View metadata, citation and similar papers at core.ac.uk

brought to you by

provided by Elsevier - Publish

PERGAMON

Vision Research 39 (1999) 3824–3833

Vision Research

www.elsevier.com/locate/visres

Section 2

Spatial frequency bandwidth used in the recognition of facial images

Risto Näsänen

Brainwork Laboratory, Section of Clinical Neurosciences, Finnish Institute of Occupational Health, Topeliuksenkatu 41 aA, FIN-00250 Helsinki, Finland

Received 6 October 1998; received in revised form 6 April 1999

Abstract

The purpose of the study was to find out what spatial frequency information human observers use in the recognition of face images. Signal-to-noise ratio thresholds for the recognition of facial images were measured as a function of the centre spatial frequency of narrow-band additive spatial noise. The relative sensitivity of recognition to different spatial frequencies was derived from these results. The maximum sensitivity was found at 8–13 c/face width and the bandwidth was just under two octaves. Qualitatively similar results were obtained with stimuli in which Fourier phase was randomised within a narrow band of different centre spatial frequencies. This resulted in a considerable increase of energy threshold around 8 c/face width and less elsewhere. Further, contrast energy thresholds were measured as a function of the centre spatial frequency of band-pass filtered face images. As a function of object spatial frequency (c/face width), energy threshold first decreased and then increased. The lowest energy thresholds found around 10 c/face width were lower than the energy threshold for unfiltered images. This is what one would expect if face recognition is narrow-band, since band-pass filtered images of optimal centre spatial frequency do not contain unused contrast energy at low and high spatial frequencies. In conclusion, the results suggest that the recognition of facial images is tuned to a relatively narrow band (<2 octaves) of mid object spatial frequencies. © 1999 Elsevier Science Ltd. All rights reserved.

Keywords: Visual recognition; Spatial frequency; Bandwidth; Face; Noise

1. Introduction

Several studies suggest that the recognition of facial images only depends on a limited range of spatial frequencies. Fiorentini, Maffei and Sandini (1983) studied the recognition of low-pass and high-pass filtered face images. They found that the accuracy of recognition was worse for images only containing spatial frequencies below 5 c/face width than for images only containing spatial frequencies above 5 c/face width. Since they adjusted the viewing distances so that spatial frequencies above 15 c/face width were not visible, their result means that information between 5 and 15 c/face width is more useful for recognition than information below 5 c/face width.

Costen, Parker and Craw (1996) studied identification accuracy for faces low- or high-pass filtered with different cut-off frequencies. Their results suggested that face identification preferentially depends on spatial frequencies between 8 and 16 c/face width.

Hayes, Morrone and Burr (1986) studied the recognition of band-pass filtered face images shown from three different viewpoints. Hayes et al. used an ideal (sharp edged) 1.5 octave band-pass filter of various centre spatial frequencies. The results were expressed in the proportion of correct recognitions as a function of the centre spatial frequency of the band-pass filter. According to their results the band of spatial frequencies most useful for face recognition is located around 20 c/face width. This finding was independent of viewing distance. Therefore, the relevant dimension of spatial frequency for object recognition is cycles per object rather than cycles per degree of visual angle.

E-mail address: risto.nasanen@occuphealth.fi (R. Näsänen)

0042-6989/99/$ - see front matter © 1999 Elsevier Science Ltd. All rights reserved.
PII: S0042-6989(99)00096-6

Tieger and Ganz (1979) used a plaid masking technique to study the significance of different spatial frequencies in a recall task for face recognition. The images were superimposed with a plaid (vertical plus horizontal sinusoidal gratings) of different spatial frequencies. The results were described as the relative sensitivity of recognition to different spatial frequencies. They also derived an 'attentional filter' of the recognition system by taking into account the amplitude spectrum of the face images and the contrast sensitivity function of the visual system. Spatial frequency was expressed in cycles per degree but it can be concluded that the maximum masking effect and the filter maximum occurred at about 15 c/face width (2.2 c/deg). According to their results the retino-cortical spatial frequency transduction alone could not account for their findings, and they argued that at least one higher level linear filter stage is necessary. The limited range of plaid spatial frequencies used in their study unfortunately does not allow exact estimation of the bandwidth of the 'attentional filter'.

Peli, Lee, Trempe and Buzney (1994) studied the recognisability of low-pass filtered images of celebrities. They found that spatial frequencies at about 8 c/face height are critical in the sense that they are sufficient for correct recognition.

According to these studies, there is a spatial frequency range that has a larger weight in determining the face identity than other spatial frequencies. However, these studies do not tell us what the bandwidth for face recognition is. On the other hand, the above mentioned studies do not agree where this critical range is located; some of them suggest that it is around 10 c/face and others that it is closer to 20 c/face width. For the recognition of letters it has been shown by Solomon and Pelli (1994) that the spatial frequency bandwidth used is only about two octaves. They used low-pass and high-pass filtered noise masks of different cut-off frequencies. Assuming a linear relation between noise spectral density and the effect of the mask on the contrast energy threshold for recognition, it was possible to determine the 'filter function' used in this task. Braje, Tjan and Legge (1995) studied the recognition of low-pass filtered three dimensional objects in spatial noise and found that their results were qualitatively similar to the behaviour of a model consisting of a narrow band-pass filter followed by an ideal pattern classifier.

The purpose of the present study was to estimate quantitatively the bandwidth used in the recognition of facial photographs by measuring the relative sensitivity of the visual system to different spatial frequencies in this task. Four different experiments were run. In each of them, the stimuli were degraded in a spatial frequency specific way, which resulted in either a selective reduction or selective preservation of available stimulus information within narrow spatial frequency bands of different centre spatial frequencies. The first two experiments were specifically designed for determining the relative sensitivity (or gain) function used in the recognition of facial images. The third and fourth experiments were designed for further testing of the findings of the first two experiments.

In the first two experiments, the threshold signal-to-noise ratio for the recognition of facial images was determined as a function of the centre spatial frequency of narrow-band additive spatial noise. Narrow-band noise reduces available image information only at those spatial frequencies with which it overlaps. Therefore, its effect on threshold signal-to-noise ratio should correlate with the importance of that spatial frequency band to the recognition performance.

In the third experiment, the Fourier phase information of stimulus images was removed within a narrow band of spatial frequencies by replacing the phase coefficients by random numbers, while the amplitude spectrum was left unaltered. Contrast energy thresholds were measured as a function of the centre spatial frequency of the phase randomisation band. The adverse effect of phase randomisation should be largest close to the spatial frequencies that have the greatest contribution to recognition.

In the fourth experiment, contrast energy thresholds were measured as a function the centre spatial frequency of images band-pass filtered with a two octave Gaussian Fourier filter. Lowest energy thresholds were expected to occur at spatial frequencies optimal for recognition. The results of all these experiments showed a clear band-pass nature of the recognition of facial images.

2. Methods

2.1. Equipment

The stimuli were generated by using a PC computer with a 200 MHz AMD K6 processor and a 15 in. monitor (Nokia 449Xi). The graphics board (Hercules Dynamite) was used at a resolution of 800×600 pixels and a frame rate of 90 Hz. The pixel size of the display was 0.034×0.034 cm^2, and the average photopic luminance of the stimuli and background was 53 cd/m^2. The non-linearity of the luminance response of the display was corrected (gamma correction) by using its inverse function when the stimuli were computed. The measurements were done in a dim room, where the only light source was the monitor.

The graphics board could produce 256 grey levels. It is well known that this number of grey levels is not always sufficient for presenting images at threshold contrast. In recognition studies this is not as large a

problem as in detection studies, since contrast thresholds for recognition are much higher than for contrast detection. In this study, the Michelson contrast thresholds varied between 0.024 and 0.15. Contrast detection thresholds for gratings can be one tenth of the lowest thresholds measured here. However, at low contrasts the low number of grey levels produces quantization errors in the signal. In this study the low contrast information was increased and the effects of quantization errors on the information contents of the displayed images were reduced using a quasi-periodic dithering technique, which utilises the Bayer (1973) dither matrix. The dither in the displayed images was completely invisible in all stimulus conditions. At the Michelson contrast of 0.024, the RMS contrast of the error signal (the intended signal minus the actual signal) related to dither was 0.0032. At this contrast level the number of true grey levels was five. The amplitude spectrum of the error signal for these broad band stimuli was relatively flat across spatial frequencies up to at least 100 c/image. Assuming that the error signal is white noise, its spectral density would be $N = 0.0106 \times 10^{-6}$ deg^2 for a viewing distance of 60 cm. At the contrast of 0.024 the energy (E) signal-to-noise ratio ($s/n = \sqrt{(E/N)}$) for the synthetic face images used in experiment two would then be $s/n \approx 511$. The dithering technique has been explained in detail in Näsänen and O'Leary (1998).

2.2. Facial images

2.2.1. Faces with variable pose, expression, and lighting

In the first, third, and fourth experiment facial photographs of twenty females, one photograph per person, were used. The faces were not previously known to the two subjects of the study; the photographs were taken by a third person. The digitised images were cropped so that the image width was approximately the same as the face width. Otherwise, the images were let vary freely in pose, facial expression, hair style, and lighting direction. The poses varied from nearly frontal to the so called three quarters views. Some faces were tilted towards one side. Therefore, these facial images contained plenty of non-facial, photograph specific, features. The amplitude spectra of the images obeyed a power function with an exponent of -1.41 on average. The size of the digitised images was 200×200 pixels (6.72×6.72 cm^2). At the viewing distance of 60 cm used in the experiments, this corresponds to a size of 6.4×6.4 deg^2. Two examples of the images with various spatial frequency manipulations (explained later) are show in Fig. 1.

2.2.2. Synthetic faces

In the second experiment, the images were synthetic faces, where the above non-facial variables (pose, expression, and lighting) were tried to keep as constant as possible. This set of images is shown in Fig. 2. The faces in Fig. 2B–H, used in experiment two, have been obtained from the face in Fig. 2A by image warping. Therefore, the resulting images differ mainly in shape but have a similar texture. In image warping, correspondence maps between the face in Fig. 2A and seven other facial photographs were used. The photographs were chosen so that the poses were highly similar and the facial expressions were neutral. The faces in the original images were not known to the subjects of this study.

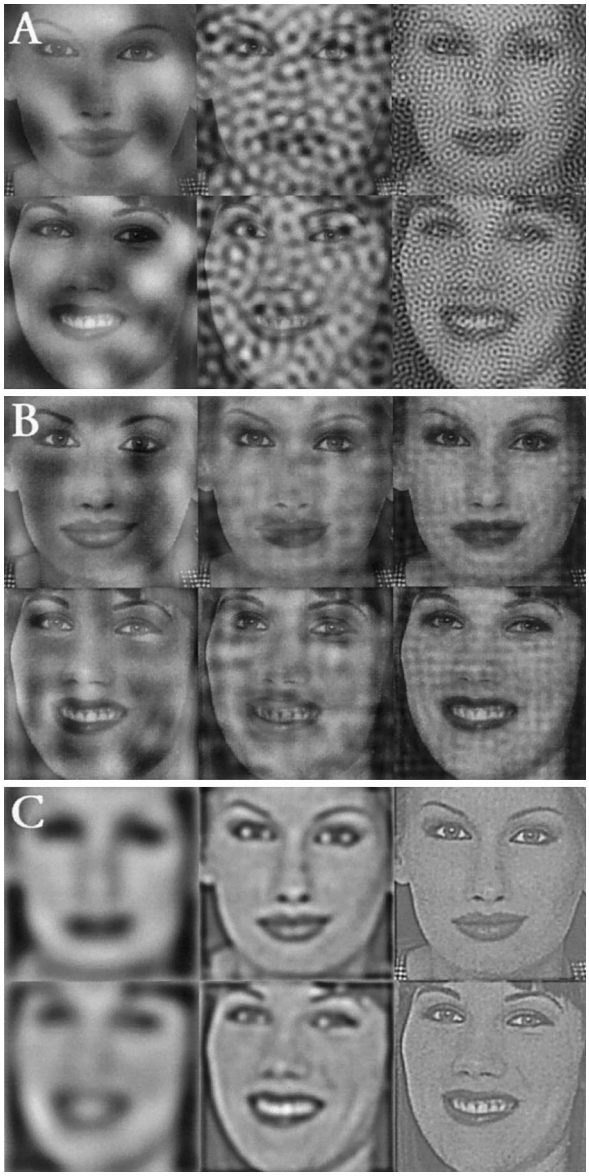

Fig. 1. Examples of stimulus manipulations used in the experiments of Figs. 3, 5 and 6. (A) Images with additive narrow-band spatial noise of different centre spatial frequencies of the noise band. In the above example the centre spatial frequencies of the noise band are 2, 11, and 32 c/face width from left to right. (B) Images with a narrow phase randomisation band. The centre spatial frequencies of the phase randomisation band in the above examples are 3.3, 11, and 16 c/face width. (C) Band pass filtered images with centre spatial frequencies of 3.7, 9.7, and 29 c/face width.

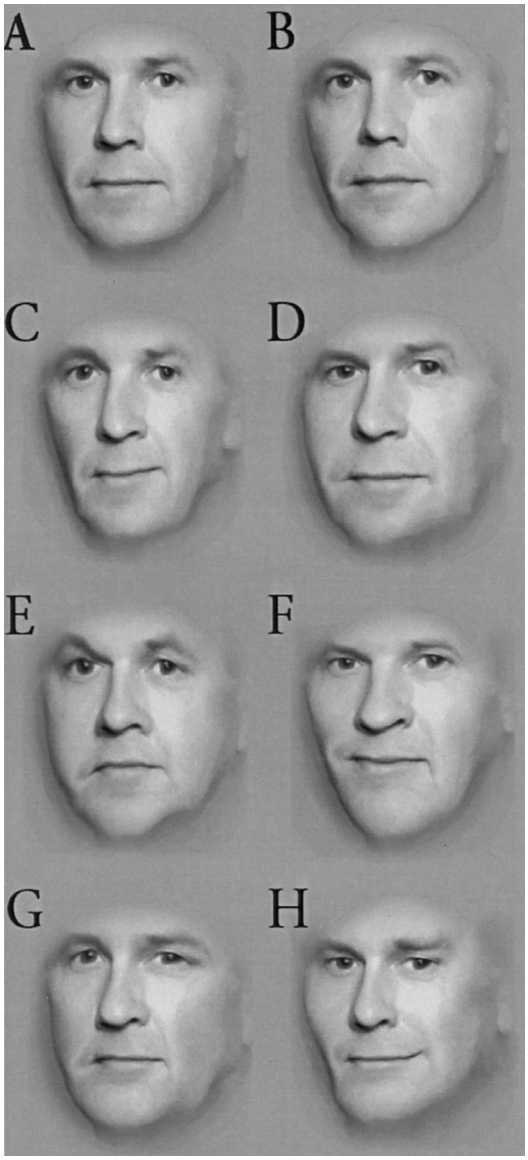

Fig. 2. Synthetic faces. The faces from B to H, used in the experiment of Fig. 4, have been obtained from face A by image warping. See the text for details.

The correspondence maps were obtained by using a multi-resolution algorithm, which searches for correspondence points between two face images. The correspondence points are expressed in displacements in horizontal and vertical directions $(\Delta x, \Delta y)$ for each pixel. The criterion for similarity between points in the two images was the Euclidean distance computed over 3×3 pixel areas in both images. Band-pass filtered (two-octave Gaussian band-pass filter) images of different resolutions were used. The image sizes (pixels) and filter centre spatial frequencies (c/image) were 32×32 pixels and 10 c/image, 64×64 pixels and 20 c/image, 128×128 pixels and 40 c/image, and 256×256 pixels and 80 c/image. The search at the lowest resolution gives a rough correspondence map, which is refined by searches at higher resolutions. The maps were smoothed always after the search at each resolution using median filtering.

The mean slope of the amplitude spectrum of the synthetic faces was -1.86. The size of the displayed images was 300×300 pixels, which corresponds to 10×10 cm^2 on the screen, and 9.5×9.5 deg^2 at the viewing distance of 60 cm, and 2.4×2.4 deg^2 at the viewing distance of 240 cm.

2.3. Spatial frequency manipulations

2.3.1. Narrow-band noise masks

In the first two experiments, narrow-band additive spatial Gaussian noise was used. The bandwidth of noise was 4 c/image width. The centre spatial frequency of the noise band was 2, 2.8, 4, 5.6, 8, 11, 16, 23, 32, or 45 c/image width. Narrow-band noise was obtained by filtering white Gaussian noise. For each centre spatial frequency one hundred different noise images were generated. For each contrast level of each face image, one of these noise images was chosen at random. The final stimulus was the sum of the face image and the chosen noise image. For filtering noise an 'ideal' (sharp edged) Fourier band-pass filter was used. Therefore, the noise was white within the band and had zero power elsewhere. The maximum contrast of the noise waveforms was always equal to 0.2. The RMS contrast (c_{RMS}) of noise, which was computed from the filtered noise samples, varied from 0.082 to 0.046 when the centre spatial frequency increased from 2 to 45 c/image. The noise spectral density (N) was computed as

$$N = c_{\mathrm{RMS}}^2/[\pi(f_2^2 - f_1^2)] \quad (1)$$

where f_1 and f_2 are the lower and higher cut-off frequencies of the noise band, respectively. In Eq. (1), term $[\pi(f_2^2 - f_1^2)]$ represents the area of the annulus in the spatial frequency domain to which the mean square contrast (c_{RMS}^2) of noise is evenly distributed. Examples of stimulus images with band-pass noise are shown in Fig. 1A.

Thresholds were expressed in signal-to-noise ratio (s/n), which is defined as the square root of the ratio of contrast energy threshold (E) and spectral density of noise (N)

$$s/n = \sqrt{(E/N)} \quad (2)$$

Contrast energy was computed as follows

$$E = p^2 \Sigma_x \Sigma_y c^2(x, y) \quad (3)$$

where $c(x, y)$ is the contrast waveform, and p^2 is the pixel area. Contrast waveform is defined as

$$c(x, y) = (l(x, y) - l_{\mathrm{o}})/l_{\mathrm{o}} \quad (4)$$

where $l(x, y)$ is the luminance waveform and l_{o} is the mean luminance (Legge, Kersten & Burgess, 1987).

2.3.2. Fourier phase randomisation

In the third experiment, the stimuli were images in which the phase spectrum of a narrow band of spatial frequencies was replaced by random numbers drawn from an even distribution with a range of 360°. Examples of this kind of manipulations are shown in Fig. 1B. Outside this spatial frequency band the phase spectrum was unaltered. The amplitude spectrum was the same as in the original images. Phase spectrum is essential for the shape information of images (Oppenheim & Lim, 1981; Piotrowski & Campbell, 1982). Therefore, randomising phase spectrum destroys the original shape information within the spatial frequency band in question. The width of the phase randomisation band was 5.3 c/face width. The centre spatial frequencies of the phase randomisation band were 3.32, 4.80, 6.20, 8.43, 11.3, 16.2, and 32.1 c/face width. The phase randomisation was different for each face but did not vary with contrast, i.e. each face always had the same phase randomisation. It was expected that the contrast energy within the phase randomisation band would not support recognition. Therefore, recognition at threshold would have to rely on spatial frequency information at other spatial frequencies. The largest loss of information should occur when the phase randomisation band is close to the centre of the spatial frequency band used for recognition. To compensate for the loss of shape information a spatial frequency specific increase in contrast energy would be required at recognition threshold.

2.3.3. Band-pass filtering

In the fourth experiment, the stimuli were band-pass filtered with a two octave (full bandwidth at half height) Fourier band-pass filter. A Gaussian filter ($H(f) = \exp(-(f-f_o)^2/b^2)$) of different centre spatial frequencies ($f_o = 4$, 5.6, 8, 11, 16, 23, or 32 c/face width) was used. Examples of these are shown in Fig. 1C. The centre spatial frequencies of the filtered images were computed using the following formula (Parish & Sperling, 1991):

$$f_c = [\Sigma_u \Sigma_v f|F(u,v)|^2]/[\Sigma_u \Sigma_v |F(u,v)|^2] \quad (5)$$

where $|F(u,v)|$ is the Fourier amplitude spectrum of the filtered image, $f = \sqrt{(u^2+v^2)}$, and u and v are spatial frequencies in the horizontal and vertical directions, respectively. The average centre spatial frequencies were found to be 3.66, 4.64, 6.58, 9.65, 14.1, 20.9, and 29.4 c/face width.

If only a limited band of spatial frequencies contributes to face recognition, the unfiltered broad band facial images will contain unused contrast energy at high and low object spatial frequencies. Band-pass filtering at optimal centre spatial frequency will reduce this unused contrast energy. Therefore, at recognition threshold the amount of contrast energy needed should be smaller for band-pass filtered images of optimal centre spatial frequency than for unfiltered images.

2.4. Procedure

Contrast energy thresholds were determined using a multiple-alternative forced-choice method. In the first, third, and fourth experiment there were 20 alternatives, and in the second experiment there were seven alternatives. The stimuli were presented for 1000 ms in the centre of the screen. The fixation target was a small graphical cross, which was switched off during the stimulus presentation. The task of the observer was to indicate which one of the faces was shown. Close to the left-hand edge of the screen, there was an array of graphical buttons, one button for each face. In experiments one, three, and four, the buttons were marked by icons, which were the original full contrast faces presented at a resolution of 50×50 pixels. In experiment two, the buttons were marked by numerals from 1 to 7.

To indicate her/his response, the observer pointed and clicked one of the buttons with mouse. This required that the observer first moved fixation from the fixation target to the button array, placed the mouse cursor on the appropriate button, moved fixation back to the fixation target and then pressed the mouse button. The response started a new presentation after a delay of 500 ms. The presentation of the stimulus was indicated by a sound signal. Another sound signal gave feedback about the correctness of the choice of the observer.

After four consecutive correct responses the signal contrast was decreased by a factor of 1.26, and after each incorrect response the contrast was increased by the same factor. A threshold estimate at the probability level of 0.84 of correct answers (Wetherill & Levitt, 1965) was obtained as the mean of eight reversals. The number of trials needed for one threshold estimate was 60 on average. Each data point shown in Figs. 3–6 represents the arithmetic mean of three threshold estimates.

Before the experiments, there was a training phase. This served two purposes. Firstly, good performance in the task required that the subjects learned well the appearance of each face and which button corresponded to which face. Secondly, the training phase was necessary for avoiding performance improvement during the final experiments. The task of the observer was the same as in the final experiments except that, at first, contrast was kept constant. As in the final experiments, the observers were given feedback about the correctness of their responses. The face stimuli used were the unmanipulated faces. After the subject felt that she/he could perform the task smoothly, a few contrast threshold measurements were done before the final experiments. No objective criterion was used to determine when the training was complete. However, no improvement of performance was found during the final experiments, which suggests that the subjective

criterion was adequate. The training took place during 3 consecutive days preceding the final experiments.

2.5. Subjects

Two persons served as observers; one of them was the author. Both had normal or corrected to normal vision. Subject RN had extensive experience as an observer in psychophysical experiments and VN only had very limited previous experience.

3. Results

3.1. Masking by narrow-band noise

In the first experiment, contrast energy thresholds were measured as a function of the centre spatial frequency of a narrow band of additive spatial noise. Facial images with varying pose, expression, hair style, and lighting across faces (persons) were used in this experiment. Fig. 1A shows examples of these images. The results shown in Fig. 3A are expressed in signal-to-noise ratio (s/n) thresholds. For both subjects, the signal-to-noise ratio threshold first increased and then decreased as the centre object spatial frequency of the noise band increased. The maximum signal-to-noise ratio threshold occurred at 11 c/face width.

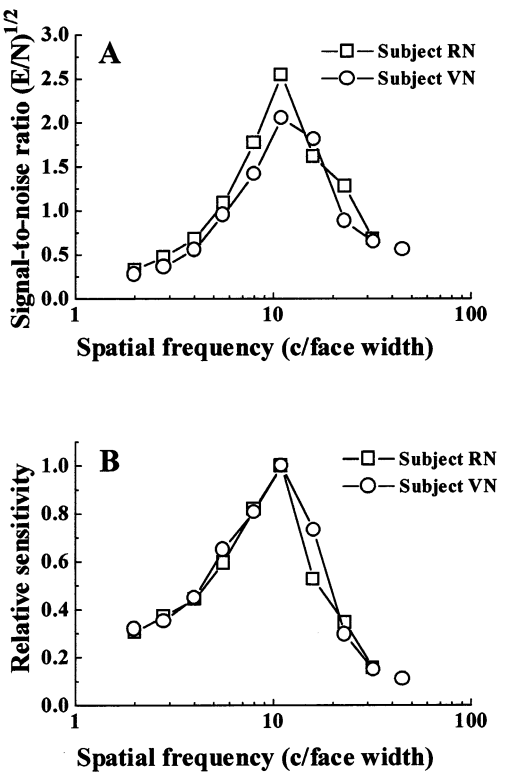

Fig. 3. Signal-to-noise ratio thresholds for face recognition as a function of noise centre spatial frequency (A), and estimated relative sensitivity of face recognition as a function of object spatial frequency (c/face width) (B). The relative sensitivity was calculated by using Eq. (6). The viewing distance was 60 cm. The spatial frequencies expressed in c/deg were 0.31, 0.44, 0.62, 0.87, 1.2, 1.7, 2.5, 3.6, and 5.0. The dashed line in A and B show respectively the performance and relative sensitivity of the white noise ideal observer.

From the signal-to-noise ratio results it is possible to estimate the relative sensitivity of the pattern recognition mechanism to different object spatial frequencies. It was assumed that the increase of contrast energy threshold (ΔE) is directly proportional to the total power of filtered noise obtained by integrating the product of noise spectral density and the square of the sensitivity function across spatial frequencies (Solomon & Pelli, 1994). The relative sensitivity function $S(f_o)$ is calculated by using Eq. (6) (see Appendix A for derivation).

$$S(f_o) \approx \{\Delta E/[a\pi N(f_o)(f_2^2 - f_1^2)]\}^{0.5} \quad (6)$$

where f_o is the centre spatial frequency, f_2 and f_1 are the upper and lower cut-off frequencies, $N(f_o)$ is the spectral density of the noise band, and a is a proportionality constant.

Fig. 3B shows the estimated relative sensitivity (gain) functions for face recognition normalised so that the maximum sensitivity is equal to unity. The estimated relative sensitivity functions peaked at 11 c/face width, and their full bandwidths at half height were just under 2 octaves for both subjects. The above results suggest that in this task most of the contrast information is collected from a two octave band around 11 c/face width.

Since in the first experiment there was only one face per person and the faces varied in pose, expression, hair style, and lighting, there were photograph specific and non-facial cues available. To find out whether the bandwidth is similar when such cues are not present, these variables were kept constant across images in the second experiment. The image set used in the second experiment were synthetic faces (Fig. 2) generated from a single face image by image warping (geometric transformation). Therefore, the resulted images differed mainly in shape.

In the first experiment, the response buttons contained the faces in a smaller size. Therefore, one might think that the first experiment is some sort of image matching task rather than an identification task. This could not be the case in the second experiment, since the response buttons were marked by numerals.

Fig. 4A and B show, respectively, the signal-to-noise ratio thresholds and estimated relative sensitivity functions for the synthetic faces at the viewing distance of 60 cm. The results were highly similar to those in the first experiment (Fig. 3A and B). The peak sensitivity (in Fig. 4B) occurred between 8 and 13 c/face width and the bandwidth was about two octaves.

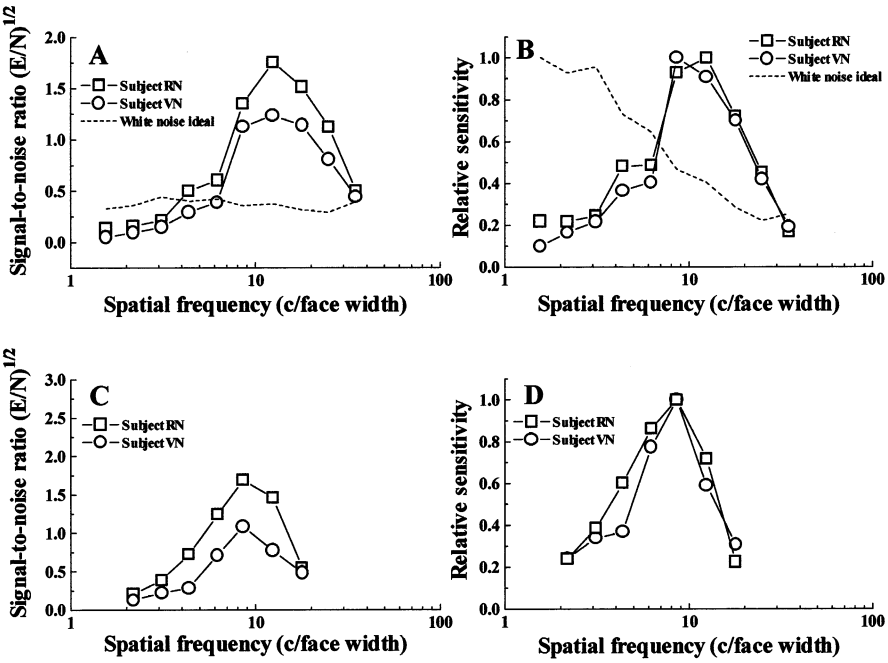

Fig. 4. Signal-to-noise ratio thresholds and estimated sensitivity functions for recognition of the synthetic face images shown in Fig. 2. In A and B the viewing distance was 60 cm, and in C and D it was 240 cm. Therefore, in c/deg the spatial frequencies were higher by a factor of four in C and D (11 c/face corresponds to 6.9 c/deg) than those in A and B (11 c/face corresponds to 1.71 c/deg).

Do these results reflect the properties of human perception or the inherent characteristics of the visual task and stimuli? To find the answer to this question human performance was compared with the performance of a white noise ideal observer in the narrow band noise masking experiment (Fig. 4A and B). An ideal observer uses all available information. Therefore, the use of different spatial frequencies by the ideal observer reflects the characteristics of the task and stimuli. The white noise ideal observer for this particular task computes the Euclidean distance ($D = \Sigma_x \Sigma_y [s(x, y) - m_i(x, y)]^2$) between received signal ($s(x, y)$) and a set of templates ($m_i(x, y)$). The templates are identical copies of the face stimuli. To identify the face, the ideal observer searches for the shortest Euclidean distance. The performance of the ideal observer was estimated by computer simulations using the same threshold estimation algorithm as in the experiments with human observers.

The signal-to-noise ratio performance of the white noise ideal observer, shown by the dashed line in Fig. 4A, is nearly independent of the centre spatial frequency of the noise band and, therefore, is very different from human performance. In Fig. 4B, the relative sensitivity (Eq. (6)) of the white noise ideal observer (the dashed line) decreases with spatial frequency while human relative sensitivity exhibits band-pass behaviour. A similar finding was made by Solomon and Pelli (1994) for letter recognition. The ideal observer simulations, therefore, suggest that the band-pass behaviour reflects the properties of human perception, not the inherent characteristic of the identification task.

Fig. 4C and D show the experiment of Fig. 4A and B repeated at a viewing distance of 240 cm, four times the distance used in the experiment of Fig. 4A and B. Again the signal-to-noise ratio thresholds (Fig. 4C) and estimated relative sensitivity functions (Fig. 4D) had a limited bandwidth. The only differences are that the peak sensitivity point (in Fig. 4D) has moved to a slightly lower object spatial frequency (about 8 c/face width), and the high spatial frequency fall-off is steeper than for the shorter viewing distance (in Fig. 4B). This is probably due to the increased attenuation of high spatial frequencies by the optics of the eye. For the viewing distance of 240 cm the absolute spatial frequencies expressed in c/deg are four times higher (11 c/face corresponds to 6.9 c/deg) than for the shorter viewing distance (11 c/face corresponds to 1.71 c/deg). Since the shift of the relative sensitivity function in terms of object spatial frequency (cycles per face width) is small in comparison to the difference in absolute spatial frequency values, the use of object spatial frequency instead of absolute spatial frequency (c/deg) seems to be appropriate. This is in agreement with the findings of Hayes et al. (1986).

3.2. Effect of randomising Fourier phase

In the third experiment, contrast energy thresholds were measured as a function of the centre spatial

frequency of a band in which Fourier phase information was randomised. In this experiment, the first set of images with variable pose, expression, and lighting was used (see Fig. 1B for examples of the stimuli).

The results of the third experiment are shown in Fig. 5. As expected, there was a clear spatial frequency specific increase in contrast energy threshold. The highest energy threshold was obtained at 8 c/face width. At low and high centre spatial frequencies, there was only a small increase in energy threshold relative to the energy threshold for the original unprocessed images shown by the solid horizontal line. The results were rather similar for the two subjects except that the energy thresholds were systematically higher for subject RN than for subject VN. The results of the third experiment are in agreement with the prediction based on the first two experiments.

3.3. Recognition of band-pass filtered faces

In the fourth experiment, contrast energy thresholds for recognition were measured as a function of the centre object spatial frequency of band-pass filtered face images (examples shown in Fig. 1C). The half height bandwidths of the filtered images were between one and two octaves.

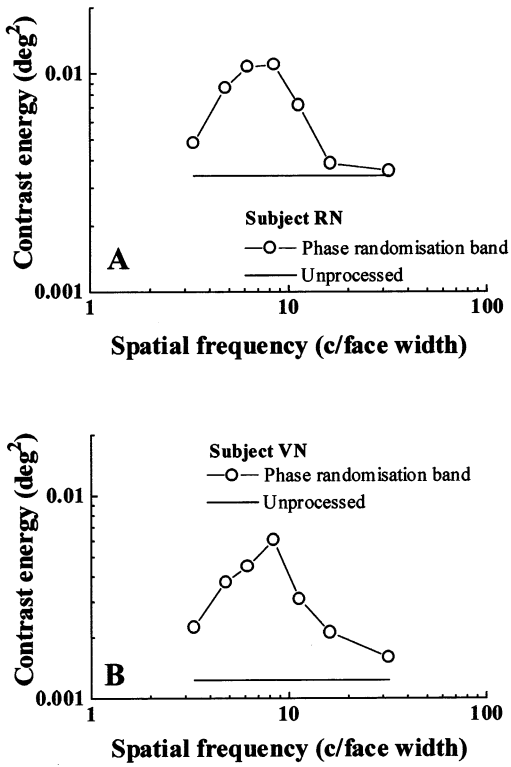

Fig. 5. Energy threshold for the recognition of facial images as a function of the centre spatial frequency of the band of Fourier phase randomisation. The horizontal lines show the thresholds for unprocessed images.

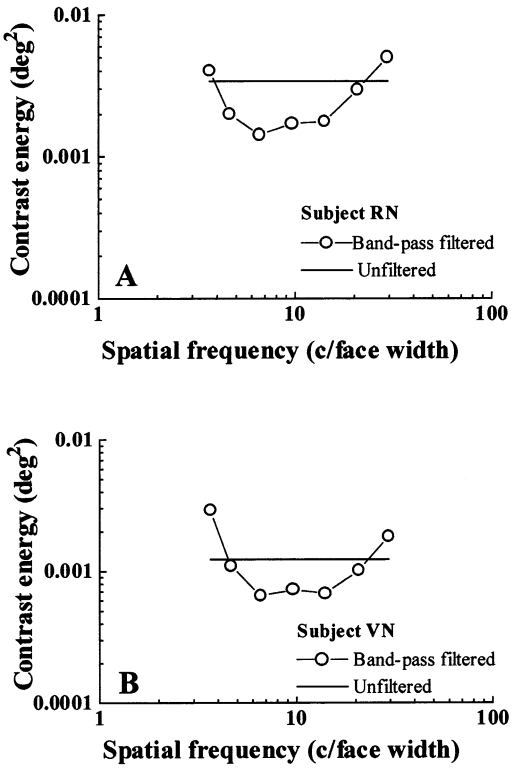

Fig. 6. Energy threshold for face recognition as a function of the centre spatial frequency of band-pass filtered facial images. The bandwidth of the filtered images was about 1–2 octaves. The horizontal lines show the thresholds for unfiltered images.

The results of the fourth experiment are shown in Fig. 6. Smallest contrast energy thresholds were obtained at centre object spatial frequencies of 6–14 c/face width. At these object spatial frequencies, contrast energy thresholds were clearly lower than for unfiltered images. The result, therefore, provides additional evidence supporting the claim that only a narrow band of spatial frequencies contributes to recognition of facial images. At the lowest and highest object spatial frequencies, energy thresholds are higher than for unfiltered images showing that the recognition of facial images is relatively insensitive to these object spatial frequencies. The results for the two subjects are rather similar except that the contrast energy thresholds for subject RN are higher than for VN.

In the experiments above, the average standard errors were about 10 and 20% of the mean for subjects RN and VN, respectively.

4. Discussion

Masking by narrow-band spatial noise resulted in a spatial frequency specific increase in threshold signal-to-noise ratio. The bandwidth estimate for the recogni-

tion of facial images was slightly less than two octaves (at half height) and the peak spatial frequency of the band was 8–13 c/face width. This finding was independent of whether the facial images of different persons varied in pose, expression, and lighting (Fig. 3) or not (Fig. 4). Computer simulations of the white noise ideal observer, which uses all available information, showed that a very broad range of spatial frequencies contain information useful for recognition of face images. Therefore, the band-pass behaviour found with human observers does not seem to be based on the inherent characteristics of the recognition task, but is a property of human visual perception.

As a function of the centre spatial frequency of the phase randomisation band, energy threshold first increased and then decreased. Phase randomisation had the most adverse effect when the centre spatial frequency was 8 c/face width. Phase randomisation selectively destroys shape information within the phase randomisation band. When the phase randomisation band is close to the centre of the spatial frequency band used for recognition, the observer has to use information at spatial frequencies for which recognition is not optimally sensitive. This results in an increase in contrast energy threshold. The results of the phase randomisation experiment are, therefore, in qualitative agreement with the findings of the first two experiments.

In a further experiment it was shown that the energy threshold for band-pass filtered images is lower than for unfiltered images when the centre spatial frequency of the images is close to the mid spatial frequencies (6–14 c/face width). This is exactly what is expected if low and high object spatial frequencies have little contribution to face recognition as the first two experiments suggested.

It should be noted that although at half height the bandwidth of the sensitivity function is slightly less than 2 octaves, the tails of the sensitivity function are rather long, and, therefore, a wide range of spatial frequencies can have at least some contribution to the recognition of facial images.

4.1. Comparison with other studies

The less than 2 octave bandwidth for the recognition of facial images estimated in the present study is similar to the bandwidth estimate for letter recognition by Solomon and Pelli (1994), who used low-pass and high-pass filtered noise masks. However, the centre object spatial frequency for letter recognition (3 c/letter) is much lower than for face recognition. The higher centre frequency for the recognition of faces suggests that adequate analysis of the shape of facial features, such as mouth, nose and eyes, requires the use of a smaller scale relative to the size of the whole object.

The present estimate of the centre spatial frequency (8–13 c/face width) is lower than the finding of Hayes et al. (1986) (20 c/face width), who measured the recognisability of band-pass filtered faces. The study of Tieger and Ganz (1979), who used plaid masks (vertical plus horizontal sinusoidal gratings) of different spatial frequencies, suggested a somewhat higher critical spatial frequency than what was found here. The maximum masking effect occurred at about 15 c/face width. The present estimate is in good agreement with Costen et al. (1996), who used low-pass and high-pass filtered faces in a face identification study. Their estimate of the critical spatial frequency range was from 8 to 16 cycles/face. The study of Fiorentini et al. (1983) showed that spatial frequencies between 5 and 15 c/ face width were more useful for recognition than spatial frequencies below 5 c/face width. Exact comparison of these studies is not meaningful because of the relatively large differences in the stimuli and methods used. However, all of these studies agree that the mid spatial frequency range is the most important one. What is new in the results of the present study in comparison to the studies cited above is that it provides a clear quantitative bandwidth estimate for the recognition of facial images.

The use of a narrow band of spatial frequencies may be a general property of human object recognition since it seems to apply to the recognition of different kinds of objects, not only faces, but also letters (Solomon & Pelli, 1994), and, as suggested by the study of Braje et al. (1995), three dimensional geometrical objects as well. Band-pass filtering, therefore, may be an initial stage of feature analysis in human object recognition.

4.2. Off-frequency looking

The use of narrow-band noise does not allow the evaluation of the possible effect of off-frequency looking (e.g. Solomon & Pelli, 1994). By off-frequency looking it is meant that, in the presence of band limited noise, the observer could use spatial frequencies that are outside the noise band and for which the signal-to-noise ratio is better. Solomon and Pelli (1994) used low-pass and high-pass filtered noise of various cut-off frequencies. The effects of off-frequency looking should be different for the low-pass and high-pass noise conditions. However, the sensitivity functions derived from these two conditions were highly similar. Therefore, off-frequency looking did not have any essential effects in their study. For narrow-band noise, off-frequency looking would broaden the estimated bandwidth. If there were any off-frequency looking effects in the present results, the true bandwidth would be even narrower than estimated here.

5. Conclusion

The present results suggest that in the recognition of previously learned facial images most of the information is collected from a spatial frequency band that is just under two octaves wide and centred around 8–13 c/face width.

Acknowledgements

This study was supported by Otto A. Malm Foundation.

Appendix A. Derivation of the relative sensitivity function for recognition in narrow-band noise

It is assumed that the increase of contrast energy threshold is directly proportional to noise spectral density (e.g. Burgess, Wagner, Jennings & Barlow, 1981). Specifically, at each spatial frequency the effect of noise on the increase of energy threshold is assumed to be directly proportional to the product of noise spectral density and the square of the sensitivity of the pattern recognition system to that spatial frequency. The total effect is obtained by integrating this product across spatial frequencies. The method used by Solomon and Pelli (1994) is based on the same assumption. The equation for the increase of energy threshold ($\Delta E = E - E_0$, where E and E_o are thresholds with and without the noise, respectively) can be written in the following way:

$$\Delta E = a2\pi N(f_0)\int_{f_1}^{f_2} fS^2(f)\,\mathrm{d}f \quad \text{(A1)}$$

where a is a proportionality constant, f is radial spatial frequency ($f=\sqrt{(f_x^2+f_y^2)}$, where f_x and f_y are spatial frequencies in the horizontal and vertical directions, respectively), f_1 and f_2 are respectively the lower and higher cut-off frequencies, and f_o is the centre spatial frequency of the noise band, $N(f_o)$ is the noise spectral density, which is constant between f_1 and f_2 and 0 elsewhere, and $S(f)$ is the sensitivity function.

Since the bandwidth of noise is narrow, we may assume that $S(f)$ is nearly constant within that band. Then the equation becomes

$$\Delta E \approx a2\pi N(f_0)S^2(f_0)\int_{f_1}^{f_2} f\,\mathrm{d}f \quad \text{(A2)}$$

from which we get

$$\Delta E \approx a\pi N(f_0)S^2(f_0)(f_2^2-f_1^2) \quad \text{(A3)}$$

From this the sensitivity function can be solved as

$$S(f_o) \approx \{\Delta E/[a\pi N(f_o)(f_2^2-f_1^2)]\}^{0.5} \quad \text{(A4)}$$

References

Bayer, B. E. (1973). An optimum method for two-level rendition of continuous-tone pictures. In *IEEE International Conference on Communication*, *Conference Record* (pp. 11–15), Session 26.

Braje, W. L., Tjan, B. S., & Legge, G. E. (1995). Human efficiency for recognizing and detecting low-pass filtered objects. *Vision Research*, *32*, 2955–2966.

Burgess, A. E., Wagner, R. F., Jennings, R. J., & Barlow, H. B. (1981). Efficiency of human visual signal discrimination. *Science*, *214*, 93–94.

Costen, N. P., Parker, D. M., & Craw, I. (1996). Effects of high-pass and low-pass spatial filtering on face identification. *Perception and Psychophysics*, *58*, 602–612.

Fiorentini, A., Maffei, L., & Sandini, G. (1983). The role of high spatial frequencies in face perception. *Perception*, *12*, 195–201.

Hayes, T., Morrone, M. C., & Burr, D. C. (1986). Recognition of positive and negative bandpass-filtered images. *Perception*, *15*, 595–602.

Legge, G. E., Kersten, D., & Burgess, A. E. (1987). Contrast discrimination in noise. *Journal of the Optical Society of America A*, *4*, 391–404.

Näsänen, R., & O'Leary, C. (1998). Recognition of band-pass filtered hand-written numerals in foveal and peripheral vision. *Vision Research*, *38*, 3691–3701.

Oppenheim, A. V., & Lim, J. S. (1981). The importance of phase in signals. *Proceedings of the IEEE*, *69*, 529–541.

Parish, D. H., & Sperling, G. (1991). Object spatial frequencies, retinal spatial frequencies, and the efficiency of letter discrimination. *Vision Research*, *31*, 1399–1415.

Peli, E., Lee, E., Trempe, C. L., & Buzney, S. (1994). Image enhancement for the visually impaired: the effects of enhancement on face recognition. *Journal of the Optical Society of America A*, *11*, 1929–1939.

Piotrowski, L. N., & Campbell, F. W. (1982). A demonstration of the visual importance and flexibility of spatial-frequency amplitude and phase. *Perception*, *11*, 337–346.

Solomon, J. A., & Pelli, D. G. (1994). The visual filter mediating letter identification. *Nature*, *369*, 395–397.

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

Wetherill, G. B., & Levitt, H. (1965). Sequential estimation of points on a psychometric function. *British Journal of Mathematical and Statistical Psychology*, *18*, 1–10.