Exp Brain Res (1994) 101:473–484

ORIGINAL PAPER

Edmund T. Rolls · Martin J. Tovee · Dean G. Purcell
Alan L. Stewart · Paul Azzopardi

The responses of neurons in the temporal cortex of primates, and face identification and detection

Received: 6 July 1993 / Accepted: 23 June 1994

Abstract The ability of a human observer to detect the presence of a briefly flashed picture of a face can depend on the picture's spatial configuration, that is on whether its features are rearranged (jumbled) or are in their normal configuration. The *face-detection effect* (FDE) is found under conditions of backward masking, when the presence of a face can be detected with shorter masking intervals when it is in the normal than when in the rearranged configuration. A similar effect is found when the subject is asked to classify the face as rearranged or not – the *face-classification effect* (FCE). Part of the interest of the FDE and the FCE is that they show how the configuration of a stimulus can be an important factor in the perceptual processing which leads to detection and classification of the stimulus. To analyse these effects we recorded from single neurons in the cortex in the superior temporal sulcus of macaques when they were shown (in a visual fixation task) normal and rearranged faces under backward masking conditions shown in experiments 2 and 3 to produce, with the same apparatus, the FCE, and also to produce comparable effects on the identification of which face was present (called hereafter the face-identification effect), and also of the clarity of the face. We found in experiment 1 that there are some face-selective neurons which respond to faces only, or better, when the features in the faces are in their normal configuration rather than rearranged. We also showed in this experiment that the difference in the response to the normal as compared to the rearranged faces became greater when the masking stimulus was delayed more. Thus, at intermediate delays, there are more neurons active for the normal than for the rearranged face. We therefore propose that the FDE, the FCE, and the *face-identification effect* arise because the total number of neurons activated by faces in their normal configuration is greater than that activated by rearranged faces, because of the sensitivity of some face-selective neurons to the spatial arrangement of the features. The experiments also show that backward visual masking does produce abrupt termination of the firing of neurons in the temporal cortical visual system, so that the duration of a neuronal response is very short when visual stimuli can just be perceived.

Key words Face recognition · Masking
Inferior temporal cortex
Spatial configuration of features · Monkey · Human

E. T. Rolls (✉) · M. J. Tovee[1] · D. G. Purcell[2] · A. L. Stewart[3]
P. Azzopardi
University of Oxford, Department of Experimental Psychology,
South Parks Road, Oxford OX1 3UD, UK;
Fax no: +44-865-310447

Present addresses:
[1] Department of Psychology,
University of Newcastle Upon Tyne, Ridley Building,
Newcastle Upon Tyne NE1 7RU, UK
[2] Department of Psychology, Oakland University, Rochester,
MI 48309–4401, USA
[3] Department of Management and Engineering Management,
Stevens Institute of Technology, Castle Point, Hoboken,
NJ 07030, USA

Introduction

The ability of a human observer to detect the presence of a briefly flashed picture of a face can depend on the picture's spatial configuration. Configuration refers here to whether the features of the face are rearranged (jumbled) as compared to their normal configuration with respect to each other, or to whether the face is in its normally viewed upright configuration or is inverted. In particular, under conditions of visual backward masking a picture of an upright face can be detected better than a picture of an inverted face, and an inverted face can be detected better than a rearranged face (Purcell and Stewart 1986, 1988). This phenomenon is called the face-detection effect (FDE). A phenomenon which is related to the FDE is the face-classification effect (FCE), where the classification threshold for normal faces is shorter than the classification threshold for rearranged faces. Part of the interest of the FDE and the FCE is that

they show how the configuration of the features in a stimulus can be an important factor in the perceptual processing which leads to the ability to detect and classify the stimulus. Part of their interest is also that they show what perceptual processing can be performed when the processing time available is limited, by the use of a masking visual stimulus. In the experiments described here we investigated the neural bases of these configuration-sensitive perceptual effects and of the perception that can occur when stimuli are shown for only short periods, by analysing the responses of single neurons which respond selectively to faces in the temporal lobe visual cortical areas of macaques. This is one of a series of investigations being performed in order to advance our understanding of the neural bases of perception and of the disorders produced by brain dysfunction (Rolls 1991, 1992, 1994a, b; Tovee and Cohen-Tovee 1993).

The experimental paradigms in which these perceptual phenomena have been investigated use visual backward masking and are described next. Visual masking uses a briefly presented stimulus (10–20 ms) as a visual target. A masking visual stimulus follows the onset of the target stimulus. The second (masking) stimulus often consists of a complex pattern of black lines on a white background. The interval from the onset of the target to the onset of the masking pattern is called the stimulus onset asychrony (SOA). The dependent variable is taken as the SOA threshold time interval for performance of the appropriate task (detection or classification). When comparing the temporal detection threshold of normal faces with either inverted or rearranged versions of that face the detection threshold time is shortest for the normal face, next for the inverted face, and longest for the rearranged version of the face. The masking stimulus limits the time in which the target is available for processing by the visual system. With one exception (Gorea and Julesz 1990), all demonstrations of the FDE have been under conditions of visual backward masking. The neurophysiological experiments described here were performed using a masking stimulus like those used in the original FDE experiments (Purcell and Stewart 1986, 1988).

The classification threshold for a normal face is longer than the detection threshold for a normal face. The same relationship holds for rearranged faces. This is consistent with work which demonstrates that classification latencies are shorter for normal faces than for rearranged faces (Valentine and Bruce 1986; Valentine 1991). Surprisingly, it takes longer to detect the presence of a face with rearranged features than it does to correctly classify a normal face. It takes still longer to classify a rearranged face (Purcell and Stewart 1988). The SOA for classifying a normal face is 10 ms longer than for its detection. For rearranged faces an observer needs an additional 60 ms to classify the face as rearranged (Purcell and Stewart 1988). The classification advantage for normal faces has been called the FCE. Thus, while normal faces have an initial advantage in terms of detectability, the magnitude of this advantage for the classification task is directly correlated with the dissimilarity of the stimulus from the optimum normal face configuration. The classification effects considered in this paper all refer to decisions about the configuration of a face (normal vs rearranged).

The investigations described here were designed to analyse the neural bases of face detection and identification by combining the psychophysics of face perception with the techniques of neurophysiology, using the same stimuli and similar procedures with both human beings and monkeys. The neurons recorded from were those in macaques with responses selective for faces (see Bruce et al. 1981; Desimone et al. 1984; Rolls 1984, 1992; Tovee et al. 1993), some of which only respond to faces when the face features are in the correct configuration and are not rearranged (Perrett et al. 1982). [We note that monkeys can discriminate between jumbled and normal faces (Perrett et al. 1988), and that the temporal cortex face-responsive neurons in monkeys are sensitive to a number of different parameters which specify the spatial configuration of the features in a face (Yamane et al. 1988; Young and Yamane 1992)]. While recordings were made from these neurons, the stimuli were presented for short durations and short SOA values of the masking stimulus, comparable with those used in establishing the FDE (Purcell and Stewart 1988). The neurophysiological studies described in experiment 1 were designed to investigate first how the face-selective neurons respond to the short stimuli (e.g. 16 ms) used in the psychophysical experiments, and especially how large the responses are to such short stimulus presentations, and whether the neuronal responses continue for longer than the duration of the stimuli. Second, the neurophysiological experiments were designed to show how masking stimuli affect the responses of these neurons to the 16-ms presentation of a face stimulus, and especially how both the magnitude and duration of the response of the face-selective neurons was affected by the non-face pattern mask. It was possible, for example, that a non-face mask might not terminate abruptly the response of a face-selective neuron, but simply attenuate the magnitude of the response. In all the experiments described here the test and masking stimuli had the same intensity and covered the same part of the retina (cf. Humphreys and Bruce 1989). The only previous single neuron analysis of the effects of backward visual masking we know was an investigation of cat lateral geniculate neurons in which stimuli of unequal intensity and in some cases area were used (Schiller 1968), so that the results are not directly comparable with the findings described here. Third, the neurophysiological experiments were designed to show how the configuration of the features of stimuli in a face influences the responses of the face–selective neurons under these conditions of short stimulus duration and backward masking. In analysing the responses of these neurons under conditions in which face detection and face classification can just occur, we may well be investigating the earliest stages at which stimuli

are treated as faces. In experiments 2 and 3 we describe psychophysical studies in human beings performed with the same apparatus used for the neurophysiological experiments, to allow direct comparison between the neurophysiological and psychophysical findings. It was also necessary to perform Experiments 2 and 3 in order to extend the previous psychophysical observations to the conditions of stimulus size and duration used in the neurophysiological experiments. Another aim of experiments 2 and 3 was to investigate other perceptual phenomena than the FDE and FCE, which might occur with similar stimulus presentation conditions. In particular, because some populations of the temporal lobe face-selective neurons respond differently to the faces of different individuals (Baylis et al. 1985; Hasselmo et al. 1989a, b), even in some cases relatively independently of perceptual transforms (see Hasselmo et al. 1989b; Rolls 1992), we wished to investigate face identification by humans under these conditions, to determine whether this could also be related to the responses of the neurons recorded.

Materials and methods

Experiment 1 – neurophysiological investigations

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 two alert macaque monkeys (*Macaca mulatta*, weight 3.0 kg) seated in a primate chair using techniques that have been described previously (Rolls et al. 1976, 1990; Tovee et al. 1993). 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 Micro VaxII 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. Eye position was measured to an accuracy of 0.5° with the search coil technique, and fixation of the visual stimuli was ensured by use of a blink version of a visual fixation task, in which the fixation spot was blinked off 100 ms before the test stimulus appeared. The timing of the task is described below. The stimuli were static visual stimuli subtending 8° in the visual field presented on a video monitor at a distance of 1.0 m. The fixation spot position was at the centre of the screen. The monitor was viewed binocularly, with the whole screen visible to both eyes.

Task timing

This is shown diagrammatically in Fig. 1. Each trial started at −500 ms (with respect to the onset of the test image) with a 500-ms warning tone to allow fixation of the fixation point, which appeared at the same time. At −100 ms the fixation spot was blinked off so that there was no stimulus on the screen in the 100-ms period immediately preceding the test image. The screen in this period, and at all other times including the inter-stimulus interval (ISI) and the interval between the test image and the mask, was set at the mean luminance of the test images and the mask, so that pattern masking with equally intense test and mask stimuli was investigated (see Bruce and Green 1989). At 0 ms, the tone was switched off and the test image was switched on for 16 ms. This 16-ms period was the frame interval of the video framestore with which the images were presented. The image was drawn on the monitor from the top to the bottom in the first 16 ms of the frame period by the framestore, with the remaining 4 ms of the frame period being the vertical blank interval. The monitor had a persistence of less than 3 ms, so that no part of the test image was present at the start of the next frame. (The PAL video system was in use. When we refer in this paper to a stimulus duration of 16 or 20 ms we refer to the display system just described. The display system used in a tachistoscope does not use a rater display method, and so the stimulus duration periods will not be directly comparable.) SOA values of 20, 40, 60, 100 or 1000 ms (chosen in a random sequence by the computer) were used (see Fig. 1). The duration of the masking stimulus was 300 ms. At the termination of the masking stimulus the fixation spot reappeared, and then after a random interval in the range 150–3350 ms it dimmed, to indicate that lick responses to a tube in front of the mouth would result in the delivery of juice reward. The dimming period was 100 ms, and after this the fixation spot was switched off, and reward availability terminated 500 ms later.

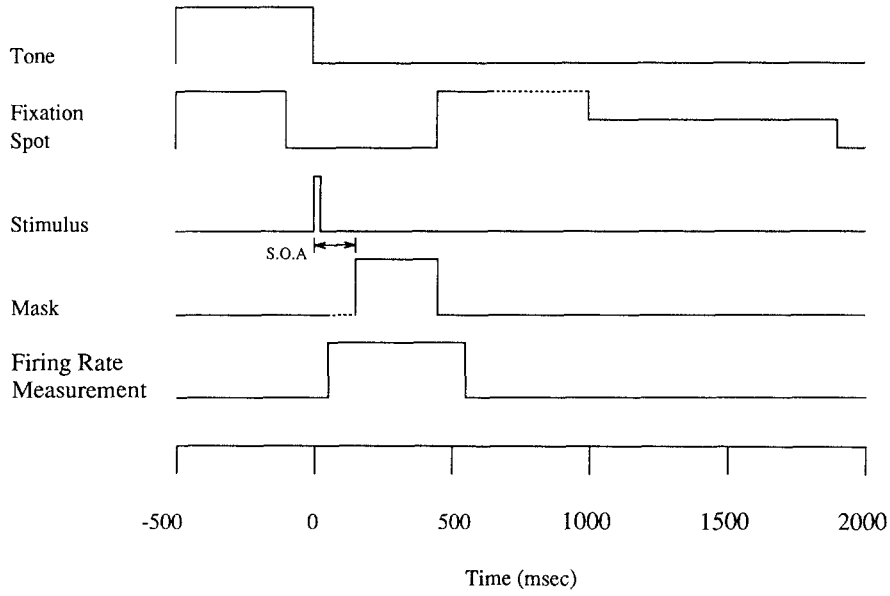


Fig. 1 The timing used in the backward-masking, visual-fixation blink task. (*S.O.A.* stimulus onset asynchrony)

The monkey was required to fixate the fixation spot in that if he licked at any time other than when the spot was dimmed by a small amount, saline instead of fruit juice was delivered from the tube; also, if the eyes moved by more than 0.5° from zero time until the start of the dimming period, then the trial was aborted. (When a trial aborted, a high-frequency tone sounded for 0.5 s, the stimuli were removed from the screen, no reinforcement was available for that trial, and the inter-trial interval was lengthened from 8 to 11 s.)

Stimuli

Faces were used as test stimuli. The faces were pictured looking into the video camera, and included both macaque and human faces. The rearranged version of each face was made by interchanging features, as illustrated in Fig. 2. Both the face and the rearranged face were presented for a duration of 16 ms. The masking stimulus was made up of letters of the alphabet (N,O; see Fig. 2). The masking pattern consisted of overlapping letters, and this masking pattern was used because it is similar to the mask used in the previous psychophysical experiments on the FDE.

The stimuli were stored in digital form on a computer disk and displayed on a monochrome video monitor using a video framestore (Advanced Electronic Design 512). The resolution of these images was 256 wide by 256 high with 256 grey levels. At all times the mean luminance of whatever was displayed on the monitor (even when it was a blank screen) was set to grey level 128, so that there were no luminance changes during each trial. The monitor provided maximum and minimum luminances of 6.0 and 0.13 foot-lamberts, respectively, and was adjusted internally and by use of a look-up table 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.

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 five 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 neurone responded differently to face and to non-face stimuli. 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 four to ten times in order to provide sufficient data for an analysis of variance (ANOVA), so as to determine whether the neurone responded differently to the different stimuli within the set.

Procedure

As tracks were made into the cortex in the superior temporal sulcus, the responses of each neurone 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 three-dimensional (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 response to the most effective face was more than 5 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) and Rolls (1992). The non-face stimuli from which the optimal was chosen included sine wave gratings, boundary curvature descriptors, complex two-dimensional (2-D) stimuli, and complex 3-D junk objects, as described above.) If the neuron satisfied the criterion, then it was tested with a version of one of the effective face stimuli for that neuron in which the features had been rearranged in the way illustrated in Fig. 2. In the masking experiment, there were ten stimulus conditions (the five SOA values of 20, 40, 60, 100 and 1000 ms for each of two images, the normal and rearranged face). These conditions were presented in a random sequence by the computer, and then the computer presented a new random sequence of the set until the set had been

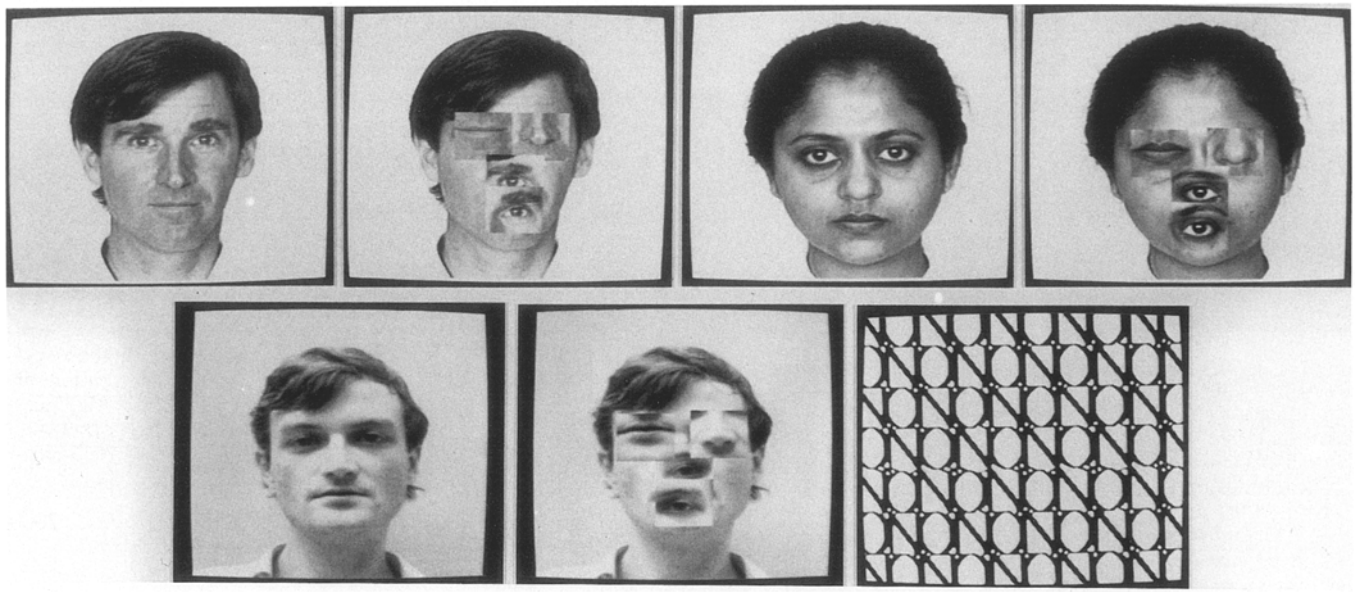


Fig. 2 Examples of the test images used in the normal spatial configuration and with rearranged features. The mask is also shown

presented eight to ten times. All the neurophysiological data were saved to disk.

Data analysis

The default period for which the firing rate was calculated was a 200-ms period starting 50 ms after the onset of the test stimulus. (The period was chosen to start at 50 ms because none of the neurons started to respond with a latency shorter than this, and all started to respond strongly from some time in the post-stimulus period between 50–100 ms.) The mean firing rate over the 8–10 trials together with its standard error was calculated by the computer for each of the stimulus conditions for graphical presentation. In addition, a two-way ANOVA was performed over the same set of data, with one factor image configuration (normal or rearranged) and the other factor SOA. We were particularly interested in the interaction term from this ANOVA, to show whether the response of the neuron at different SOAs depended on whether the face was in the normal or rearranged configuration. This was to allow comparison with the human psychophysics, in which it was known that at some SOAs only the normal, and not the rearranged face, can be detected; *t*-tests using the appropriate error term from the ANOVAs were then used if particular pairs of neuronal responses were to be compared.

Recording sites

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 then reconstructed from the X-ray co-ordinates taken together with serial 50-μm histological sections which showed the reference electrodes and micro-lesions made at the end of some of the microelectrode tracks (Feigenbaum and Rolls 1991).

Experiment 2 – magnitude estimation

Having shown how the responses of these face-selective neurons can vary over a range of SOAs, we next wished to investigate, using the identical apparatus, stimuli, and time parameters, how human face perception varies as a function of the SOA. Given that the stimuli were just above detection threshold in the neurophysiological experiments (i.e. human observers could detect the stimuli, given the 16-ms face stimulus duration and its contrast, which were a little greater than in experiments in which the FDE was investigated; see Purcell and Stewart 1986, 1988), we investigated in experiment 2 not human detection threshold, but instead how the clarity of the face stimuli (as shown by magnitude estimation) varies with the SOA. (The clarity of the face was one perceptual effect which did vary with the stimulus parameters used when the SOA was increased.) Further, the face stimuli in the neurophysiological experiment subtended larger angles (8°) than in the previous psychophysical experiments (Purcell and Stewart 1986, 1988), and it was thus important to establish the psychophysical findings with the viewing conditions used for the neurophysiological experiments.

Stimuli and observers

The stimuli were images of two different faces, and the features in these faces were normal or rearranged, so that there were four images. The stimuli were stored in digital form on a computer disk and displayed on a Microvitec video monitor using an AED 512 video framestore. Luminance was measured with a photometer. The monitor provided maximum and minimum luminance of 6 and 0.13 foot-lamberts and was adjusted internally for linearity, within an error of no more than 3%. The two faces used as stimuli are shown on the left of Fig. 2. The faces were pictured looking straight into the video camera; their expressions were neutral. The rearranged version of each face was made by interchanging the mouth and right eye, and the nose and left eye. Both the face and the rearranged face were presented for a duration of 16 ms (as described under the neurophysiological methods). The observers viewed the faces from a distance of approximately 1 m. The faces subtended approximately 8° of visual angle.

The masking stimulus was made up of letters of the alphabet (N,O; shown in Fig. 2). The masking pattern consisted of overlapping letters and was similar to the mask used in the previous psychophysical experiments on the FDE. The mask duration was constant at 300 ms. Four of the authors (all except M.J.T.) served as observers. The independent variable was the delay between the onset of the picture of a face (or rearranged face) and the onset of the masking stimulus (the SOA). Five SOA values were used (20, 40, 60, 180, 1020 ms).

A magnitude estimation task was run. The dependent variable was the number representing the clarity of the face under conditions of visual backward masking. The face, when presented without being followed by a masking stimulus, was assigned the number 10. If the observer was not able to see the features of the face, that was to be considered 0. The observers assigned a number from 0 to 10 to represent the perceived clarity of the face. One series of training trials was run, consisting of 20 randomly selected presentations. Then 200 trials were run, representing 10 presentations for each condition (image × SOA). The conditions were randomized in blocks of the 20 combinations of image and SOA. The experiment lasted approximately 1 h. A warning tone came on with a fixation spot. The fixation spot indicated the location on the monitor where the face would appear, which was always located centrally on the screen. The tone remained on for 500 ms and the spot for 400 ms. One hundred milliseconds after the offset of the fixation spot the stimulus came on, followed, after an appropriate delay, by the masking stimulus. The observers then wrote down their rating of the face's clarity and prepared for the next stimulus presentation. The inter-trial interval was approximately 13 s. The experiment was conducted in a lighted room.

Experiment 3 – psychophysical investigation with judgements as to configuration and identification

In experiment 3 judgements of face type (normal, rearranged) and of face identity were made. Part of the reason for this experiment was that it forced the observers to choose a response, that is, which face from the set had been seen and whether the configuration was normal or rearranged. Another reason was that we knew that the face-selective neurons reflect the identity of the face seen (Baylis et al. 1985; Rolls 1992) and might also be configuration-sensitive, so that it was reasonable to expect that face identification might also be particularly difficult with rearranged faces at short SOAs. Although classification of type of face (normal versus rearranged) has been investigated in previous experiments under similar conditions, determination of the identity of a face has not been previously investigated.

Stimuli and observers

Stimuli were presented using the same apparatus as in experiment 1. Five different faces were used as stimuli (four men and one woman). All the faces were well known to each of the eight observers used in the experiment. The stimuli were either normal or rearranged faces, comparable to those of experiment 1. The faces were chosen so that each person's hair appeared to be about the same grey level on the monitor and was approximately symmetrical with respect to the face.

Procedure

The observers made two judgements. They first determined whether the face was a normal or a rearranged face; they then identified whose face they thought had been presented. Both judgements were forced responses. Even if the observers were unsure of their judgement they were told to respond with their best guess. All of the subjects were familiar with psychophysical procedures and volunteered to take part in the experiment. Two of the authors ran as observers (D.G.P., A.L.S.); all other observers were unaware of the purpose of the experiment. Three hundred stimulus presentations (trials) were made, randomized in blocks of 50 presentations. Each face (and its rearranged counterpart) was shown 6 times at each of the SOA values; this gave a total of 30 replications per condition if the individual images are treated as being only normal or rearranged pictures of faces. The mean intertrial interval was 13 s.

Observers were given as much practice as they wanted. Two observers were given two practice sessions of 20 presentations each; two observers had no practice session. All other observers had one practice session. Each observer wrote down whether the stimulus appeared to be normal or reorganized, using the letter N or R; they then wrote down the identity of the face using the first letter of the person's first name. The experimental session and training lasted approximately 1 h.

Results

Experiment 1

The responses of a large sample of cells (approximately 400) in the temporal cortical visual areas were recorded. The responses of 40 of these cells with face-selective responses according to the criteria described in the Materials and methods section were investigated with the normal and rearranged face stimuli in the experiments described here. (Face-selective cells comprise approximately 20% of the cells with visual responses in some temporal lobe cortical areas – see Baylis et al. 1987). Of these 40 cells, 15 responded approximately equally to faces with features in the normal and in the rearranged configuration. These cells were not investigated further in this study. No cells in this population responded more to the faces in the rearranged than in the normal configuration. The responses of the remaining 25 cells that had greater responses in the 500-ms firing rate measurement period to the normal as compared to the rearranged configuration faces were investigated in the masking experiments described here.

Peristimulus time histograms to show the time course of the masking effects found are shown in Fig. 3 for one cell. The solid lines show the response to the normal face, and the dashed lines to the face with rearranged features. The top pair of traces show the response of the cell with no masking stimulus present in the time period shown. (With an SOA of 1000 ms, the mask did not appear until after the end of the period shown). The response of the cell to the normal face was greater than that to the rearranged face. Of particular interest was the fact that there was a reasonable response, with a peak firing rate of more than 100 spikes/s, to the single frame (16 ms) presentation of the (normal)

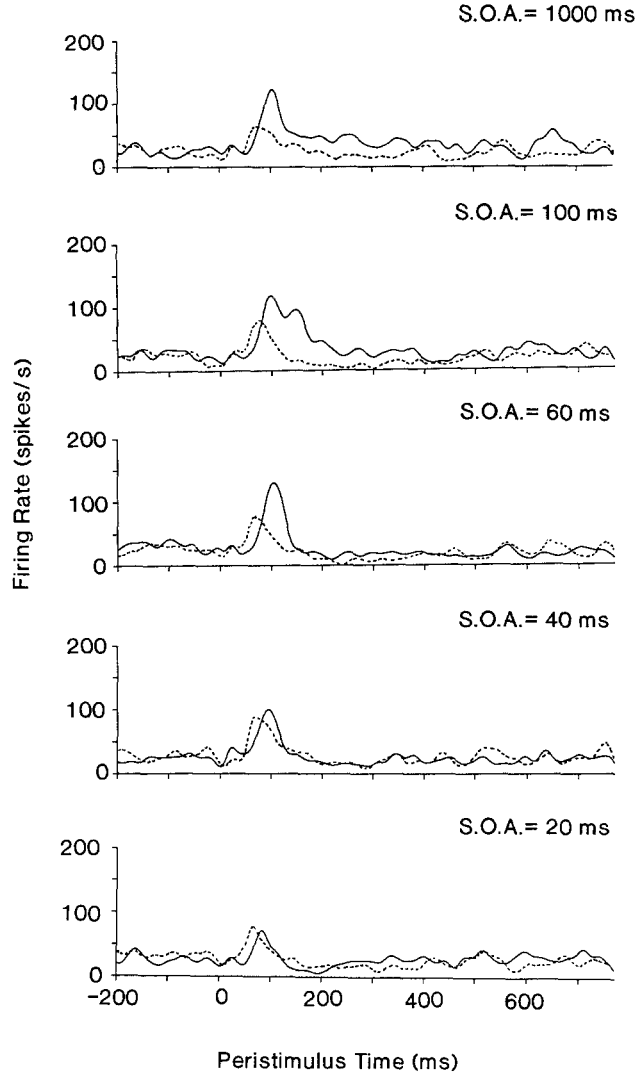


Fig. 3 Peristimulus time histograms to show the time course of the masking effects on the responses of a single neuron (AM126). The *solid lines* show the response to the normal face, and the *dashed lines* to the face with rearranged features. The test face stimulus appeared at time zero, and its presentation took 16 ms. The *top pair of traces* show the response of the cell to the test stimulus with no masking stimulus present in the time period shown. (With a stimulus onset asynchrony, *S.O.A.*, of 1 000 ms, the mask did not appear until after the end of the period shown.) With an S.O.A. of 20 ms, the 300-ms mask stimulus started 20 ms after the start of the test face stimulus. The peristimulus time histograms are based on eight to ten trials with each condition (run in random sequence), and the histograms are Gaussian smoothed with a width σ of 10 ms

face. Also of especial interest is the demonstration that the cell continued to fire at above its spontaneous rate for 300–400 ms after the end of the 16-ms stimulus, provided that no masking stimulus was present. The lower trace pairs show that the masking stimulus very effectively cut off the response of the neuron, with a latency that was only a very little greater than the onset latency of the neuronal response to the face. The result was that, with an SOA of 100 ms, the neuron fired for a little more than 100 ms, with an SOA of 60 ms the neuron fired for a little more than 60 ms, etc. (In fact, these durations of

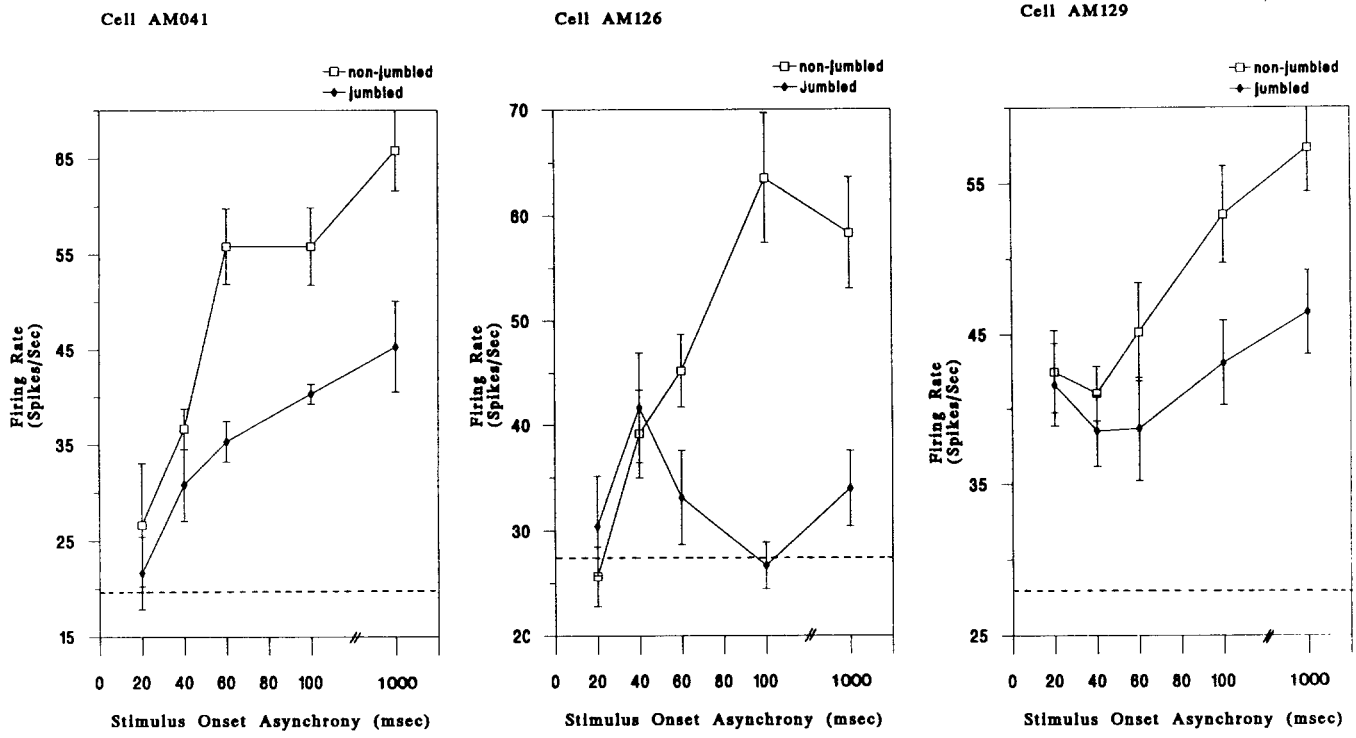
firing are if anything a little exaggerated in Fig. 3, because of the Gaussian smoothing with a width σ of 10 ms that was applied to the peristimulus time histogram.) The significant effect with respect to the FDE is that at short SOAs (e.g. 20 ms), the neuron responded hardly differently to the normal and rearranged face, and that as the SOA was increased through 60 ms the neuron gradually came to respond more to the normal than to the rearranged face. The effects shown in Fig. 3 for a single cell were typical of the effects for the population of cells described here (see further Rolls and Tovee 1994).

The data obtained from three of these neurons are shown in firing rate histogram form in Fig. 4, and described next. The mean firing rate and its standard error under each of the stimulus conditions are shown. The firing rates are measured in the 200-ms period starting 50 ms after the onset of the test face stimulus, because a typical latency for the start of firing in these cells is a little longer than this, and because humans could not use more than 200 ms of the firing to determine their behavioural response, given typical reaction times (Thorpe and Imbert 1989; and observations in experiments 2 and 3). (Nor would monitoring the neuronal firing for longer provide more information about the stimulus for SOAs of 100 ms or less, for, as shown in Fig. 3, the neurons did not respond for as long as 200 ms with such SOAs. It may also be noted that the exact duration of the period over which the firing rate was measured did not qualitatively alter any of the findings described here, as indicated by the peristimulus histogram shown in Fig. 3.)

Cell AM041 responded to the normal (non-jumbled) face when the mask was delayed by 1000 ms (effectively a no-mask condition) with a mean rate in the 200-ms period of approximately 66 spike/s (see Fig. 4). There was typically a large transient response with a latency of 60–90 ms and a duration of typically 40–60 ms followed by a period of firing in the next 200–500 ms at a rate that was typically above the spontaneous firing rate. As the delay between the mask and the test stimulus was decreased to nearly 0 ms (SOA of 20 ms), the neuronal response to the test stimulus decreased to approximately 26 spikes/s, which was close to the spontaneous rate of 19 spikes/s. For comparison, there was a smaller response to the rearranged (jumbled) face even with the SOA of 1000 ms, and this response became less at shorter SOAs until, at 20 and 40 ms SOAs, there was no significant difference between the response to the rearranged and to the normal face. The fact that the effect of SOA was dependent on which stimulus was shown was confirmed by the significant interaction term in the two-way ANOVA ($F_{(4,49)}=2.53$, $P<0.05$). Also consistent with this, the firing rates to the non-jumbled and jumbled faces became significantly different at SOAs of 60 ms and above (as indicated also by the fact that the standard error bars in Fig. 4 for cell AM041 do not overlap at these SOAs).

Cells AM126 and AM129, also shown in Fig. 4, showed similar effects, in that at short SOAs there was

Fig. 4 The neuronal responses of three different cells (AM041, AM126, AM129) to the test face stimulus in the normal spatial configuration (*non-jumbled*) and in the rearranged configuration (*jumbled*) as a function of the stimulus onset asynchrony. The mean and the standard error of the mean firing rate measured in a 200-ms period starting 100 ms after stimulus onset is shown for each condition, and the *dashed horizontal line* shows the spontaneous firing rate



no significant difference between the response to jumbled and non-jumbled faces, whereas there was a significant difference at longer SOAs, and in that there were significant interaction terms (respectively $F_{4,200}=8.9$, $P<0.001$ and $F_{4,324}=2.8$, $P<0.03$). (The SOA at which the difference between the responses to non-jumbled and jumbled faces became significantly different depends on a number of factors, including the extent to which the cell responds differently to jumbled and non-jumbled faces at long SOAs, and the length of the period over which the response of the single neuron was measured. For example, for cell AM041 the cell responded significantly differently at an SOA of 40 ms to the normal and rearranged face if the response of the cell was taken over a 500-ms period starting 50 ms after the onset of the stimulus.)

It was possible to perform the masking experiment, which implied finding a configuration-sensitive, face-selective cell that had a reasonable response to a single frame (16 ms) presentation of the test image, and running eight to ten presentations of the stimulus set, for 25 neurons in two monkeys. Qualitatively similar and statistically significant effects to those already illustrated were found in 9 of the 25 cells fully tested (as shown for example by a significant interaction term in the two-way ANOVAs for all 9 cells). To document this, in Fig. 5 we show the mean response of these 9 neurons to jumbled and to non-jumbled faces as a function of SOA. It is seen that on average the neurons have greater responses to non-jumbled than to jumbled faces at all SOAs, and that this difference is larger at longer than at shorter SOAs. In one of the cells the responses to the non-jumbled face became different to those to the jumbled face not only because the response to the non-jumbled face became greater at the longer SOAs but also because the firing rate of the neuron was inhibited below its spontaneous value by the jumbled face at long SOAs. For the remaining cells, it was not possible to demonstrate similar significant effects with the particular stimulus conditions (test image, etc.) which had been chosen for the experiment. For example, for some of these cells, with a single frame (16 ms) presentation of the face stimulus, the difference in the response of the single cell to the normal and rearranged stimuli was not sufficiently great that a statistical treatment showed significance even with long SOAs. Although for these cells the responses to the normal configuration were not sufficiently greater than to rearranged with the 16-ms stimulus duration, nevertheless all 25 cells did show backward masking effects. For none of the neurons was there any effect of the mask itself, and for none of the neurons was there a greater response to the rearranged face than to the face with the normal configuration of features. The recording sites of the neurons analysed in this study are shown in Fig. 6. The neurons were in the cortex in the depths of the superior temporal sulcus (STS).

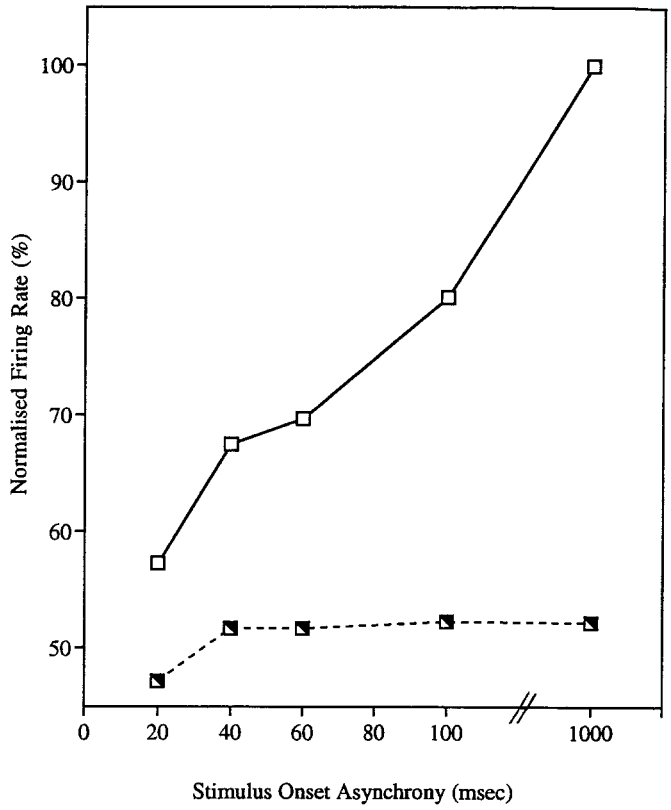


Fig. 5 The mean response of nine neurons to faces in the normal spatial configuration (non-jumbled) (*open squares*) and in the rearranged configuration (jumbled; *half-filled squares*) as a function of the stimulus onset asynchrony (SOA). The results for each cell were scaled to a maximum rate indicated at 100%. The spontaneous firing rates were not subtracted. All nine cells had statistically significant interaction terms between SOA and jumbled versus non-jumbled

Experiment 2

The mean magnitude estimates given by subjects for the clarity of the faces in normal configuration and with rearranged features are shown in Fig. 7. For simplicity only the first four SOA values are plotted. The unplotted values are for an SOA of 1020 ms and yielded mean estimates of 8.73 and 8.05 for normal and rearranged faces.

A repeated-measures ANOVA showed that the main effect of face type was not significant ($F_{1,3}=4.4$, $P=0.127$), although face type did interact with SOA ($F_{4,12}=3.81$, $P=0.032$). The statistical results are probably due to floor and ceiling effects. The normal face was judged clearer than the rearranged face at all SOA values. At very short SOAs normal faces were judged to be almost as unclear as rearranged faces. At the longest SOAs, rearranged faces were judged to be almost as clear as normal faces.

Planned comparisons showed that at SOAs of 40 and 60 ms the normal faces were judged clearer than rearranged faces ($P<0.01$). The main effect of SOA was also statistically significant ($F_{4,12}=39.52$, $P<0.001$). Thus, as the SOA increased, clarity increased monotonically.

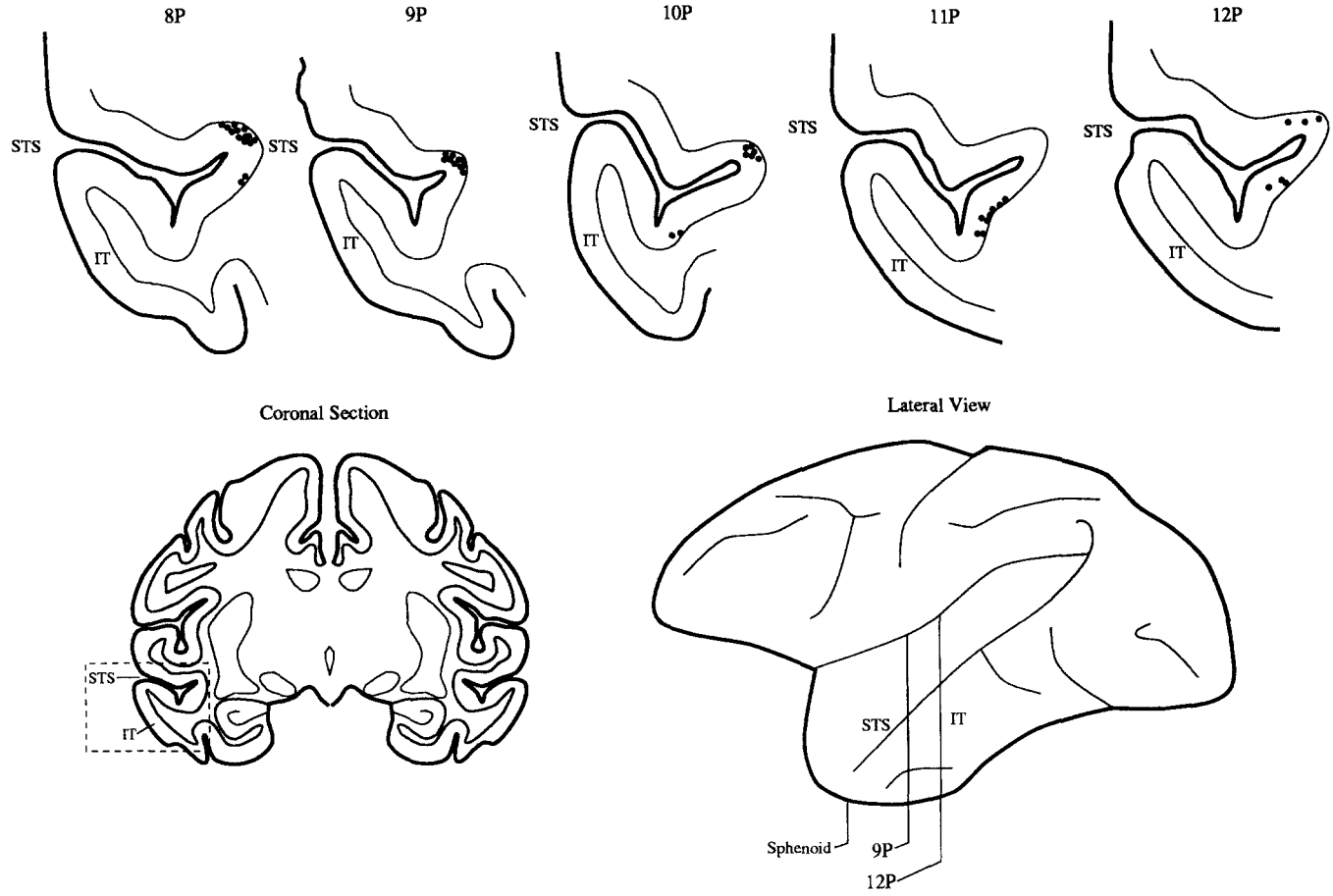


Fig. 6 The recording sites shown on coronal sections of the neurons included in this study. The positions of the coronal sections are shown on a lateral view of the macaque brain. The distances refer to millimetres posterior (*P*) to the sphenoid reference plane (see text). The neurons were in the cortex in the superior temporal sulcus (*STS*). (*IT* inferior temporal visual cortex)

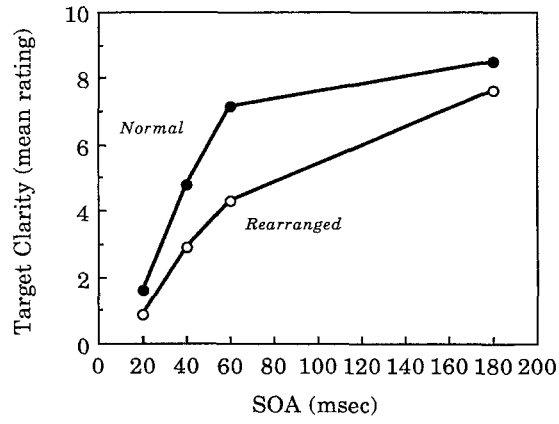


Fig. 7 Rating of the clarity of faces when in the normal or rearranged spatial configuration as a function of stimulus onset asynchrony (*SOA*). The test stimulus was presented for 16 ms. The mean of the ratings is shown

This parallels the increased firing rate we observed with many face-sensitive neurons.

Experiment 3

The data were corrected for guessing, to aid the comparison between classification and identification. (This correction arranged that chance performance would be shown as 0% correct on the graphs and perfect performance as 100% correct.) The mean proportion of correct data for the classification and identification tasks of experiment 3 are displayed in Fig. 8. For clarity data are not plotted for SOA equal to 1020 ms. At this SOA value the performance in all conditions was greater than 97% correct. The proportion correct data was submitted to an arc sin transformation (to normalise the data) and a repeated-measures ANOVA was performed. This analysis showed statistically significant effects of task ($F_{1,7}=18.901$, $P=0.0034$), ISI ($F_{4,28}=61.52$, $P<0.0001$), and configuration (normal vs rearranged; $F_{1,7}=16.718$, $P=0.0046$). Surprisingly performance on the identification task was better than on the classification task. The effect of SOA and face configuration were in the expected directions.

The interaction of task × ISI was statistically significant ($F_{4,28}=7.122$, $P=0.0004$), as was the interaction of configuration × ISI ($F_{4,28}=6.886$, $P=0.0005$). The interaction of task × configuration was not significant

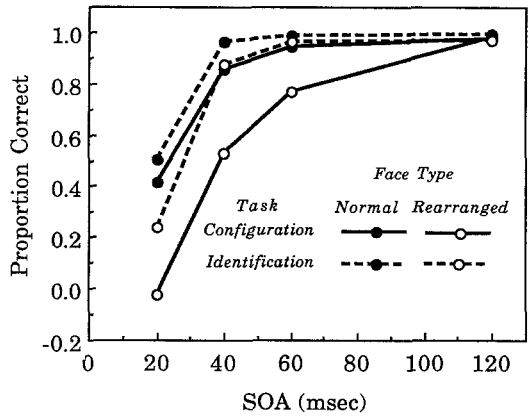


Fig. 8 Proportion correct on the tasks of classification of spatial configuration (i.e. whether the face features were normal or rearranged) and of determination of the identity of the faces for faces in the normal or rearranged spatial configuration of face features, as a function of stimulus onset asynchrony (*SOA*). The data have been corrected for guessing, to facilitate comparison between the two tasks. The means of the proportions correct are shown. The test stimulus was presented for 16 ms

($F_{1,7}=0.696$), nor was the second order interaction of task × configuration × ISI ($F_{4,28}=2.348$, $p=0.079$). Generally, the significant interactions are because of ceiling effects, with performance under all conditions converging on 100% correct as SOA increases. Face identification may have been better than face classification because of the highly trained observers we used. Each picture was of a person well known to each observer.

Discussion

Experiment 1

The results of the neurophysiological experiment show first that there is a population of neurons in the temporal cortical areas that respond to a single frame (16 ms) presentation of the test stimulus, a face. Second, the results show that the duration of the responses of these neurons to the short (16 ms) presentation of the stimulus is much longer than the stimulus, with, after an initial transient, a neuronal response that continued for several hundred milliseconds (see, e.g. Fig. 3). Third, the responses of these neurons do show backward masking. It was found that the effect of the mask, even though it was a non-face pattern mask which did not activate this population of neurons, was to cut off or interrupt the neuronal response to the face (see, e.g. Fig. 3). Fourth, for the neurons that respond to faces with the features in the normal configuration but have much smaller responses when the face features are rearranged, there is no difference in the firing rate to normal faces and rearranged faces at short SOAs, but at larger SOAs the responses to the normal faces become larger than those to the rearranged faces. Equally, at short SOAs there are no significant differences in the responses of the neurons to different faces, but at longer SOAs the neurons do have significantly different responses to different faces. This suggests that at short SOAs differentiation of the identity of the face, and whether the features are in the normal or rearranged configuration, will be impaired, but will improve as the SOA increases.

The neurophysiological results thus not only provide direct evidence that visual pattern masks produce their effects by interrupting neuronal firing, which is normally of much longer duration than a brief (e.g. 16 ms) presentation of a visual stimulus, but also provide evidence on a neuronal substrate which could underlie the face detection and related effects. Before considering the detailed explanation of those effects, we consider further psychophysical investigations of visual perceptual performance under conditions similar to those used in the neurophysiogical experiments, so that we can relate the neurophysiological findings to a wide range of perceptual function. The psychophysical experiments considered next were performed with the identical apparatus to that used for the neurophysiology, and the stimuli used were some of those used in the neurophysiological experiments.

Experiments 2 and 3

The conclusions from the psychophysical experiments are as follows. First, with the same apparatus used for the neurophysiology experiment, and in which some neurons showed a continual improvement in responses as SOA increased and responded better to faces in their normal configuration than when rearranged, human subjects on a number of measures found a corresponding advantage as SOA was gradually increased, and for normal as compared to rearranged faces. Second, the psychophysical effects just noted were found when the judgement was for the clarity of the stimuli. Third, the psychophysical effects were found when the judgement was for whether the face was normal or rearranged. Fourth, the psychophysical effects were found when the judgement was for the identity of the face, a considerable extension of earlier psychophysical findings (Purcell and Stewart 1986, 1988).

General

The findings in this paper lead us to suggest a neurophysiological basis for the FDE, as follows. There are some neurons in the temporal lobes which respond preferentially to faces (see Rolls 1992). As noted elsewhere and documented here, some of these neurons are configuration sensitive, in that they respond more to faces when the features are in the normal than in the rearranged configuration (Perrett et al. 1982; Yamane et al. 1988). The activation of these configuration-sensitive neurons will be added to the other populations of neu-

rons that can be activated by the stimuli irrespective of whether the features are in the normal configuration or not. (These other neurons are likely to include neurons which respond at different stages of visual processing to edges, texture and the component features in faces.) The suggestion then is that the FDE arises because the total number of neurons activated by faces in their normal configuration is greater than that activated by the rearranged face. This would be consistent with the results of visual evoked potential studies. Several studies have suggested the existence of neural activity in the temporal lobes which can be evoked by tachistoscopic presentation of a face (Botzel and Grusser 1989; Jeffreys 1989). This evoked activity is largely independent of factors such as image size (Jeffreys 1989), but is affected by the configuration of facial features (Jeffreys and Tukmachi 1992). These results also suggest that there may be a greater number of neurons in the temporal cortex responsive to normal faces than to faces with rearranged facial features. If this hypothesis is correct then an observer has more neurons active with which to determine whether a face has been seen or not. Given that, as shown here (e.g. Fig. 5), the response of the configuration-sensitive neurons is only just greater to the normal configuration at short SOAs (e.g. 20 ms), but becomes greater at longer SOAs, the FDE will be particularly evident when the SOA is just above 20 ms. This matches the psychophysical results found by Purcell and Stewart (1986, 1988) and extended in experiments 2 and 3. This explanation of the FDE is a new result of the work described here. We note that the mean response of nine neurons became gradually larger to non-jumbled faces than to jumbled faces as the SOA was increased from 20 ms (see Fig. 5). The exact value at which this difference becomes significant depends of course on the number of neurons over which the difference is estimated. The human observer could use many neurons of the type described here when performing psychophysical judgements.

The results in this paper enable this argument to be extended to other psychophysical measures of human performance for faces, including clarity of a face, detection of whether it is rearranged or not, and its identification, all of which it is suggested arise because there are more face-selective neurons activated by face stimuli with their features in their normal configuration, so that more "analysers" on the basis of which a behavioural judgement can be made are activated with faces in their normal configuration. The proportion of neurons added by using stimuli in their normal configuration may be especially significant in the parts of the temporal lobe cortex concerned with processing faces, for it is here that neurons sensitive to configuration (Perrett et al. 1982; Yamane et al. 1988; experiment 1) and identity (Hasselmo et al. 1985; Tovee et al. 1993) are found, and it is presumably on the basis of the responses of these neurons that behavioural responses about the identity, clarity, etc. of faces may be made (Rolls 1992, 1994a, b).

The neurophysiological results described here also show that there is a population of neurons in the temporal cortical areas which respond to a single frame (16 ms) presentation of the test stimulus, a face. Second, the neurophysiological results show that the duration of the responses of these neurons to the short (16 ms) presentation of the stimulus is much longer than the stimulus, with, after an initial transient, a neuronal response that continued for several hundred milliseconds (see, e.g. Fig. 3). Third, the responses of these neurons do show backward masking. It was found that the effect of the mask, even though it was a non-face pattern mask which did not activate this population of neurons, was to cut off or interrupt the neuronal response to the face (see, e.g. Fig. 3). Fourth, for the neurons that respond to faces with the features in the normal configuration but have much smaller responses when the face features are rearranged, there are SOAs at which there is a large response to the normal but not to the rearranged face.

Another fascinating aspect of the results described here is that at the SOAs at which humans can just perform face identification (20 ms; see Fig. 8), the neurons in the temporal cortical areas respond for periods which are limited by the masking stimulus with this SOA to be 20–30 ms (see Fig. 3). This provides direct evidence that as little as 20–30 ms of processing time is sufficient for the temporal cortex to perform sufficient computation to just allow identification of a stimulus (although of course perception and identification improve if more processing time is available; see Fig. 7 and 8). Consistent with this, neurons in this region are not only known to have different responses to different stimuli early on in their response periods (Thorpe and Imbert 1989; Oram and Perrett 1992; Tovee et al. 1993), but it has recently been shown that more than half the total available information in the spike train of one of these neurons, taken over a long period, is available in a period as short as 20–40 ms taken near the start of the neuronal response to the visual stimulus (Tovee et al. 1993). The fact that sufficient processing for identification can be performed in such short times has important implications for the type of computation which could be being performed in a cortical area (Treves et al. 1995).

Acknowledgements This research was supported by Medical Research Council Grant PG8513790 to Dr. E.T. Rolls.

References

Baylis GC, Rolls ET, Leonard CM (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 GC, Rolls ET, Leonard CM (1987) Functional subdivisions of temporal lobe neocortex. J Neurosci 7:330–342

Botzel K, Grusser O-J (1989) Electric brain potentials evoked by pictures of faces and non-faces: a search for "face-specific" EEG-potentials. Exp Brain Res 77:349–360

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

Desimone R, Albright TD, Gross CG, Bruce C (1984) Stimulus-selective properties of inferior temporal neurons in the macaque. J Neurosci 4:2051–2062

Feigenbaum JD, Rolls ET (1991) Allocentric and egocentric spatial information processing in the hippocampal formation of the behaving primate. Psychobiology 19:21–40

Gorea A, Julesz B (1990) Context superiority in a detection task with line-element stimuli: a low-level effect. Perception 19:5–16

Hasselmo ME, Rolls ET, Baylis GC (1989a) The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. Behav Brain Res 32:203–218

Hasselmo ME, Rolls ET, Baylis GC, Nalwa V (1989b) Object-centered encoding by face-selective neurons in the cortex in the superior temporal sulcus of the monkey. Exp Brain Res 75:417–429

Humphreys GW, Bruce V (1989) Visual cognition. Erlbaum, London

Jeffreys DA (1989) A face-responsive potential recorded from the human scalp. Exp Brain Res 78:193–202

Jeffreys DA, Tukmachi ESA (1992) The vertex-positive scalp potential evoked by faces and by objects. Exp Brain Res 91:340–350

Merrill EG, Ainsworth A (1972) Glass-coated platinum-plated tungsten microelectrodes. Med Biol Eng 10:662–672

Oram MW, Perrett DI (1992) Time course of neural responses discriminating different views of the face and head. J Neurophysiol 68:70–84

Perrett DI, Rolls ET, Caan W (1982) Visual neurons responsive to faces in the monkey temporal cortex. Exp Brain Res 47:329–342

Perrett DI, Mistlin AJ, Chitty AJ, Smith PAJ, Potter DD, Broenniman R, Harries M (1988) Specialized face processing and hemisphere asymmetry in man and monkey: evidence from single unit and reaction time studies. Behav Brain Res 29:245–258

Purcell DG, Stewart AL (1986) The face detection effect. Bull Psychonomic Soc 24:118–120

Purcell DG, Stewart AL (1988) The face-detection effect: configuration enhances detection. Percept Psychophys. 43:355–366

Rolls ET (1984) Neurons in the cortex of the temporal lobe and in the amygdala of the monkey with responses selective for faces. Hum Neurobiol 3:209–222

Rolls ET (1991) Neural organisation of higher visual functions. Curr Opin Neurobiol 1:274–278

Rolls ET (1992) Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas. Philos Trans R Soc Lond Biol 335:11–21

Rolls ET (1994a) Brain mechanisms for invariant visual recognition and learning. Behav Proc (in press)

Rolls ET (1994b) Learning mechanisms in the temporal lobe visual cortex. Behav Brain Res (in press)

Rolls ET, Burton MJ, Mora F (1976) Hypothalamic neuronal responses associated with the sight of food. Brain Res 111:53–66

Rolls ET, Sanghera MK, Roper-Hall A (1979) The latency of activation of neurons in the lateral hypothalamus and substantia innominata during feeding in the monkey. Brain Res 164:121–135

Rolls ET, Tovee MJ (1994) Processing speed in the cerebral cortex and the neurophysiology of visual masking. Proc R Soc B 257:9–15

Rolls ET, Yaxley S, Sienkiewicz ZJ (1990) Gustatory responses of single neurons in the orbitofrontal cortex of the macaque monkey. J Neurophysiol 64:1055–1066

Schiller PH (1968) Single unit analysis of backward visual masking and metacontrast in cat lateral geniculate nucleus. Vision Res 8:855–866

Thorpe SJ, Imbert M (1989) Biological constraints on connectionist models. In: Pfeifer R, Schreter Z, Fogelman-Soulie F (eds) Connectionism in perspective. Elsevier, Amsterdam, pp 63–92

Tovee MJ, Cohen-Tovee EM (1993) The neural substrates of face processing models: a review. Cogn Neuropsychol 10:505–528

Tovee MJ, Rolls ET, Treves A, Bellis RP (1993) Information encoding and the responses of single neurons in the primate temporal visual cortex. J Neurophysiol 70:640–654

Treves A, Rolls ET, Tovee MJ (1995) On the time required for recurrent processing in the brain. Proc Natl Acad Sci USA

Valentine T (1991) A unified account of the effect of distinctiveness, inversion, and race in face recognition. Q J Exp Psychol 43A:161–204

Valentine T, Bruce V (1986) The effects of distinctiveness in recognizing and classifying faces. Perception 15:525–535

Yamane S, Kaji S, Kwano K (1988) What facial features activate face neurons in the inferotemporal cortex of the monkey? Exp Brain Res 73:209–214

Young MP, Yamane S (1992) Sparse population coding of faces in the inferotemporal cortex. Science 256:1327–1331