further the information which leads these neurons in the cortex in the superior temporal sulcus to respond, their responses have been investigated to faces in which defined amounts of information had been parametrically varied by spatial frequency filtering (Rolls *et al.*, 1985). Lowpass and highpass spatial frequency filtered stimuli were used. It was found that many of the neurons could respond

to blurred images of faces, with a mean half-maximum amplitude of the neuronal response to the series of lowpass filtered images of faces at 3.3 cpf. Almost all the neurons had lowpass cutoff frequencies defined in this way below 8 cpf. 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 half-maximum amplitude of the neuronal response to the series of highpass filtered images of faces was 29.7 cpf. Almost all the neurons had high-pass cutoff frequencies above 8 cpf. 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. A comparison which suggests that these neurons are related to face perception is that face recognition in man can be performed with lowpass filtered images which contain only information up to 8 cpf, or with highpass filtered images which contain only information down to 8 cpf (Fiorentini *et al.*, 1984).

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 (Rolls *et al.*, 1985). This suggests 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.

In the experiments described here, the responses of these neurons in the cortex in the superior temporal sulcus with responses selective for faces were measured to band-pass spatial frequency filtered faces, and to combinations of different pass-bands of spatial frequency. The widths of the band-passed stimuli were one octave. The experiments had two main aims. The first was to analyze in more detail than was possible with the low-pass and high-pass filtered images of faces (Rolls *et al.*, 1985) the importance of small regions of the spatial frequency spectrum normally present in faces in the responses of these neurons to faces. For example, with the low-pass filtered stimuli, it was found that the majority of the neurons responded well when frequencies up to 8 cpf were present, but as the band 4–8 cpf was

always shown in combination with the bands 2–4 and 1–2 cpf, it was not clear to what extent the responses were due to the 4–8 band, or were due to a combination of the frequencies 1–8 cpf (Rolls *et al.*, 1985). This could be investigated with octave band-pass filtered stimuli, as described here. The aim was to investigate quantitatively using parametric variation of the stimuli the information in faces which is sufficient to activate these neurons. The second aim was to investigate whether the responses to a combination of bands of the spatial frequencies present in faces could be predicted from a linear addition of the responses of the neuron to each of the bands of frequencies presented separately. (It may be noted that on this model a smaller response to a combination than to some of the components presented separately might result if one or more of the components presented alone produced inhibition.) The combinations were provided both by specially prepared combinations of some of the band-pass spatial frequency filtered images, and by low-pass and high-pass spatial frequency filtered images of faces, which are also of course combinations of some of the octave band-pass filtered images. The aim was to investigate the responses of these neurons at the contrast levels normally present in faces to test the utility of computations by the brain in the spatial frequency domain under normal viewing conditions.

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 (1 *Macaca mulatta* and 2 *Macaca fascicularis*) (weight 3.5–4.8 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 histological sections which showed the reference electrodes and micro-lesions made at the end of some of the microelectrode tracks. Alternate 50 and 25 μm sections were stained for cytoarchitecture with cresyl violet and for myeloarchitecture by the Gallyas method (Baylis *et al.*, 1986).

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 frame-store (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 foot-lamberts 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 visual stimuli, and allowed image processing tech-

niques such as spatial frequency filtering to be applied to the stimuli presented. The monitor on which the images were displayed was placed 1 m from the monkey, and subtended 12 deg at the retina.

Second, stimuli could be 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 1.0 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 any stimulus other than a square appeared (such as a grating, a 3-D object, a face, or circle), then if the monkey licked he obtained fruit juice (i.e. all stimuli except the square were treated as positive discriminative stimuli, S+). If a square of the same area and luminance as the circle, 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 sec) period in which the stimulus was on. This procedure was designed to ensure fixation of the stimuli (Rolls *et al.*, 1979). 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 was 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, 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 included octave band-pass images of a face, and another set low-pass and high-pass filtered images of a face. 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, and also blank trials on which an image set with the same gray level as the other images occurred with a probability of 0.06.

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 $256 \times 256 \times 8$ 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 frequency filtering was performed by a digital Fourier transform convolution with filters circularly symmetric in the spatial frequency domain. The octave spatial frequency pass-bands were 2–4, 4–8, 8–16, 16–32, 32–64 and 64–128 cpf. Several combinations of different pass-bands were also prepared, as shown in Fig. 1(A–E). All these images for one face comprised a set of images, of which the members were shown repeatedly in random sequence. A separate set of images contained low-pass and high-pass filtered images of the same face. The cutoff frequencies of the low-pass filters were 2, 4, 8, 16, 32 and 64 cpi (see Fig. 1). The cutoff frequencies of the high-pass filters were 4, 8, 16, 32 and 64 cpi (see Fig. 1). In the frequency domain, the profile of the low-pass and band-pass filters was flat with a gaussian taper which occurred over two adjacent frequency points (cf. Fig. 1 of Rolls *et al.*,

1985). For the band-pass and high-pass filters the d.c. 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, the frequencies represent cycles per face (cpf). No adjustment was made to the contrast of the filtered images, as we 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 128 cpf 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 cpi). The maximum contrast achieved by the display system was 0.94, corresponding to gray level values of 255 and 0. The mean gray level value of all images was 127, and the standard deviation of the pixel range in each of the unfiltered faces was 45.

Non-face stimuli. The responses of the cells were tested to a wide range of non-face stimuli, including sine wave gratings, boundary curva-

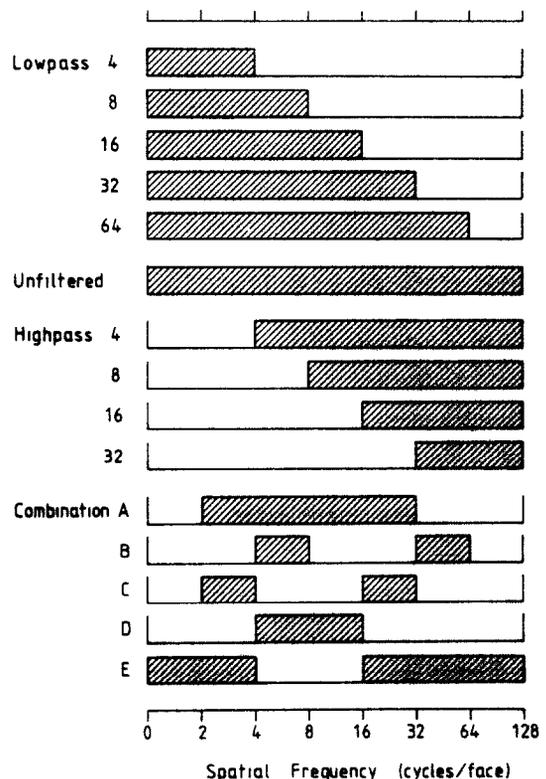


Fig. 1. The pass bands (shaded) of the low-pass and high-pass filtered images of faces, and of the combinations (A–E) of different octaves present in faces.

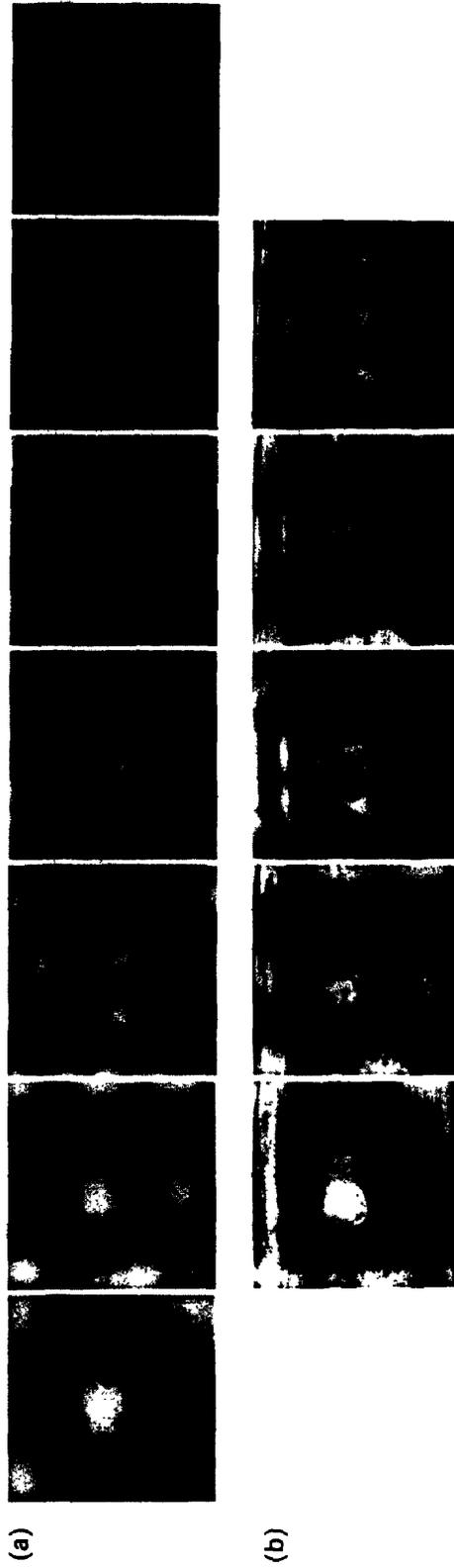


Fig. 2. Examples of the stimuli. (a) The seven single octave bandpassed images of one face. 0-2 cycles per face is the leftmost stimulus, and 64-128 cycles per face is the rightmost image. (b) Combinations A-E (see text) of octaves for the same face.

ture descriptors, complex 2-D images, and three-dimensional junk objects, as described previously (Baylis *et al.*, 1985).

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, and significantly different ($P < 0.05$). (In fact, for half these neurons, their response 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, as shown by Baylis *et al.*, 1985. 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 Rolls, 1984 and by Baylis *et al.*, 1985.) If the neuron satisfied the criterion, then the octave band-pass series for 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 sets of images were shown, including a set of low-pass and high-pass images of the same face, and a set of octave band-pass filtered images of a non-optimal face.

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 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 to different band-pass images of faces and combinations of these, and to 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.

RESULTS

It was possible to measure the responses to the octave band-pass spatial frequency filtered images of faces of 48 neurons. Examples of the responses of one neuron in rastergram and peristimulus time histogram form to the octave band-pass filtered images of a face are shown in Fig. 3.

The mean responses of the population of neurons to the octave band-pass stimuli of optimal faces, and for comparison to non-optimal faces, are shown in Fig. 4. It is clear that large responses are obtained to some of the individual octaves present in a face, with the octaves 4–8, 8–16 and 16–32 being particularly effective, and 2–4 and 32–64 somewhat effective. [It may be noted that the median response of the neurons to the most effective octave of the optimal face stimulus was 30 spikes/sec (interquartile range 23–36), with corresponding values for the non-optimal face of 19 (12–28). The ratio of the response to the most effective octave in the non-optimal face stimulus to the most effective octave in the optimal face stimulus had a median of 0.71 (0.58–0.85).]

The responses of 3 neurons to combinations of band-pass filtered images, to low-pass and high-pass filtered images of the same face, and to the unfiltered face are shown in Fig. 5. A comparison of the response to the unfiltered face with the responses to individual octaves shows that the latter can in some cases approach the magnitude of the response to the unfiltered face. To show whether the responses to the unfiltered face, to the combinations of octaves, and to the low-pass and high-pass images can be predicted by the response to a linear addition of

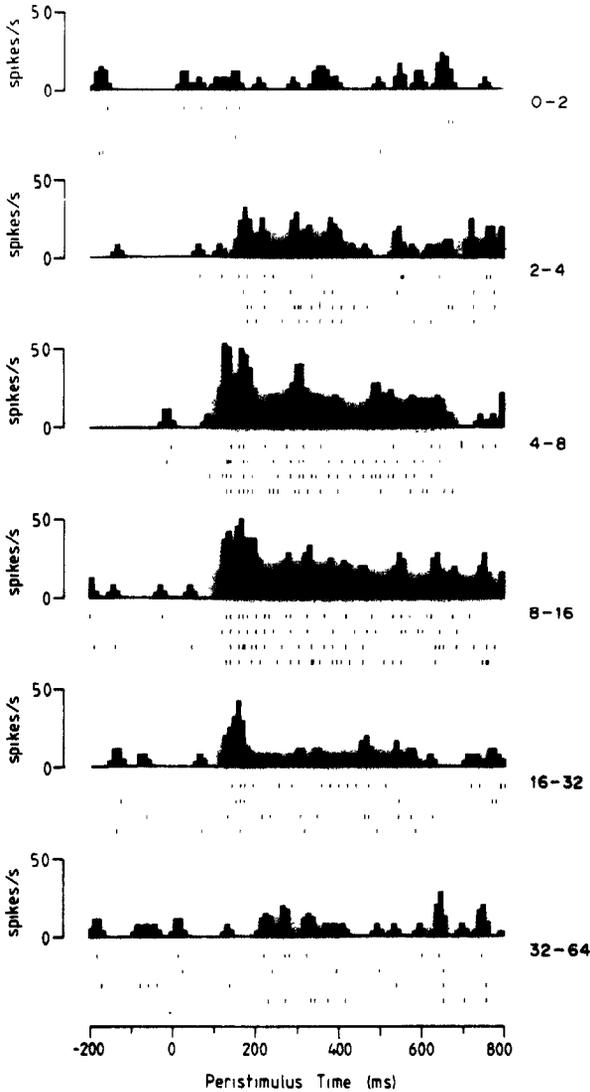


Fig. 3. Examples of the responses of one of the neurons to some of the stimuli used. Four peristimulus time rastergrams, and peristimulus time histograms based on 6 presentations of the stimulus, are shown for each stimulus. The onset of the visual stimulus was at time 0. The bin width was 10 msec.

the responses to individual octaves, the latter is also shown for the 3 neurons in Fig. 5 (open symbols). For two of these neurons (GG332 and DD089), it is clear that a linear addition of the neuronal responses to each octave separately produces a value which is greater than the actual neuronal response obtained. For one of the neurons (GG439) illustrated in Fig. 5, the neuronal response could be predicted approximately from an addition of the responses to each octave presented separately.

The results for the population of neurons tested are shown in Fig. 6(a), as the mean

response obtained to the different low-pass and high-pass images, and to the combinations of octaves (normalized for each neuron to 100% as the response to the unfiltered face), and the mean response predicted from the responses to the octaves presented separately. It is clear that for the population as a whole, a much greater response was predicted than was obtained. The responses shown in Fig. 6(a) are with the optimal face stimulus. For comparison, the responses obtained when a non-optimal face stimulus was used are shown in Fig. 6(b). It is shown that the actual response obtained from the neurons was even less well predicted by a linear addition of the responses to single octaves, particularly near the center of the curve (i.e. with an unfiltered or little filtered face). This implies that with a non-optimal face stimulus,

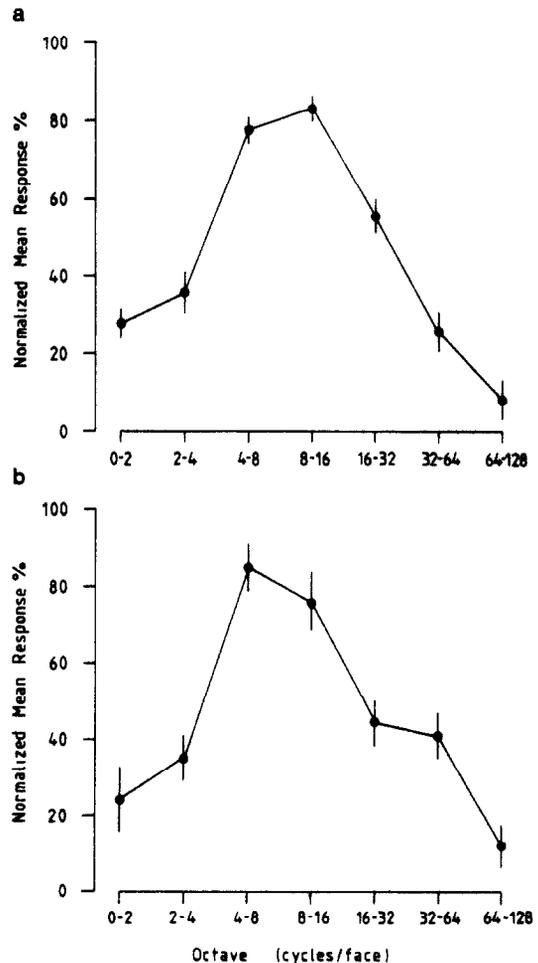


Fig. 4. The mean of the responses of the population of neurons to the different octaves in the optimal face (a) and in a non-optimal face (b). The response of each neuron was first normalized relative to the response to the best single octave stimulus for that neuron.

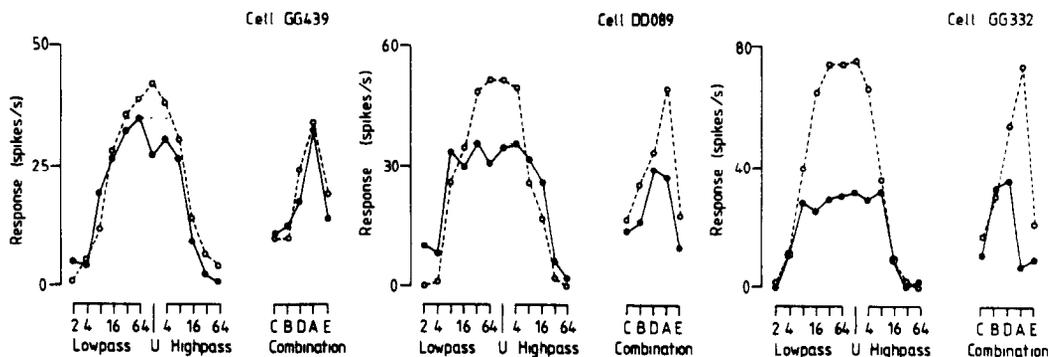


Fig. 5. The responses (solid symbols) of three cells to the octave stimuli, to the combinations of octave stimuli, and to the low-pass and high-pass filtered images of the same face. Also shown (open symbols) are the responses predicted from a linear addition of the responses to each of the octave stimuli presented separately.

the response to the whole face results from processes different from addition, and it is suggested that in fact inhibition by part of the information present in a whole (unfiltered) face accounts for the lower response obtained to a non-optimal than to an optimal face stimulus.

To provide an indication of the degree of non-linearity of each of the neurons, the deviation from non-linearity, estimated from the ratio of the mean response obtained with the unfiltered image to that predicted from linear addition of the responses to each octave is shown in Fig. 7(a) for the population of 38 neurons for which this information is available. (To provide estimates with little variance for

each neuron, this ratio for each neuron was calculated from the actual and predicted responses not only to the unfiltered face, but also to the LP64 and HP4 images, which are almost unfiltered.) It is clear that for the majority of the neurons, non-linearity occurred, in that the response to the unfiltered face and to the two most similar images from the LP and HP series was much less than that which was predicted by a linear summation of the response to each of the octaves. The results for Fig. 7(a) are for neuronal responses to the optimal face stimulus. For comparison, the results with a non-optimal face stimulus are shown in Fig. 7(b). It is clear that even lower levels of linearity were obtained,

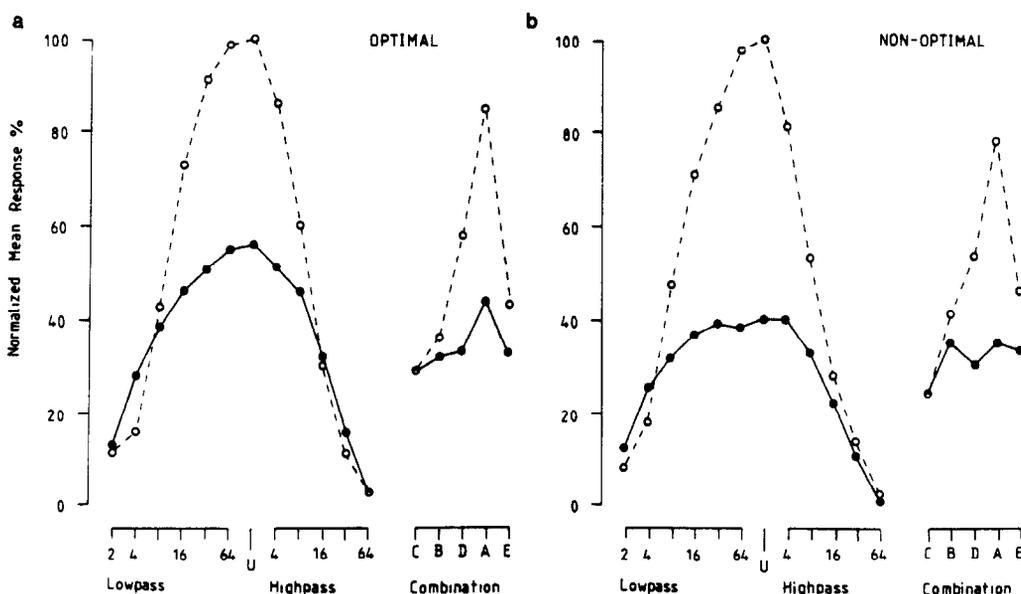


Fig. 6.(a) The mean response obtained, and the mean response predicted, to the different images for the population of 38 neurons analyzed with optimal face stimuli, and (b) of 19 neurons analyzed with non-optimal face stimuli.

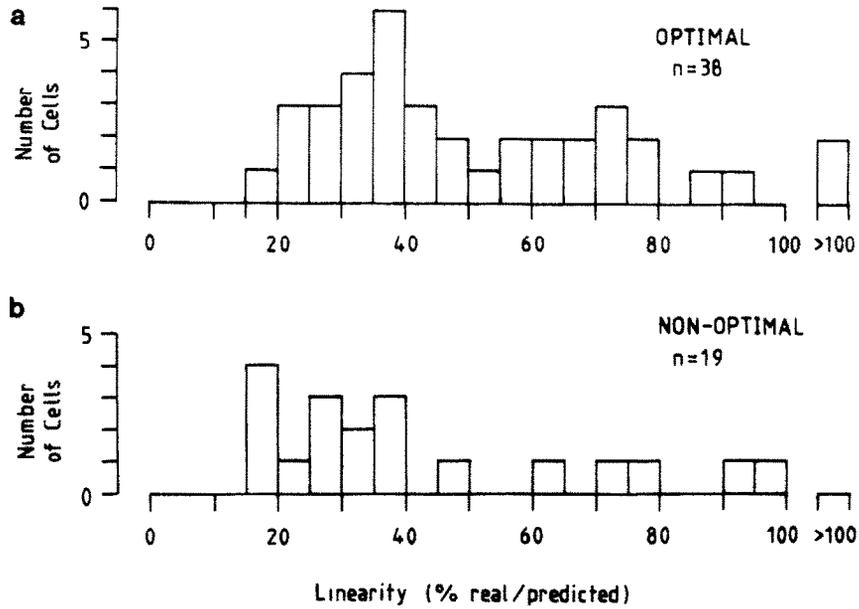


Fig. 7. The proportion of the responses obtained to an unfiltered face to that predicted from linear addition of the responses to each of the octaves presented separately, labelled as linearity, for the population of neurons analyzed. The results are plotted separately for responses of the neurons to optimal (a) and non-optimal (b) face stimuli.

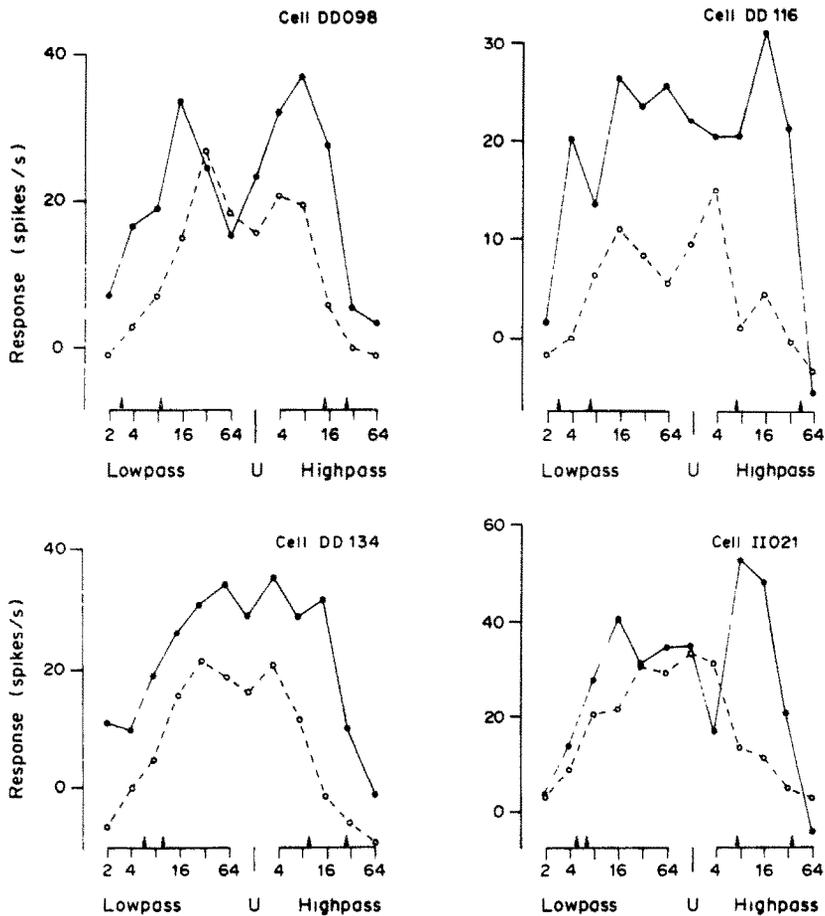


Fig. 8. The responses of four cells to high-pass and low-pass filtered images of faces at normal (solid symbols) and at low (one quarter, open symbols) contrast.

consistent with the hypothesis that more than just linear summation, and in fact inhibition as suggested above, accounted for the lower responses of the neurons to a non-optimal than to an optimal face.

These experiments indicated that at normal contrast, the responses of the neurons to a face could not in most cases be predicted from the responses of the neurons to each octave contained in the face. However, it was possible that saturation had occurred, with a ceiling of firing rate produced as soon as two or three octaves were combined, resulting in non-linearity. In order to determine whether the responses of the neurons might behave more linearly at low contrast, the above experiments were repeated with all stimuli reduced to one quarter (or in a few cases to one eighth or one half) of the normal contrast.

Examples of the responses of the neurons to the reduced contrast low-pass and high-pass filtered images of a face, and of the unfiltered face are shown in Fig. 8. It is clear that decreasing the contrast did result in less saturation, in that there is a smaller range of stimuli for which the firing rate has reached a plateau. The result of this is that the frequencies at which the half-maximal firing rates occurred (called the cutoff frequencies here) shifted towards the centre of Fig. 8. That is, the low-pass cutoff frequency became higher, and the low-pass cutoff frequency became lower. This effect was found for the majority of the neurons tested, as shown by the comparison of the cutoff frequencies at normal and at low contrast in Fig. 9.

It was possible to measure the responses of 10 neurons to both the reduced contrast octave images and to the reduced contrast images of combinations of band-pass filtered images, of low-pass and high-pass filtered images of the same face, and of the unfiltered face. The responses of one neuron (DD141) to the reduced contrast images of combinations of band-pass filtered images, of low-pass and high-pass filtered images of the same face, and of the unfiltered face are shown in Fig. 10. Also shown are the values of firing rate predicted from the neuronal response to each octave presented separately. It is clear that even at reduced contrast, it is not possible to predict the responses to combinations of octaves from the response to each octave alone, although the neuronal response approaches linearity for a greater range than at high contrast. The nor-

malized mean response found for the population of neurons analyzed and the predicted response is shown in Fig. 11. To provide an indication of the degree of non-linearity of all 11 of the neurons at reduced contrast, the deviation from non-linearity, estimated from the ratio of the mean response obtained with the unfiltered image to that predicted from linear addition of the responses to each octave is shown in Fig. 12. (To provide estimates with little variance for each neuron, neuronal responses obtained to LP64 and HP4, and the predicted values, were used to give further estimates of the non-linearity, and the values from these data points are included in the estimate for the deviation from non-linearity of each neuron shown in Fig. 12. This is the same procedure used for Fig. 7.) It is clear that for the majority of the neurons even at reduced contrast, non-linearity occurred, in that the response to the unfiltered face and to the two most similar images from the LP and HP series was much less than was predicted

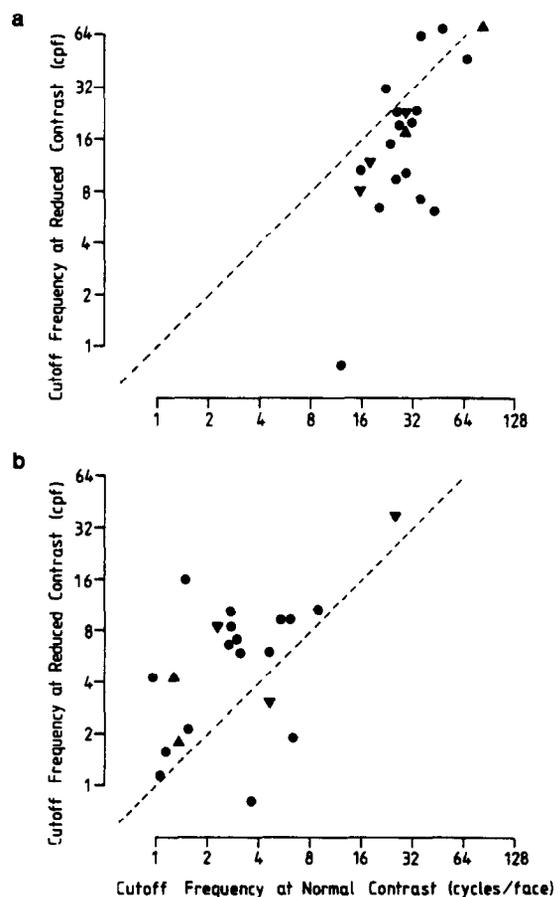


Fig. 9. Low-pass (a) and high-pass (b) cutoff frequencies at normal and at reduced contrast. Each point represents data for one neuron. Inverted triangles: half contrast; circles: quarter contrast; triangles: one eighth contrast.

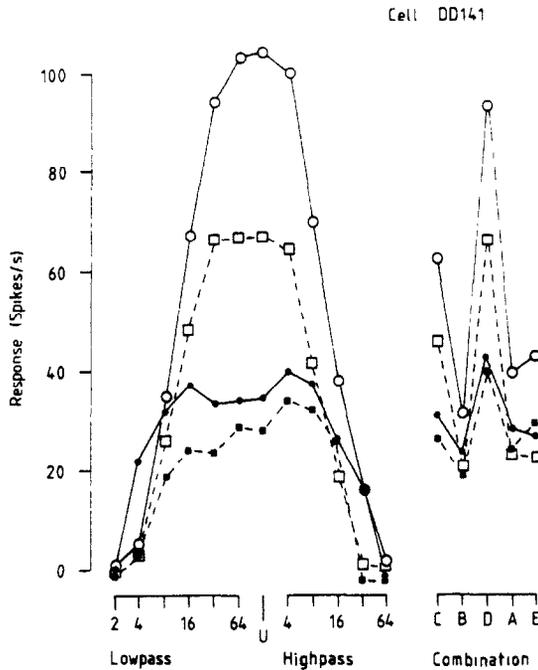


Fig. 10. Example for one neuron of the responses to low contrast as well as to normal contrast images. The actual response obtained is shown by the closed symbols, and the predicted response by open symbols. Squares—low contrast. Circles—high contrast.

by a linear summation of the response to each of the octaves.

It has been observed previously that for non-optimal face stimuli, there is in some cases inhibition of the neuronal response as more frequencies are added into a low-pass series of face images (Rolls *et al.*, 1985). A corresponding finding occurs when a high-pass series includes more low spatial frequency information (Rolls *et al.*, 1985). To investigate to what extent this effect might be predictable from the responses to the octave stimuli, the experiments described above were performed for some neurons with stimuli which were non-optimal for the neurons. An example of the results obtained from one neuron are shown in Fig. 13. To show whether the responses of neurons to a low-pass and high-pass series of faces might differ if responses to optimal and to non-optimal face stimuli were measured, and whether these could be related to the responses to octaves presented alone, the actual and predicted responses for optimal and non-optimal stimuli for the set of neurons tested in this way have been shown in Fig. 6(b), and the linearity of the set in Fig. 7(b). To analyze further whether some non-linear process makes it more difficult to predict the neuronal re-

sponses to a non-optimal than to an optimal face, the correlation between the predicted and the actual responses (calculated across the population of neurons tested) to the optimal and non-optimal face stimuli are shown in Fig. 14. It is clear that in the central region of the figure (unfiltered and adjacent little filtered stimuli) there is a smaller correlation between the actual and predicted response for the non-optimal than for the optimal face stimulus. Thus non-linear processes appear to contribute especially to the lack of response produced in these neurons by non-optimal face stimuli.

The recording sites of the neurons from which this subpopulation was drawn are shown elsewhere (Baylis *et al.*, 1987), and were in most cases in region TPO, TEa or TEM (see Seltzer and Pandya, 1978).

DISCUSSION

First, it can be seen from Fig. 3 that neurons can respond vigorously to images containing a very small proportion of the spatial frequency information normally present in a face. The most effective octave in many cases led to a response as great as that to the unfiltered face. The mean response of the population of neurons to the octaves 4–8, and 8–16 cpf, was more than

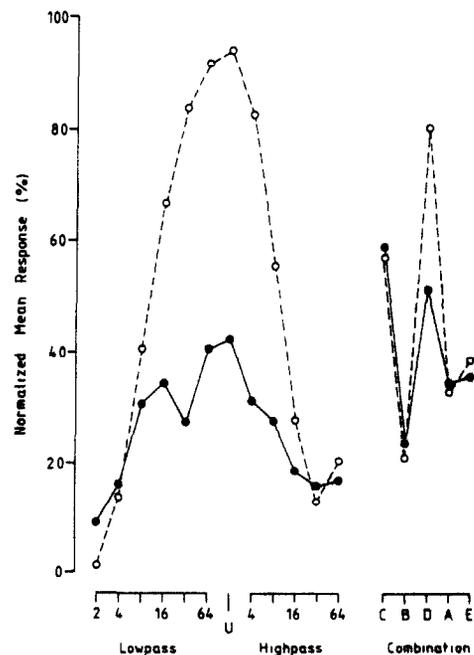


Fig. 11. The mean response obtained (solid circles), and the mean response predicted (open circles), to the different images presented at low contrast for the population of 10 neurons analyzed.

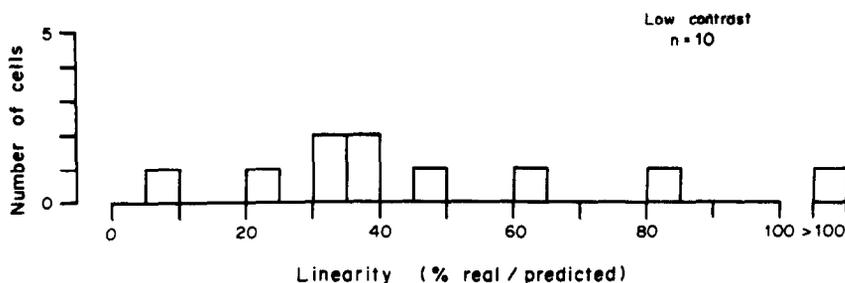


Fig. 12. The proportion of the response obtained to an unfiltered face to that predicted from linear addition of the responses to each of the octaves presented separately, labelled as linearity, for the population of 10 neurons analyzed at low contrast.

half that to the unfiltered face. Inspection of Fig. 1 shows that these neurons had half maximal responses to very degraded stimuli. For each cell the spatial frequency tuning was determined from the responses to individual octaves as follows. The response of the neuron to each octave was plotted at the geometric mean frequency on a logarithmic scale (in the same way as is shown for the whole population in Fig. 4). By linear interpolation on this scale the point at which the response dropped to half that to the most effective octave could be calculated. There are clearly two such points—representing low and highpass cutoff frequencies. The median and interquartile range of these frequencies for the optimal face stimulus was as follows. For optimal face stimuli the lowpass was 3.9 cpf (2.1–5.4), and the highpass was 24.5 cpf (18.3–37.1). Although derived somewhat differently, these estimates of the spatial frequency tuning of neurons with responses selective for faces agree closely with those determined by Rolls *et al.* (1985).

Second, the spatial frequency tuning for non-optimal face stimuli was determined in the same way. The mean responses for the population are shown in Fig. 4. A similar pattern of response is seen to that to the optimal stimuli, and this is reflected in the values of highpass and lowpass cutoff frequencies. The lowpass had a median of 3.8 cpf (1.3–4.6), and the highpass 26.4 cpf (16.4–65.1). There was no significant difference between the distribution of such values for the optimal and non-optimal stimuli (as shown by a Kolmogorov–Smirnov test). Thus the tuning of cells to *individual octaves* of the spatial frequency spectrum does not appear to depend on whether the stimulus was optimal for the neuron.

Third, it is clear that the responses of most of these neurons to unfiltered faces or to combina-

tions of spatial frequency bands cannot be predicted by linear addition of their responses to octaves of the spatial frequency information present in faces (see Figs 5–13). This is not just because the firing rate reaches a limiting high value, as shown by the observation that non-linearity occurred even with the low contrast images (see Figs 8–12), and at values of firing rate which were below those attained with the high contrast images (see e.g. Fig. 10). Nevertheless, at low contrasts, the responses of the neurons were somewhat more linear than at high contrasts (cf. Figs 7 and 12). However, this result emphasises the point that linearity, for these neurons at least, is a property which is

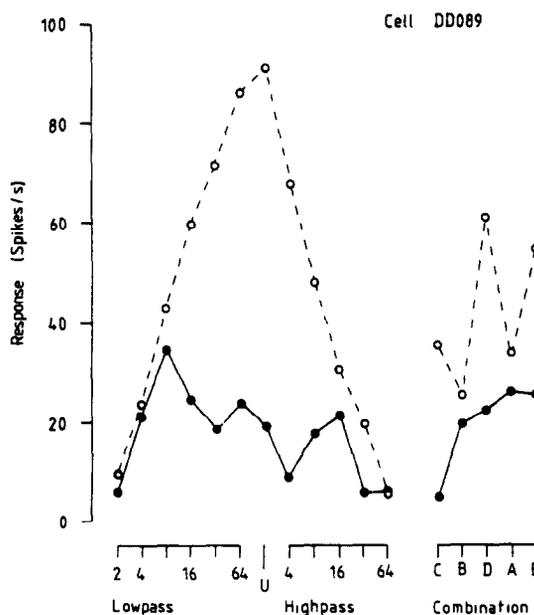


Fig. 13. The responses (solid symbols) of one cell to the combinations of octave stimuli, and to the low-pass and high-pass filtered images of the same face, which was non-optimal for the cell. Also shown (open symbols) are the responses predicted from a linear addition of the responses to each of the octave stimuli presented separately.

approached only at low contrast. That is, under normal viewing conditions even of a video image, which is by no means the highest contrast or brightest image normally viewed by the visual system, there is clear evidence for non-linearity, and we believe that a linear approach to visual information processing in the spatial frequency domain could not work at this level of the visual system. The extent to which linearity is a property which is found most clearly in earlier parts of the visual system (Campbell, 1983; Andrews and Pollen, 1979; Kulikowski and Bishop, 1981; Kulikowski *et al.*, 1982; Schiller *et al.*, 1976; DeValois *et al.*, 1982) only when contrast is very low (that is at contrasts which are below those at which the visual system normally functions) deserves consideration. Further, for the neurons at the level of the visual system described here, we note that even at low contrast, the responses of the neurons do not reflect linear operations in the spatial frequency domain (see Figs 8–12).

It is shown in Fig. 14 that the correlation between the response predicted from a linear combination of the responses to individual octaves and the response actually obtained from the neuron was poorer for the non-optimal than for the optimal face stimuli. This difference was particularly prominent for the stimuli with little filtering, that is those near the middle part of Fig. 14. This is consistent with the hypothesis that inhibition produced by information in some parts of the spatial frequency spectrum of

a non-optimal face contributes to the small response of that neuron to an unfiltered version of the non-optimal face, and that this inhibition can be detected once a neuron is responding, but is not evident relative to the spontaneous activity of the neuron when that octave is presented alone. Indeed, in a number of cases a single octave presented alone led to an increase of firing rate of the neuron, but when added to another stimulus (such as a lowpass, or highpass image), led to a significant reduction (as tested by Neuman-Keuls analysis) in the firing rate. For example, in neuron DD236 the octave 8–16 cpf produced a response of 16.0 spikes per sec, but the effect of adding this to a lowpass 8 image (thus making it a lowpass 16) was to reduce the firing rate from 48.6 to 26.4 spikes per sec, suggesting that *in this combination* the octave 8–16 cpf was inhibitory. A total of 7 neurons (out of a total of 38, or 18%) showed this effect for optimal faces, and 11/19 (58%) for non-optimal faces. A chisquare test showed that such inhibition was more likely to occur for non-optimal than for optimal faces ($\chi^2 = 7.4$, d.f. = 1, $P < 0.01$). The octaves at which such reversal was found for non-optimal faces were often those with high frequencies, as shown in Table 1. These results provide a further indication that these neurons are not linear with respect to individual octaves, and that there is less linearity when the face stimulus is not optimal for the cell.

In conclusion, in quantitative investigations it has been shown that these neurons can respond to simplified images of faces containing as little as one octave of the spatial frequency information normally present in a face. It has also been shown that the responses of these neurons can not be predicted by linear addition of their responses to each of the octave bands shown separately, and that this non-linearity was particularly pronounced for responses to non-optimal faces. This indicates that at least part of the differential responsiveness of these neurons

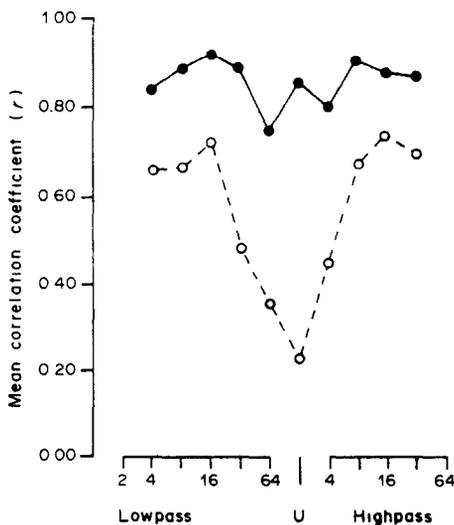


Fig. 14. The correlation between the response obtained and that predicted with optimal (solid symbols) and non-optimal (open symbols) face stimuli.

Table 1 Percentage (number) of cells showing reversal

Octave	Optimal face	Non-optimal face
0–2	0	0
2–4	5.1 (2)	5.3 (1)
4–8	7.7 (3)	15.7 (3)
8–16	7.7 (3)	10.5 (2)
16–32	0	10.5 (2)
32–64	0	21.0 (4)
64–128	0	10.5 (2)

to different faces (Baylis *et al.*, 1985) arises because of inhibition produced by information present in non-optimal faces. It also shows that linear operations are not used at this stage of the visual system, although of course it does not exclude the possibility that at earlier stages of the visual system there is linear spatial frequency processing, with its possible advantages for size and translation invariance.

Acknowledgement—This work was supported by the Medical Research Council.

REFERENCES

- Andrews B. W. and Pollen D. A. (1979) Relationship between spatial frequency selectivity and receptive field profile of simple cells. *J. Physiol., Lond.* **287**, 163–176.
- Baylis G. C. and Rolls E. T. (1987) Responses of neurons in the inferior temporal visual cortex in short term and serial recognition memory tasks. *Expl. Brain Res.* In press.
- 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.* **342**, 91–102.
- Baylis G. C., Rolls E. T. and Leonard C. M. (1987) Functional subdivisions of temporal lobe neocortex. *J. Neurosci.* In press.
- Benton A. (1980) The neuropsychology of facial recognition. *Am. Psychol.* **35**, 176–186.
- Bruce C., Desimone R. and Gross C. G. Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *J. Neurophysiol.* **46**, 369–384 (1981).
- Bruning J. L. and Kintz B. L. (1977) *Computational Handbook of Statistics*, 2nd edn. Scott, Foresman, Glenview, Ill.
- 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, N.Y.* **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.
- Klüver H. and Bucy P. C. (1939) Preliminary analysis of functions of the temporal lobes in monkeys. *Archs Neurol. Psychiatr.* **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. Psychiatr.* **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. (1984) 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.). Churchill Livingstone, London.
- Rolls E. T. (1984) Neurons in the cortex of the temporal lobe and in the amygdala of the monkey with responses selective for faces. *Human Neurobiol.* **3**, 209–222.
- 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.
- Rolls E. T., Baylis G. C. and Leonard C. M. (1985) Role of low and high spatial frequencies in the face-selective responses of neurons in the cortex in the superior temporal sulcus. *Vision Res.* **25**, 1021–1035.
- 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.