Perception & Psychophysics
1996, 58 (4), 602-612

Effects of high-pass and low-pass spatial filtering on face identification

NICHOLAS P. COSTEN, DENIS M. PARKER, and IAN CRAW
University of Aberdeen, Aberdeen, Scotland

If face images are degraded by block averaging, there is a nonlinear decline in recognition accuracy as block size increases, suggesting that identification requires a critical minimum range of object spatial frequencies. The identification of faces was measured with equivalent Fourier low-pass filtering and block averaging preserving the same information and with high-pass transformations. In Experiment 1, accuracy declined and response time increased in a significant nonlinear manner in all cases as the spatial-frequency range was reduced. However, it did so at a faster rate for the quantized and high-passed images. A second experiment controlled for the differences in the contrast of the high-pass faces and found a reduced but significant and nonlinear decline in performance as the spatial-frequency range was reduced. These data suggest that face identification is preferentially supported by a band of spatial frequencies of approximately 8–16 cycles per face; contrast or line-based explanations were found to be inadequate. The data are discussed in terms of current models of face identification.

The questions of whether the information concerning the identity of faces is carried by a limited range of spatial scales and whether the potential information from different regions of the spatial spectrum is given equal weight in the determination of identity have been approached in a number of different ways. One method of considering these issues has been to make use of spatial-frequency filtering techniques (Harmon, 1973). However, variations in this method have produced contradictory results, with notably different conclusions about the relative importance of different spatial-frequency bands specified in terms of cycles per face. The term *cycles per face* is defined as the number of sinusoidal repetitions of a given width that can be placed within the eye-level width of the face. The use of this metric to describe the information present in stimuli allows discussion of the degree of detail necessary for recognition, perhaps by defining the scale of facial configuration. A class of objects has a configuration if there is a consistent set of features all arranged in the same order. Thus, if a set of examples are superimposed, normalizing for scale and viewpoint, another example of the class is produced that is closer to the prototype. Clearly, faces have this property, since all have two eyes, a nose, and a mouth—and these are consistently arranged.

Harmon (1973), who used images created by local block averaging, was the first to consider spatial scale and identification; examples of this technique, known as *pixelizing*, can be seen in Figure 2. The images are formed by placing a regular square grid across the image and setting the pixel value at each grid square to the average gray level within it. This work suggested that the minimum image quality that allows effective identification corresponds to a 16×16 pixel image; however, since the images did not take up the whole of the screen, the number of pixels per face was slightly lower. Harmon also used a smooth low-pass filtering technique. This type of filtering operation does not introduce additional spatial frequencies (noise), as the pixelization procedure does (see Figure 1, for a comparison of low-pass and pixel spectra). With this procedure, the minimum image quality for identification was found to be the surprisingly low value of 2.5 cycles per face, measured at eye level. Harmon mentions a bald subject as being consistently well recognized. This figure of 2.5 cycles per face corresponds with 5 pixels per face. The smallest detail retained after pixelization will be that retained after Fourier (smooth) low-pass filtering, which removes sinusoidal components, with a cut-off wavelength twice the width of a pixel. Thus, the smooth-filtered images were recognized with a lower frequency cut-off than were the pixelized versions, probably because of the introduction of additional irrelevant spatial-frequency noise in the latter versions that masked the preserved face-specific information.

When Fiorentini, Maffei, and Sandini (1983) considered the best range of the spatial frequencies at which identification is supported, they found a sharp decrease in accuracy when the highest spatial frequency in the image dropped from 8 to 5 cycles per face. When the lower spatial frequencies were removed, a similar decline was seen between 8 and 12 cycles per face. Bachmann (1991) considered the identification of pixelized faces, quantized at levels between 15 and 74 pixels per face. Faces with 15 pixels (7.5 cycles) were correctly identified on approximately 45% of the occasions, but those with 18 pixels (9

N.P.C. was employed on ESRC/MRC/SERC Joint Council Initiative in Cognitive Science and Human–Computer Interaction Grant 9002054 to I. Craw, H. D. Ellis, and J. W. Shepherd, and SERC Grant GR/J04951 to V. Bruce, A. M. Burton, and I. Craw. N.P.C. is now at ATR Human Information Processing Research Laboratories, 2-2 Hikaridai, Seika-cho Soraku-gun, Kyoto 619-02 Japan. D.M.P. is now at Glasgow Caledonian University. Correspondence should be addressed to D. M. Parker, Department of Psychology, Glasgow Caledonian University, Cowcaddens Road, Glasgow G4 0BA, Scotland (e-mail: costen@hip.atr.co.jp).

cycles) and above were all identified on approximately 80% of trials. These studies concur in indicating that below 8 cycles per face for smoothly filtered images, or 9 cycles (18 pixels) per face for pixelized images, there are problems. None of these studies give identification times, and, thus, it is impossible to tell if efficiency varied for the scales at which identification was possible or if processing strategies changed at any point.

Costen, Parker, and Craw (1994) measured both speed (response time, RT) and accuracy for 42-, 21-, and 11-pixel faces (21, 11.5, and 5.5 cycles per face), as well as Fourier and Gaussian low-pass versions calculated to remove the same range of spatial frequencies. A reduction of the maximum detail from 22 to 10.5 cycles per face produced no change in accuracy and an increase in RT only for the pixelized condition. A further drop to 5.5 cycles per face both reduced accuracy and increased RT for all conditions. The reduction in performance was more pronounced for the pixelized images, replicating the discrepancy between Fiorentini et al. (1983) and Bachmann (1991) and also that seen by Harmon (1973) between pixelized and Fourier-filtered images. Harmon and Julesz (1973) suggested that this discrepancy might be due to masking by adjacent spatial-frequency components introduced by the pixelizing process. Morrone, Burr, and Ross (1983) suggested spatial-domain mispositioning of facial contours used for the determination of shape as an additional problem with the pixelized images. In a second experiment, Costen et al. (1994) also measured the RT and accuracy for images with the same noise spectrum as the pixelized images, but with random orientations of the Fourier components. Performance on the unstructured-noise images was intermediate between the pixelized and Fourier low-pass images, suggesting that, while a substantial portion of the decrement could be accounted for by unstructured noise, the edge elements of the pixels did cause additional interference.

Although these results are relatively consistent, other studies show lesser effects or higher optimal frequencies. Hayes, Morrone, and Burr (1986) considered the identification, or matching (it is unclear which), of band-passed images, some of which had 180° phase shifts, with 1.5-octave-wide frequency bands, centered between 3.2 and 50 cycles per face. They found 80% accuracy with 25 cycles per face and about 60% accuracy with 6.4 or 50 cycles per face. Schuchard and Rubin (1989) also measured performance on a *same–different* task, with similar bands centered on 4.0, 11.2, and 31.7 cycles per face, and they found no difference between the three spatial-frequency levels. If both of these studies are classed as matching tasks rather than identification tasks, the results suggest a higher optimal spatial band for the former task.

Tieger and Ganz (1979) approached the problem by combining faces with two-dimensional sine-wave gratings whose frequencies lay between 3.2 and 31.2 cycles per face and found that the mask with 17.6 cycles per degree (cpd) had the most detrimental effect on identification. Moscovitch and Radzins (1987; see also Bryer, 1988, and Moscovitch, 1988, on the proper analysis to be applied) used random dots with groupings of between 1.1 and 52.8 cpd as masks and found no differential effect. Keenan, Witman, and Pepe (1989; see also Keenan, Witman, & Pepe, 1990) used masks formed by vertical square-wave bars with fundamental frequencies between 1.6 and 76.8 cpd; although it was not clear whether there was a significant effect of the spatial frequency of the mask, 76.8 cycles per face showed less masking than did lower frequencies.

This discrepancy between studies that have varied the effective spatial-frequency range of the images by altering either the upper or the lower cut-off of the filters (Bachmann, 1991; Costen et al., 1994; Fiorentini et al., 1983; Harmon, 1973) and those that have attempted to remove a constant spatial range by removing a band of information within the images' spectrum (Hayes et al., 1986; Tieger & Ganz, 1979) could have a number of explanations. The first is that it is due to variation in image contrast. The energy at a specified frequency in a real image will decline with at least the square of the frequency (Edwards, 1967). In addition, humans are differentially sensitive to spatial frequencies, with a maximum sensitivity at approximately 2–4 cpd (Campbell & Green, 1965), and RT is monotonically related to spatial frequency (Parker, 1980). Thus, it is possible that the additional contrast present in low-passed images disproportionately enhances performance and underestimates optimal spatial frequency. The matching accuracy for the faces in Hayes et al. (1986) order correctly for a contrast sensitivity effect, as do the masking results of Tieger and Ganz (1979).

A second possibility is that this discrepancy is a consequence of the range of spatial frequencies in the images. However, this explanation fails to account for the results of Fiorentini et al. (1983). Their low-pass limit of 6.5 cycles per face suggests a range of approximately 4.5 octaves of preserved information, but the high-pass limit of 10 cycles per face suggests a range of 0.5 octaves, since their images did not preserve components above 15 cycles per face. Nor can this idea explain the results of Costen et al. (1994), where the additional decline in the pixelized images over that seen with Fourier and Gaussian filtering reflected the effect of masking within this critical band.

A third explanation might be in terms of the task that the subjects were asked to perform. Costen et al. (1994), Bachmann (1991), and Fiorentini et al. (1983) taught their subjects to recognize the faces, and Harmon (1973) used familiar faces, whereas the subjects of Hayes et al. (1986) matched the filtered face to an original, as did those of Schuchard and Rubin (1989) and Moscovitch and Radzins (1987). Tieger and Ganz (1979) required a judgment of familiarity, but not of identity; only Keenan et al. (1989), whose results are rather unclear, used an identification-from-memory paradigm. In this respect, these results resemble those of Sergent (1986), who found a left cerebral hemisphere RT advantage for semantic judgments (identification and occupational category of 16 faces personally familiar to the subjects), but none for a sex judgment. Low-passing the images so that information above 6 cycles per face was lost suppressed this effect, and all three tasks then showed a right hemisphere advantage. This suggests that different tasks are supported by different spatial frequencies and, thus, that the results from the constant-frequency-range studies may not reflect face identification, but rather perceptual matching of some sort.

To more fully investigate this area, we decided to replicate the Fiorentini et al. (1983) study and to extend it by including a measure of the time to make identity decisions. Additionally, a set of images pixelized by the method used by Bachmann (1991), Costen et al. (1994), and Harmon (1973) was added with the same range of spatial frequencies removed. The measurement of RT allows the exclusion of the possibility that a discontinuity in identification is simply due to a move from ceiling performance. Addition of a time stress also allows consideration of the method and information used in identification—for example, distinguishing between a natural, configural identification process and a slower critical feature strategy. As well as allowing reconciliation of the small difference between the estimates of the lower limit upon efficient identification provided by Bachmann (1991) and Fiorentini et al. (1983), the comparison of Fourier-filtered and pixelized images will also allow consideration of the difference between these values and the higher critical range suggested by Hayes et al. (1986) and Tieger and Ganz (1979). By simultaneously considering the effects of reduction of spatial-frequency range by both high-pass and low-pass Fourier filtering and by the introduction of extraneous information by the block-averaging process in the proposed critical region, we should be able to shed further light on which spatial scales dominate the face-identification process.

EXPERIMENT 1

Method

Subjects. The subjects were 9 undergraduate students at the University of Aberdeen. The data from 1 subject were dropped because of unacceptably slow performance. This subject's performance was significantly worse than that of the others.

Equipment. The basic equipment was an Imaging Technology FG100V framegrabber card fitted to a Sun 3/160M. A maximum of 512 × 512 pixels, 8 bits deep, were accessible or visible at any one time. The images were displayed on a Phillips 35-cm color monitor. The experiment was controlled from the Sun workstation using compiled C programs and executable shell scripts under Sun UNIX. To record the RTs and choices, an Apple IIe with a specially built nine-choice button box (three of the buttons were masked off) was used. The Apple IIe was triggered by the framegrabber device driver and returned the choice and RT for recording. On the rare occasion that the two computer systems got out of step, the Apple did not return any data; 69 of the 4,504 responses are not included in the analysis. The experimenter monitored the computers continually, and resynchronization required only the pressing of a button.

Stimuli. Six photographs of males rated as typical from a collection gathered at Aberdeen University for the Facial Retrieval and Matching Equipment (FRAME) database of face measurements and ratings (Shepherd, 1986) were digitized at a resolution of 128 × 128 pixels. The faces measured about 160 × 210 mm of the screen. The subjects were seated with their heads approximately 1 m from the screen, so the images subtended about 9.2° × 12.1°.

These base pictures were used in the practice session, with the addition of six half-sized pictures that were labeled A–F (by which the images were referred to on the button box in the same consistent arrangement). The background of these latter images (the *correction* images) was black. The experimental images were formed by processing the pictures with the HIPS image-processing package (Landy, Cohen, & Sperling, 1984), running under Unix on a Sun workstation. Images were pixelized to 45, 23, 12, and 9 pixels per face, measured horizontally at eye level by gray level averaging using a grid placed upon the image and setting each pixel to the mean brightness of the block.

Low-pass images were created by taking Fourier transforms, applying exponential low-pass filters (with second-order roll-offs) of the appropriate value, and taking inverse Fourier transforms. This applied the equation

$$f = \exp\left\{-\left(\frac{r}{c_l}\right)^2\right\},$$

where f is the multiplication faction, r is the component radius, and c_l is the low-pass cut-off radius. High-pass images were created with the same parameters and equation except that r and c_l were transposed and thus lacked low spatial frequencies. The low-pass cut-offs were selected so that a checkerboard with squares of the size given by the pixel manipulation would be just noticeable as a nonuniform field when low-passed. Thus, the same information was preserved, but without introducing extra noise. This needed spatial frequencies with wavelengths of twice the pixel side length (so 9, 12, 23, and 45 pixels per face equated with maximum frequencies in the Fourier low-pass condition of 4.5, 6, 11.5, and 22.5 cycles per face). Examples of the filters are shown in Figure 1 for a 23-pixels-per-face image, and a set of examples of the images are provided in Figure 2.

Procedure. At the start of the testing session, the experimenter explained that the subject would be taught to identify six faces, and the importance of accuracy was stressed. The experiment was subject-paced; a response triggered the next trial. On a practice trial, the subject saw in sequence a fixation point, a blank white field, and then one of the images chosen at random, each for 1 sec. The screen then showed the white blank field until the subject made his/her choice, aided by a sheet with the six correction images in the same configuration as on the button box, or for 1 sec, whichever was longer. The correction stimulus associated with the image seen immediately before was then shown for 2 sec and was also identified. This procedure was followed for three blocks of 18 trials; each block randomly repeated the six images three times.

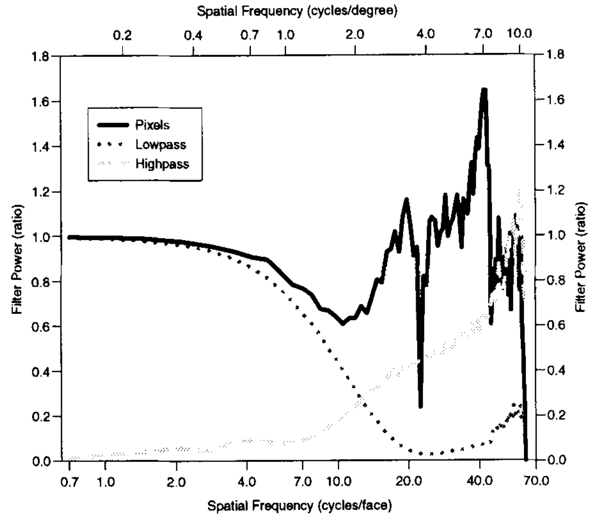

Figure 1. Fourier filter functions used on one image at 11.5 cycles per face (23 pixels per face) for high-pass, low-pass, and pixelization manipulations. Note that these are derived functions; the pixel manipulation was performed in the real domain. The upper abscissa shows the frequency in cycles per degree. The lower abscissa shows the frequency in cycles per face. The ordinates show the ratio of filter power.

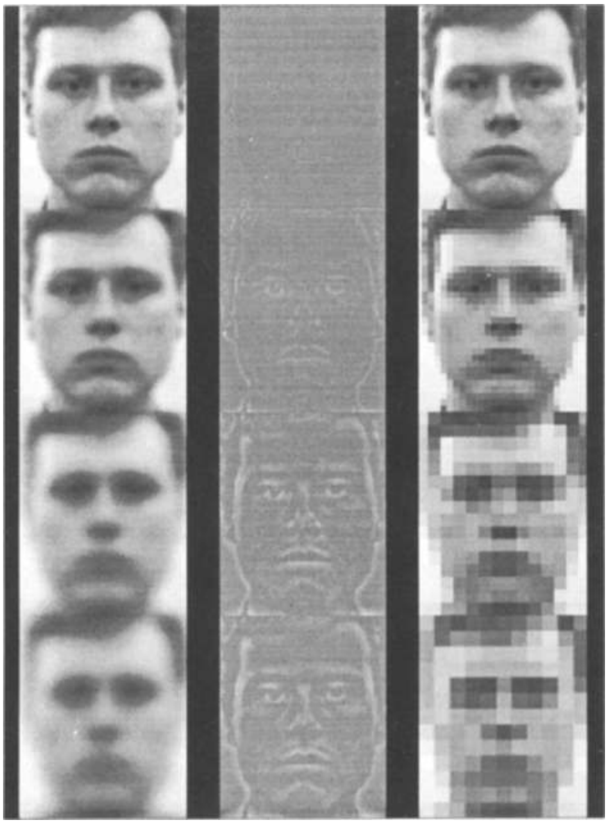

Figure 2. Low-pass, high-pass, and pixel images at 42, 23, 12, and 9 pixels per face. Note that the prints underestimate the contrast of the images displayed in the experiment, particularly the high-pass images.

Although the presence of the sheet and correction stimuli reduced the potential learning effect of response accuracy, there was a significant effect of practice across blocks [$F(2,14) = 4.08, p = .0403$] and a significant effect of the target used [$F(5,35) = 3.92, p = .0063$], but these factors did not interact ($F = 1.00$). The same pattern was found with RT [$F(2,14) = 17.13, p = .0002$; target [$F(5,35) = 6.58$, $p = .0002$; interaction, $F(10,69) = 1.54, p = .1448$]. Since the mapping from targets to buttons on the button box was not varied, some of this main effect of targets may reflect differences due to the position of the buttons.

The sheet of images was then taken from the subjects, and they were told that the experimental session would be starting. The experimenter said that he was interested in the amount of information needed to identify the images and that he had manipulated the images in a number of ways. The subjects were instructed to ignore the manipulation as much as they could and concentrate on identifying the images. The subjects were urged to respond quickly, even if this meant that a few errors would occur. Experimental trials ran in the same way as the practice ones, except that there were no correction stimuli and the stimuli were presented for only 100 msec. The experiment had eight blocks of 72 randomly ordered stimuli (6 targets × 3 manipulations × 4 spatial-frequency cut-offs), and no feedback was given.

Results

Since, in one case, spatial-frequency cut-off refers to a limit below which components were restricted (high-pass), whereas in the others, components above the cut-off were restricted, overall comparisons will show only a crude effect of spatial-frequency range. The major interest here lies in the simple effects. Accuracy data are displayed in Figure 3 and show the replication of the Fiorentini et al. (1983) and Bachmann (1991) findings. These results are derived from the inverse-sine transformation of the data, which reduced the significant correlation between means and variances of the groups. The reported means are the sine transformation of the mean values of the groups. An analysis of variance (ANOVA) of the accuracy data showed significant effects of spatial-frequency cut-off (blur) [$F(3,21) = 7.10, p = .0018$] and of type of image processing (manipulation) [$F(2,14) = 12.35, p = .0008$]. The interaction between blur and manipulation was significant [$F(6,42) = 15.53, p < .0001$].

The significant interaction between blur and manipulation indicated that simple effects should be analyzed. For the high-pass condition, there was a significant effect of blur [$F(3,21) = 25.91, p = .0001$]. A Tukey HSD test ($\alpha = .05$) showed that this effect was due to a difference between the 22.5-cycles-per-face (45-pixels-per-face) group and the other groups, which did not differ between themselves. The low-pass condition also had a significant effect of blur [$F(3,21) = 3.08, p = .0499$]. A Tukey HSD test showed that there was a significant difference between the 4.5-cycles-per-face (9-pixels-per-face) group and the 22.5-cycles-per-face (45-pixels-per-face) groups, although neither of these were significantly different from the 6- and 11.5-cycles-per-face (12- and 23-pixels-per-face) groups. The pixelized condition also had a significant effect of blur [$F(3,21) = 19.86, p < .0001$]. A Tukey HSD test showed that data could be divided into two groups: 4.5 and 6 cycles per face (9 and 12 pixels per face) and 11.5 and 22.5 cycles per face (23 and 45 cycles per face).

In summary, the accuracy data show that all three manipulations had significant effects of spatial frequency. The

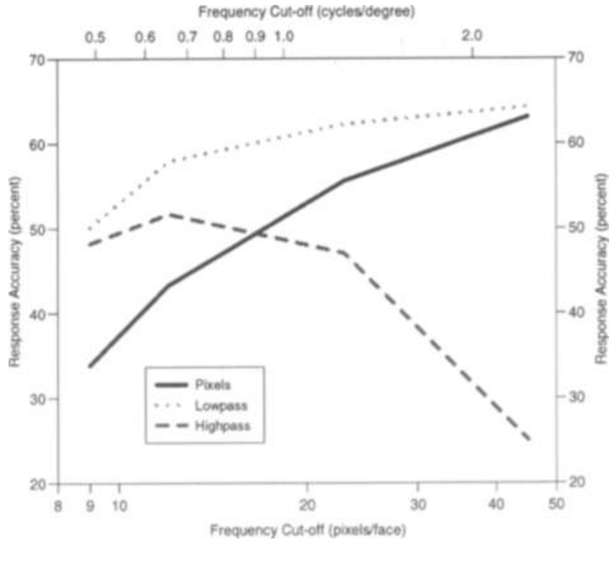

Figure 3. Response accuracy in Experiment 1. The upper abscissa shows the frequency cut-off in cycles per degree. The lower abscissa shows the frequency cut-off in pixels per face. The ordinates show response accuracy in percent.

low-pass condition showed a drop in performance when the maximum spatial frequency dropped to 4.5 cycles per face (9 pixels per face), whereas the pixelized condition showed a drop in performance when the maximum preserved spatial frequency dropped below 11.5 cycles per face (23 pixels per face). The high-pass condition, which showed a maximum identification accuracy comparable to the lowest accuracy of the low-pass condition, showed a further drop in performance when the minimum frequency present rose above 11.5 cycles per face (23 pixels per face). It is notable that when information of around 11.5 cycles per face (23 pixels per face) was present, the response accuracy was not significantly affected by the method used to restrict the spatial-frequency range [$F(2,14) = 3.41, p = .218$].

To normalize the correct RT data because of a significant positive skew, we applied a 5% upper cut-off to each subject to remove aberrant points and took a base 10 logarithm transformation that removed the positive correlations between means and standard deviations. All displayed means (see Figure 4) are harmonic means, formed by finding the averages of the logarithmic values and then taking anti-logs. Using these data and the standard ANOVA for repeated measures, there was no significant effect of blur [$F(3,21) = 1.83, p = .1720$], but there was a significant effect of manipulation [$F(2,14) = 6.56, p = .0098$], and the interaction between blur and manipulation was also significant [$F(6,15) = 3.00, p = .0157$].

For the high-pass condition, there was no effect of blur [$F(3,21) = 1.68, p > .2$]; however, for the low-pass condition, there was a significant effect of blur [$F(3,21) = 4.82$, $p = .0094$], and a Tukey HSD test showed that all four levels of blur were significantly different from each other. The pixelized condition also showed a significant effect of blur [$F(3,21) = 3.10, p = .0229$]. A Tukey HSD test showed that the 22.5-cycles-per-face (45-pixels-per-face) group was significantly different from the other three groups, which did not differ significantly.

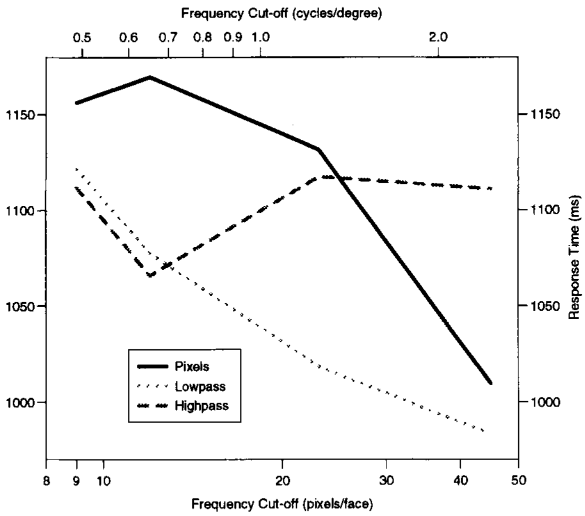

Figure 4. Response times in Experiment 1. The upper abscissa shows the frequency cut-off in cycles per degree. The lower abscissa shows the frequency cut-off in pixels per face. The ordinates show response time in milliseconds.

The data were then divided up by the values of blur. Although there were no significant effects for the 4.5- and 22.5-cycles-per-face (9- and 45-pixels-per-face) groups, the 6-cycles-per-face (12-pixels-per-face) group showed a significant effect of manipulation [$F(2,14) = 4.46, p = .0318$]. A Tukey HSD test showed that this was due to differences between the pixelized images and the high-pass and low-pass images, which did not differ between themselves. For the 11.5-cycles-per-face (23-pixels-per-face) group, there was also a significant effect of manipulation [$F(2,14) = 11.90, p = .0010$]. A Tukey HSD test showed that this effect was due to a difference between the low-pass condition and the high-pass and pixelized conditions, which did not differ between themselves.

In summary, the RT data show that the time to correctly recognize the high-pass images was not significantly affected by the spatial frequency cut-off, but it increased for the low-pass and pixelized conditions as the spatial frequency range was reduced. This occurred at a faster rate for the pixelized images than for the low-pass images, with significant additional reductions at 6 and 11.5 cycles per face (12 and 23 pixels per face).

Discussion

The first thing to be noted about Experiment 1 is that there is a comparative absence of significant differences in RTs if the high-pass manipulation is included. This derives from the relatively low accuracy with which these images were identified, dropping to 40% for the pixelized condition and 30% for the high-pass condition at their minimum spatial ranges. Although the RT measure considers only the occasions on which the subject is correct, it is not possible to say why the subject is correct on any particular trial. When accuracy drops, the proportion of occasions on which the subject guesses at random and is correct will increase. Unless the RT for guesses happens to coincide with the characteristic RT for trials upon which identification is possible at that spatial-frequency range, this will increase the variance and thus decrease the power of the design.

Within these considerations, both the RTs and the response accuracies replicate those of Bachmann (1991), Costen et al. (1994) and Fiorentini et al. (1983). The low-pass and pixel conditions showed increases in RT and declines in accuracy as image quality was reduced, and these were greater in the latter case. The results suggest that pixelization introduces noise that masks the remaining image. Costen et al. (1994) suggest that this reflects a combination of masking of the effective upper limit of the available information by medium-frequency noise and bias in locating portions of the image introduced by the sharp edges of the pixels.

The high-pass manipulation exhibits a large drop in identification accuracy, replicating the results of Fiorentini et al. (1983), who found increased errors as the cut-off frequency of their high-pass images increased, although the results here suggest a rather higher frequency band than was suggested by Fiorentini et al. Since there was no significant change in RT for the high-pass manipulation, the results of Fiorentini et al. cannot reflect a change in strategy and so support the view that there is some upper limit to the useful range of spatial frequencies.

The view that there is a band of object spatial frequencies that is preferentially used in face identification agrees with the results of Fiorentini et al. (1983), although Keenan et al. (1989) did not see such differences, and Tieger and Ganz (1979) and Hayes et al. (1986) obtained results that would lead one to expect no difference between the 11.5- and 22.5-cycles-per-face (23- and 45-pixels-per-face) conditions. From these and our experiments, this band would be situated somewhere between 8 and 16 cycles per face (16 and 32 pixels per face).

It is possible that the use of faces selected as typical, in an attempt to ensure that identification was relatively difficult and thus avoid ceiling effects, has affected the spatial frequency value obtained. However, facial measures of typicality suggest that this is unlikely. Bruce, Burton, and Dench (1994) found positive correlations at a number of size (distance) scales between ratings of distinctiveness and the deviations from the means of facial distances. The two highest correlations reported for male faces, normalized by equating the interocular distances, were the height of the eyebrows and the face width at the mouth. The measurements of small-scale distances will be dependent upon high frequencies; low-pass images will not allow accurate location of closely adjacent points. Conversely, large-scale distances will depend upon low frequencies, since their absence will increase the probability of the comparison of inappropriate contours. Thus, the typical distances should determine the typical frequency components, and vice versa. In addition, while Vokey and Read (1992) suggest typicality can be decomposed into "general familiarity" and memorability factors, O'Toole, Deffenbacher, Valentin, and Abdi (1994) suggest that, although general familiarity (or as they prefer, "attractiveness") is dependent upon the shape of faces, memorability is dependent upon small, discrete, local features. From these results, it appears that typicality and distinctiveness are probably the result of a number of different effects operating at different scales, and so there should be little interaction between typicality and an optimal spatial-frequency range.

The change in identification performance obtained here could also be attributed to a drop in contrast as the spatial range is reduced. As the frequency of Fourier components increases, the intensity of the components decreases (Edwards, 1967). Since a large proportion of the total available contrast is accounted for by very low spatial frequencies, as the frequency of the cut-off in the high-pass condition increases, the available contrast falls much faster in the low-pass conditions. Thus, a given frequency band contains a larger fraction of the contrast in an image that has been high-passed than in one that has been subjected to an equivalent low-pass.

This problem with contrast may underlie the fact that these data disagree somewhat with those collected by Fiorentini et al. (1983), since the identification rate for the 4.5-cycles-per-face (9-pixels-per-face) high-pass images was significantly lower than that for the low-pass images with the same cut-off. One explanation of this effect could be that the overall energies of our high-pass images are low. When the inverse Fourier transform was applied to the images, the output format permitted only positive values. Since the fundamental of a Fourier transform is the mean gray level of the image and all higher frequencies form deviations from this base, this removed half of the contrast of the image. This manipulation was applied to all of the images; thus, the low-pass and high-pass were computationally identical. However, the low-pass images will not be noticeably affected by this feature, since the decline in energy with increased spatial frequency ensures that negative gray levels are impossible. Although Fiorentini et al. (1983) do not state their computational method, inspection of the examples given suggests that their images did not suffer from this problem and may also have had the mean contrast enhanced. To test these options, the high-pass condition was repeated, with contrast-enhanced and uncensored *float* output conditions added. These had the effect of enhancing the remaining contrast in the high-passed images, the first retaining the requirement that the images contain only the positive deviations from the removed mean brightness, and the second with both positive and negative deviations.

EXPERIMENT 2

Method

Subjects. The subjects were 8 undergraduate students at the University of Aberdeen.

Equipment. The equipment was the same as that used in Experiment 1.

Stimuli. The base stimuli were the same as those used in Experiment 1. The high-pass condition was identical to that in Experiment 1. The stretched condition was prepared in the same way, but, before the DC component was added, the contrast of the image was enhanced. This was performed by linearly increasing the range of gray levels present in the image to the full range available (0–255) and thus equated the Michelson (1927/1962) contrast of the set of images.

The float images were constructed in the same way as the stretched images, but care was taken to ensure that, at each stage, there were no constraints upon pixel values. Only at the final stage, when the Michelson contrast was set, was the image converted for display purposes. The results of these filters are shown in Figure 5. This displays the data in the same manner as Figure 1 and shows the increased contrast for the stretched and float images relative to the high-pass images, and it shows that only the float images had the smooth, curved filter pattern expected from an exponential filter. The images were again manipulated by the HIPS image-processing package running from Unix shell scripts. A set of examples of the images are provided in Figure 6.

Procedure. The procedure was the same as that used in Experiment 1. Of a total of 4,608 trials, 66 (1.43%) were discarded because of equipment missynchronization.

The practice procedure produced a significant increase in response accuracy across blocks [$F(2,14) = 5.29, p = .0194$]. RT also showed a significant effect of practice [$F(2,14) = 21.44, p = .0001$].

Results

The accuracy results were treated in the same way as those in Experiment 1 and are displayed in Figure 7, taking the sine of the values used in the analyses. There were significant effects of spatial-frequency cut-off (blur) [$F(3,21)$

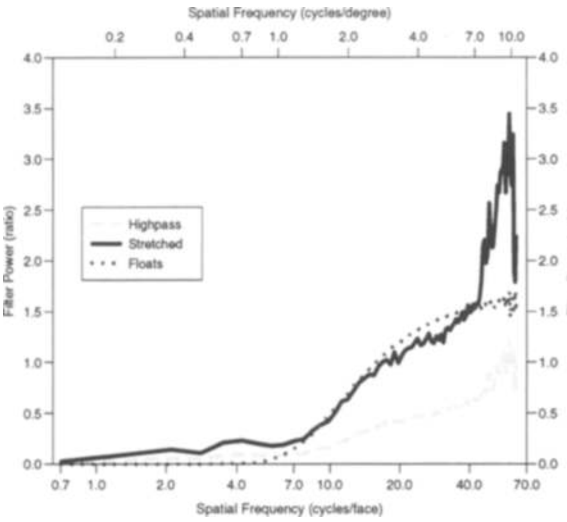

Figure 5. Fourier filter functions used on one image at 11.5 cycles per face (23 pixels per face) for high-pass, stretched, and float manipulations. The upper abscissa shows the frequency in cycles per degree. The lower abscissa shows the frequency in cycles per face. The ordinates show the ratio of filter power.

$= 40.35, p = .0001$] and of the type of filter used (manipulation) [$F(2,14) = 34.61$, $p = .0001$]. There was also a significant interaction between blur and manipulation [$F(6,42) = 15.18, p = .0001$].

The significant interaction between blur and manipulation demands an analysis of simple effects. When the data were sorted on the basis of manipulation, all three showed significant effects of blur (all ps $< .0002$). For the high-pass manipulation, a Tukey HSD test showed that the 22.5-cycles-per-face group was significantly different from the 4.5-, 6-, and 11.5-cycles-per-face groups and that, while the 4.5- and 11.5-cycles-per-face groups were significantly different from each other, neither differed significantly from the 6-cycles-per-face group. The stretched manipulation showed that, while the 22.5-cycles-per-face group was significantly different from the 4.5- and 6-cycles-per-face groups, the 11.5-cycles-per-face group did not differ from any other. For the float manipulation, the 22.5-cycles-per-face group was significantly different from the others, which did not vary among themselves.

When the data were divided on the basis of the values of the blur factor, it was found that the 4.5-cycles-per-face condition showed a significant effect of manipulation [$F(2,14) = 4.28, p = .0355$]. A Tukey HSD test showed that there was a significant difference between the stretched and float manipulations, but not between these and the high-pass condition. At the 6-cycles-per-face level, there was no significant effect of manipulation [$F(2,14) = 1.49$, $p = .2594$], but there was at the 11.5-cycles-per-face level [$F(2,14) = 9.89, p = .0021$]. The float manipulation here was significantly different from the stretched and high-pass manipulations, which did not differ significantly from each other. There was also a significant effect at the 22.5-cycles-per-face level [$F(2,14) = 64.85$, $p = .0001$]. The high-pass condition was significantly different from the stretched and float conditions, which did not vary between themselves.

In summary, then, accuracy data show that, although enhancing the contrast of the images did improve identification when the minimum frequency rose above 11.5 cycles per face, there was still a significant drop in performance at this level for both the contrast-enhanced conditions. The relatively small differences between the stretched and float manipulations suggest that total gray-level range was of relatively little importance, although it is notable that there was a significant difference between these conditions when the minimum frequency was 11.5 cycles per face.

The mean RTs, formed by taking the anti-logarithms of the means of the groups used in the analysis are given in Figure 8. Analysis found that there were significant effects of blur [$F(3,21) = 7.96$, $p = .001$] and of manipulation [$F(2,14) = 6.59, p = .0096$]. The interaction between blur and manipulation was not significant ($F < 1$). Post hoc Tukey tests showed that, although the 22.5-cycles-per-face frequency cut-off was significantly different from the 4.5-, 6-, and 11.5-cycle-per-face groups, these did not vary significantly among themselves. They also showed that the high-pass manipulation was significantly different from both the stretched and the float manipulations, which did

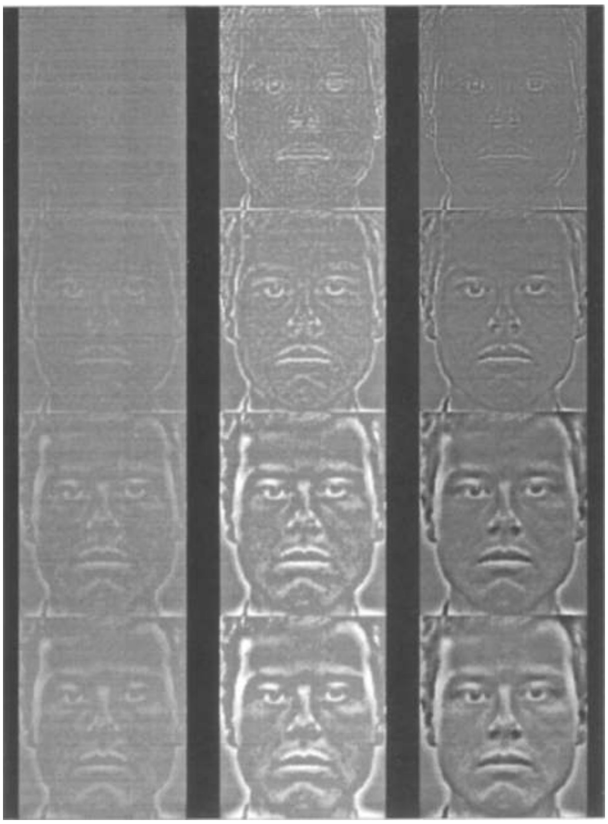

Figure 6. High-pass, stretched, and float at 42, 23, 12, and 9 pixels per face. The prints underestimate the contrast of the images used in the experiment.

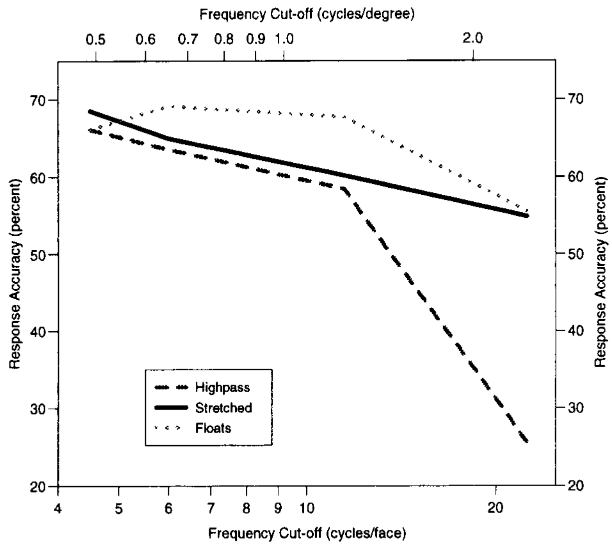

Figure 7. Response accuracy in Experiment 2. The upper abscissa shows the frequency cut-off in cycles per degree. The lower abscissa shows the frequency cut-off in cycles per face. The ordinates show response accuracy in percents.

not vary significantly between themselves. Since there was no significant interaction between these factors, simple effects were not extracted.

Discussion

The major results of Experiment 2 are relatively clear. As in Experiment 1, the high-pass condition, which did not have enhanced contrast, showed a large drop in the accuracy with which the face was identified and an accompanying increase in the time taken to do this correctly. The stretched and float conditions showed the same pattern of results, with significantly worse performance on the 22.5-cycles-per-face condition than on the others. However, the stretched and float images were identified more accurately and faster than were the high-pass images. It would appear that a proportion of the drop in performance seen in the high-pass condition as the spatial-frequency of the cut-off increased was due to the drop in the contrast of the images. However, it is not clear whether all of this decline was due to the loss of contrast or whether some of it was due to the loss of some critical band of information. This can be further investigated by considering different definitions of the contrast of the images. The images have been manipulated to control the Michelson (1927/1962) contrast, which is a function of the maximum and minimum brightness. However, other measures, such as the energy or the root mean square (Moulden, Kingdom, & Gatley, 1990), also could have been used. This statistic measures the average deviation from the mean gray level. When the root mean square of the images was measured, there were consistent effects of spatial-frequency cut-off [$F(3,15) = 250.02, p \ll .0001$], of manipulation [$F(2,10) = 212.52, p \ll .0001$], and of the interaction between these factors [$F(6,30) = 0.0039, p = .0039$], as can be seen in Figure 9.

It is clear that the variation in the root mean square values for these images will not account for the experimental results. While the relationship between the high-pass and contrast-enhanced images was approximately the same in both cases, the float images showed more accurate and faster identification than did the stretched images. However, they had significantly lower energies. In particular, the energy of the 4.5-cycles-per-face high-pass images was the same as that of the 22.5-cycles-per-face float images and was less than that of the 22.5-cycles-per-face stretched images. A one-way ANOVA of these three conditions showed that this difference was significant [$F(2,10) = 8.49, p = .007$]. A Tukey HSD test showed that this reflected a significant difference between the stretched images and the float and high-pass images, which did not vary significantly. However, the high-pass images showed more accurate identification [$F(2,14) = 15.51, p = .0003$] than did the stretched or float images, although there was no significant difference in RT for these images [$F(2,14) = 1.44, p = .2710$]. Conversely, the significant drop in contrast that is observed at 4.5 cycles per face when comparing the high-pass images with the stretched and float images produced no significant difference in the speed of response, and it produced only a slight difference between the high-pass and stretched conditions. The float condition was not significantly different from either of these conditions. The results show clearly that the available image contrast and performance were almost totally decoupled.

GENERAL DISCUSSION AND CONCLUSIONS

The data from the low-pass conditions in Experiment 1 readily confirm the evidence that there is a discontinuity in the decline in identification of faces as detail is removed (Bachmann, 1991; Costen et al., 1994). If the im-

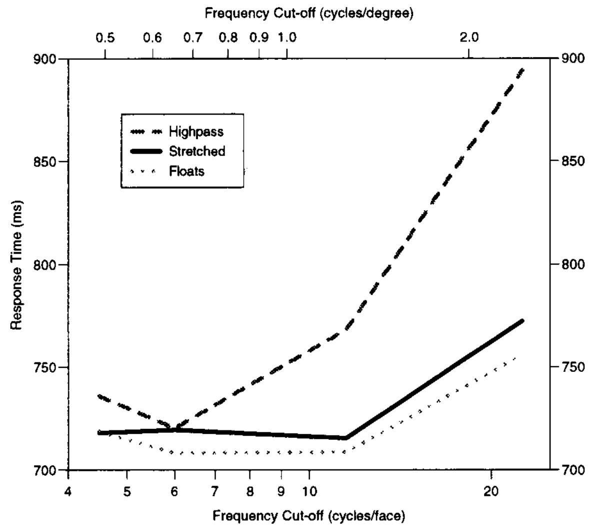

Figure 8. Response times in Experiment 2. The upper abscissa shows the frequency cut-off in cycles per degree. The lower abscissa shows the frequency cut-off in cycles per face. The ordinates show response time in milliseconds.

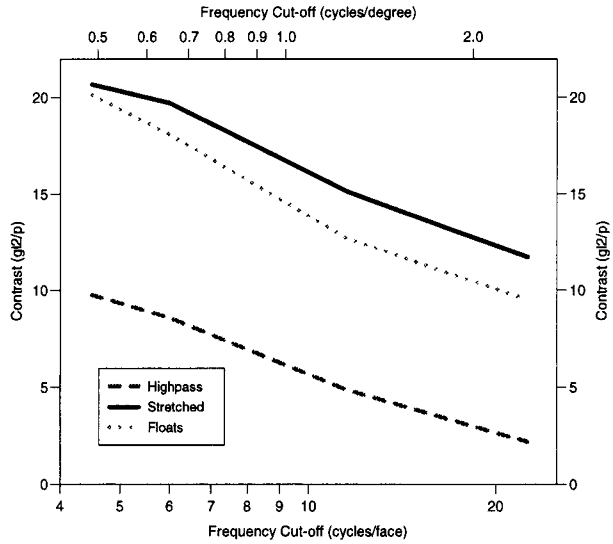

Figure 9. Image contrast in Experiment 2. The upper abscissa shows the frequency cut-off in cycles per degree. The lower abscissa shows the frequency cut-off in cycles per face. The ordinates show image contrast.

ages are degraded by quantization, or Fourier filtering, this occurs at a level between 11.5 and 6 cycles per face in Costen et al. (1994) and in Experiment 1 here. With the addition of the RT data, ceiling effects can be discounted in these investigations. The discontinuity is also clearly not due to the degradation method, since the low-pass condition, which does not introduce additional noise, shows a more marked discontinuity than does the pixel condition. The pixel condition also shows a general decrement in performance relative to the low-pass condition. This both replicates and reconciles the values from previous studies (Bachmann, 1991; Fiorentini et al., 1983), where worse identification was seen for quantized images than for smoothly blurred ones.

The data from the high-pass condition in Experiment 1 show a decrease in response accuracy and increase in RT as the frequency cut-off rises above 11.5 cycles per face. This result might have been explained by the relatively low energy or gray-level range seen in the high-pass images. Indeed, increasing the available contrast in the images did lessen the decline in accuracy and the increase in RTs in comparison with the high-pass condition in Experiment 2. Given the desire to discover the use of the spatial content of the images, it should be noted that manipulations such as equating the contrast of the filtered images are as unnatural a set of activities as, say, extracting line drawings. This will lead to one sort of information within the face predominating, and obscure the real interest, the use of different types of information within the natural frequency spectrum. However, using a range of different energy manipulations (as in the present experiments) should overcome this problem by allowing the tracing of the same underlying factor through the different manipulations.

The relative advantage for the 4.5-, 6-, and 11.5-cycles-per-face conditions increased for the enhanced-contrast conditions, with a particular increase in the float manipulation for the latter cut-off. Thus, the changes in energy present in the images do not provide an explanation of the results; there is no relation either in order or interaction or in mean identification level between the root mean square contrast of the images and the identification level. This evidence from both experiments clearly suggests that, despite changes in contrast in the different filtering conditions, the identification of these faces is preferentially supported by a band of intermediate-spatial-frequency information, which we suggest is located between approximately 8 and 16 cycles per face.

Both the differences between the low-pass and pixelized conditions in Experiment 1 and the differences between the float and stretched conditions in Experiment 2 can be explained by the effect of noise on such a critical band. If there is a critical band of spatial-frequency information of about 8–16 cycles per face, then the low-pass condition should show a particularly sharp decline as the upper limit of the available information moves below this region; the residual identification and longer processing times seen in the 4.5- and 6-cycles-per-face conditions may reflect pattern identification rather than face identification. If this explanation in terms of added noise is correct, pixel images should display the same accuracy pattern, but shifted so that equivalent levels of performance occur with higher preserved spatial frequencies. The difference arises because the lower portions of the critical region will be contaminated by the presence of edge noise even if the pixel size should allow transmission of the critical information.

A similar explanation can cover the high-pass conditions. As the minimum retained spatial frequency passes above this critical value of spatial frequencies, identification performance should again drop off. However, this decline may be accelerated either by major differences in the total contrast present in the image, as was the case in the high-pass condition, or by the addition of extra contrast by the enhancement procedure, as in the stretched condition. This enhanced-contrast manipulation was formed by taking the positive side of the high-passed images, adding a mean gray level, and then stretching the range to the maximum possible with the equipment (256 gray levels). The float manipulation took the positive and negative components of the image before adding the gray level. As a consequence, the level of distortion present in the images relative to the original (the positive and negative components) will be greater for the stretched manipulation than for the float manipulation. It is notable that, while there was a significant main effect difference between the float and stretched conditions in Experiment 2, they were significantly different at the 11.5-cycles-per-face condition in the center of the postulated critical region and not significantly different above and below that value.

This suggested band of 8–16 cycles per face, the limits of which are in agreement with that of Bachmann (1991) and Fiorentini et al. (1983), is higher than that assessed by Harmon (1973) and lower than that proposed by Hayes et al. (1986) and Tieger and Ganz (1979). The presence of these masking effects allows one to exclude explanations of the difference between these results and those of Hayes et al. (1986) based upon spatial range or image contrast. Not only is this proposed value, with a harmonic mean of 11.3 cycles per face (equating with a spatial frequency of 1.23 cpd in this case), approximately one octave below the optimal frequency for contrast sensitivity of 2.5–4.0 cycles per degree (Campbell & Green, 1965) but extraneous information in this band has a disproportionate effect upon identification for both high-pass and low-pass images, compared with information outside this range. This strongly suggests that these changes in accuracy do not reflect differences in spatial range, and reinforce the difference in results between these two studies. Hayes et al. (1986) used ideal filters (also used in Costen et al., 1994, obtaining results comparable with those here) which have a sudden change from spatial frequency components included in the image to those excluded, and thus introduce extra noise. A possible explanation of the discrepancy is that the tasks the subjects were asked to perform in these two studies were different. The base faces that the subjects processed in Hayes et al. (1986) were on display during the experiment, the images were shown for a relatively long period (contrast was nonzero for 3 sec), and no RTs were given. Interestingly, the optimal frequency value was not affected by negating the image, thus creating a frequency component impossible in a real face. As a consequence, it appears that Hayes et al. (1986) did not investigate face identification proper, but the ability to match filtered visual patterns to their base images.

An alternative explanation could involve point- or line-finding procedures operating independently of the spatial frequencies of the image and also independently of the features of a general face-identification system, and this can explain some of the data. The accelerating decline in identification performance of the low-pass and pixel conditions in Experiment 1 could reflect the range of scales at which useful information was present in these images, rather than the features of a general face-identification system. However, this explanation does not fit the data from the high-pass conditions very well: the critical information about the location of facial features will still be present in the high-pass images, obscured possibly by the lack of contrast. Under these conditions, one should expect relatively small changes in response accuracy but large changes in the time to identify images; however, the reverse was seen. In both experiments, response accuracy was affected much more than RT.

In addition, the enhanced-contrast conditions in Experiment 2 show that this explanation cannot be correct. If the relatively linear decline in identification accuracy seen in the high-pass condition reflected the failure of a point-finding algorithm under conditions of reduced contrast, then the contrast manipulation should remove, or very much reduce, the changes in identification seen as the spatial frequency cut-off increases. The continued presence of a discontinuity in identification strongly suggests that the decline reflects the selective use of information at this scale, rather than its ease of detection.

The finding that about 8 cycles per face is a critical lower region for encoding structural information about faces has an intuitive appeal. This wavelength (0.125 face width) is approximately half the distance (0.287 face width) from the side of the face to the center of the eye socket and also from the center of the eye socket to the nasion and so is suited to code a range of facial dimensions. However, this suggestion that information at this level of detail is useful in determining the configural information necessary for face identification does not explain why it is used in preference to other levels of detail—most notably, the accurate location of items as allowed by higher spatial frequencies. Perhaps the information outside this range is positively harmful to the identification process. This may happen if a major requirement in face identification is to determine the three-dimensional shape of the stimulus, as suggested when dealing with basic-level classification (Biederman, 1987; Hummel & Beiderman, 1992) and in identification via rotation to standard positions (Ullman, 1989). If this happens, neither the high- or the low-frequency information, measured relative to the object considered, will be of great use. The very low frequency information may suggest that there is a face present, but not whose face it is (Sergent, 1986). The highest spatial frequencies may also be of little use, since they cannot allow the extraction of information concerning the relative three-dimensional position of portions of the face.

These results do not require that faces be Fourier transformed during the identification process; rather, they offer an indication of the spatial scale at which critical information lies. An alternative algorithm that would capture these effects is the coding of configural information by the type of local surface distortion present in the image (Bruce, Coombes, & Richards, 1993). This calculates the ratio between directional, mean curvature and the amount of circular, Gaussian curvature within a given area. The area of the face with different ratios correlates with distinctiveness and the visibility of changes in facial shape. Alternatively, O'Toole, Abdi, Deffenbacher, and Valentin (1993) considered the accuracy of representation of faces by different ranges of a principal components spectrum, a set of orthogonal components upon which a group of images vary. The components are extracted in order of decreasing magnitude of variation; thus, as the rank number increases, they move from coding general shape to coding detail. A central band allowed the best discrimination of known from unknown faces, which again suggests that relative shape or texture is of maximum value at a medium scale.

If this suggestion, that face identification preferentially depends upon the presence of a band of spatial frequencies of about 8–16 cycles per face, is correct, comparable results should be obtained with spatial-frequency band-passed manipulations of constant width. Compatible results would also be obtained with faces transformed so that they are seen, for example, from different orientations or at different sizes. These questions are currently being addressed.

REFERENCES

BACHMANN, T. (1991). Identification of spatially quantised tachistoscopic images of faces: How many pixels does it take to carry identity? *European Journal of Cognitive Psychology*, **3**, 85-103.

BIEDERMAN, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, **94**, 115-146.

BRUCE, V., BURTON, A. M., & DENCH, N. (1994). What's distinctive about a distinctive face? *Quarterly Journal of Experimental Psychology*, **47A**, 119-141.

BRUCE, V., COOMBES, A., & RICHARDS, R. (1993). Describing the shape of faces using surface primitives. *Image & Vision Computing*, **11**, 353-363.

BRUYER, R. (1988). Some masked inconstancies about constancy. *Brain & Cognition*, **7**, 374-376.

CAMPBELL, F. W., & GREEN, D. C. (1965). Optical and retinal factors affecting visual acuity. *Journal of Physiology*, **181**, 576-593.

COSTEN, N. P., PARKER, D. M., & CRAW, I. (1994). Spatial content and spatial quantisation effects in face recognition. *Perception*, **23**, 129-146.

EDWARDS, R. E. (1967). *Fourier series: A modern introduction*. New York: Holt, Rinehart & Winston.

FIORENTINI, A., MAFFEI, L., & SANDINI, G. (1983). The role of high spatial frequencies in face perception. *Perception*, **12**, 195-201.

HARMON, L. D. (1973). The recognition of faces. *Scientific American*, **229**(5), 70-83.

HARMON, L. D., & JULESZ, B. (1973, June 15). Masking in visual recognition: Effects of two-dimensional noise. *Science*, **180**, 1194-1197.

HAYES, A., MORRONE, M. C., & BURR, D. C. (1986). Recognition of positive and negative bandpass-filtered images. *Perception*, **15**, 595-602.

HUMMEL, J. E., & BIEDERMAN, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review*, **99**, 480-517.

KEENAN, P. A., WITMAN, R., & PEPE, J. (1989). Hemispheric asymmetry in the processing of high and low spatial frequencies: A facial recognition task. *Brain & Cognition*, **11**, 229-237.

KEENAN, P. A., WITMAN, R., & PEPE, J. (1990). Hemispheric asymmetry in the processing of high and low spatial frequencies: A facial recognition task erratum. *Brain & Cognition*, **13**, 130.

LANDY, M. S., COHEN, Y., & SPERLING, G. (1984). HIPS: Image processing under UNIX. Software and applications. *Behavior Research Methods, Instruments, & Computers*, **16**, 199-216.

MICHELSON, A. A. (1962). *Studies in optics*. Chicago: University of Chicago Press. (Original work published 1927)

MORRONE, M. C., BURR, D. C., & ROSS, J. (1983). Added noise restores recognizability of coarse quantized images. *Nature*, **305**, 226-228.

MOSCOVITCH, M. (1988). Further analyses of masking functions in laterality studies of face-perception. *Brain & Cognition*, **7**, 377-380.

MOSCOVITCH, M., & RADZINS, M. (1987). Backward masking of lateralized faces by noise, pattern, and spatial frequency. *Brain & Cognition*, **6**, 72-90.

MOULDEN, B., KINGDOM, F., & GATLEY, L. F. (1990). The standard deviation of luminance as a metric for contrast in random-dot images. *Perception*, **19**, 79-101.

O'TOOLE, A. J., ABDI, H., DEFFENBACHER, K. A., & VALENTIN, D. (1993). Low-dimensional representation of faces in higher dimensions of the face space. *Journal of the Optical Society of America*, **10**, 405-411.

O'TOOLE, A. J., DEFFENBACHER, K. A., VALENTIN, D., & ABDI, H. (1994). Structural aspects of face recognition and the other-race effect. *Memory & Cognition*, **22**, 208-224.

PARKER, D. M. (1980). Simple reaction times to the onset, offset, and contrast reversal of sinusoidal grating stimuli. *Perception & Psychophysics*, **28**, 365-368.

SCHUCHARD, R. A., & RUBIN, G. S. (1989). Face identification of bandpass filtered faces by low vision observers [Abstract]. *Investigative Ophthalmology & Visual Science*, **30**(3, Suppl.), 396.

SERGENT, J. (1986). Microgenesis of face perception. In H. D. Ellis, M. A. Jeeves, F. Newcombe, & A. Young (Eds.), *Aspects of face processing* (pp. 17-73). Dordrecht: Martinus Nijhoff.

SHEPHERD, J. W. (1986). An interactive computer system for retrieving faces. In H. D. Ellis, M. A. Jeeves, F. Newcombe, & A. Young (Eds.), *Aspects of face processing* (pp. 398-409). Dordrecht: Martinus Nijhoff.

TIEGER, T., & GANZ, L. (1979). Recognition of faces in the presence of two-dimensional sinusoidal masks. *Perception & Psychophysics*, **26**, 163-167.

ULLMAN, S. (1989). Aligning pictorial descriptions: An approach to object recognition. *Cognition*, **32**, 193-254.

VOKEY, J. R., & READ, J. D. (1992). Familiarity, memorability, and the effect of typicality on the recognition of faces. *Memory & Cognition*, **20**, 291-302.

(Manuscript received February 9, 1994;
revision accepted for publication September 14, 1995.)