PERGAMON

Vision Research 39 (1999) 3537–3560

Vision Research

www.elsevier.com/locate/visres

Identification of band-pass filtered letters and faces by human and ideal observers

Jason Gold, Patrick J. Bennett *, Allison B. Sekuler

Department of Psychology, 100 St. George St., University of Toronto, Toronto, Ont., Canada, M5S 3G3

Received 22 September 1998; received in revised form 11 March 1999

Abstract

To better understand how the visual system makes use of information across spatial scales when identifying different kinds of complex patterns, we measured human and ideal contrast identification thresholds to estimate identification efficiency for 1- and 2-octave wide band-pass filtered letters and faces embedded in 2-D dynamic Gaussian noise. Varying stimulus center frequency from 1 to 70 c/object had different effects on letter and face identification efficiency. In the 2-octave conditions, identification efficiencies decreased by 0.25–0.5 log units for letters and 0.5–1.2 log units for faces as center frequency increased from 6.2 to 49.5 c/object, but only letters were identifiable at center frequencies below 6.2 c/object. In the 1-octave conditions, letter identification efficiencies increased by about 0.5 log units as center frequency increased from 1.1 to 2.2 c/object, and were nearly constant from 2.2 to 35 c/object. Letters were unidentifiable by human observers at 70 c/object. Surprisingly, face identification was impossible for human observers at all center frequencies except 8.8 c/object for one observer, and 8.8 and 17.5 c/object for a second observer. Ideal observer thresholds were obtained for both letters and faces in all conditions, so information was always available to perform the task. Thus, the failure to identify faces reflects constraints on visual processing rather than a lack of stimulus information. Selective spatial sampling may account for some of the differences between letter and face identification efficiencies.

Keywords: Ideal observer; Face identification; Letter identification; Frequency filtering

1. Introduction

Most models of visual processing assume that patterns are encoded by multiple spatial frequency-tuned channels, and these models generally provide a good account of detection and discrimination data obtained with simple visual patterns (De Valois & De Valois, 1988; Graham, 1989). However, it less clear how higher-order mechanisms make use of such a representation when performing tasks with more complex, naturalistic patterns. For example, some studies suggest that the visual system may rely upon information carried by a common band of frequencies for the identification of different kinds of complex patterns (Solomon & Pelli, 1994; Braje, Tjan & Legge, 1995; Chung & Legge, 1997; Majaj, Palomares, Mouchraud, Kotlyarenko & Pelli, 1998), but other findings imply that socially significant or well-learned patterns, such as faces, may be processed by specialized mechanisms (Yin, 1969, 1970; Gross, Rocha-Miranda & Bender, 1972; Benton, 1980; Hay & Young, 1982; Perrett, Rolls & Caan, 1982; Desimone, Albright, Gross & Bruce, 1984; Baylis, Rolls & Leonard, 1985; Damasio, 1985; Desimone & Schein, 1987; Perrett, Mistlin & Chitty, 1987; Perrett, Mistlin, Chitty & Smith, 1988; Ellis & Young, 1989; Sergent, 1989; Damasio, Tranel & Damasio, 1990; Desimone, 1991; de Haan, Young & Newcombe, 1992; Heywood & Cowey, 1992; Sergent, Ohta & MacDonald, 1992). In this article, we have attempted to address the issue of whether human observers make use of information across spatial frequencies in the same way when identifying different kinds of complex patterns. Our approach was to measure the ability of human observers to use information contained within different bands of spatial frequencies when identifying two very different kinds of complex patterns: English letters and human faces. By comparing human identification performance

* Corresponding author.

E-mail address: bennett@psych.utoronto.ca (P.J. Bennett)

0042-6989/99/$ - see front matter
PII: S0042-6989(99)00080-2

to that of an ideal observer, we were able to separate constraints imposed by the availability of stimulus information from constraints imposed by human visual processing. Differences between face and letter identification efficiencies would be consistent with the idea that either: (i) lower-level visual processing imposes different constraints upon the information available for letter and face identification; or (ii) different mechanisms that make differential use of information across spatial scales subserve face and letter identification. Alternatively, similar efficiencies for letters and faces would be consistent with the proposal that the visual system makes use of available information across spatial scales in a general fashion when identifying complex patterns.

1.1. Identification performance

Many studies have explored the role of spatial frequency in the identification of patterns within a particular stimulus category, such as letters (Ginsburg, 1978; Rubin & Siegel, 1984; Legge, Pelli, Rubin & Schleske, 1985; Parish & Sperling, 1991; Alexander, Xie & Derlacki, 1994; Solomon & Pelli, 1994; Chung & Legge, 1997), faces (Harmon, 1973; Harmon & Julesz, 1973; Tieger & Ganz, 1979; Ginsburg, 1980; Riley & Costall, 1980; Fiorentini, Maffei & Sandini, 1983; Morrone, Burr & Ross, 1983; Rubin & Siegel, 1984; Hayes, Morrone & Burr, 1986; Bachmann, 1987; Schuchard & Rubin, 1989; Bachmann, 1991; Costen, Parker & Craw, 1994; Peli, Lee, Trempe & Buzney, 1994; Costen, Parker & Craw, 1996; Parker, Lishman & Hughes, 1996), objects and geometric forms (Ginsburg, 1984; Norman & Ehrlich, 1987; Braje et al., 1995; Tjan, Braje, Legge & Kersten, 1995; Parker et al., 1996; Cannon et al., 1997), or natural scenes (Field, 1987; Parker, Lishman & Hughes, 1992; Field, 1993; Schyns & Olivia, 1994). The majority of these studies have examined whether there are 'critical identification bands': bands of frequencies that most effectively carry information about the identity of the patterns for human observers. Often, critical identification bands are defined as the range(s) of frequencies that yield optimal identification performance—as measured by accuracy, sensitivity, or response time—across differentially filtered, quantized, or frequency-masked patterns. Table 1 summarizes the stimuli, methods, and results of many experiments that have measured critical identification bands for a variety of complex patterns in terms of performance.

There is considerable consensus across studies that the critical identification band for letters is between 1 and 2 octaves wide, and centered somewhere between 2 and 6 c/letter (Legge et al., 1985; Solomon & Pelli, 1994; Chung & Legge, 1997; Majaj et al., 1998). In contrast, there is considerable disagreement regarding the critical identification band for faces (Harmon, 1973; Tieger & Ganz, 1979; Ginsburg, 1980; Fiorentini, Maffei & Sandini, 1983; Rubin & Siegel, 1984; Hayes et al., 1986; Bachmann, 1987; Schuchard & Rubin, 1989; Bachmann, 1991; Costen et al., 1994, 1996; Majaj et al., 1998; Schwartz, Bayer & Pelli, 1998). Estimates of the center frequency of the critical identification band for faces range from 1 c/face (Rubin & Siegel, 1984) to beyond 25 c/face (Hayes et al., 1986). One explanation for this disagreement is that the different tasks, methods, and conditions used in face identification studies may have tapped into different aspects of performance. Additionally, the comparison of performance levels across conditions fails to take into account differences in the amount of stimulus information available as a function of spatial frequency and stimulus set (Parish & Sperling, 1991; Solomon & Pelli, 1994; Braje et al., 1995; Liu, Knill & Kersten, 1995; Tjan et al., 1995). For example, suppose we high-pass filter two sets of faces, measure identification accuracy for the filtered images, and find that identification performance is near chance for the first set but well above chance for the second set. How are we to account for the difference in performance? One explanation is that the filtering operation affected stimulus information differently in the two sets. For instance, high-pass filtering may have rendered all of the faces in the first set *physically identical* (or very similar), but left large differences among the faces in the second set. In such a case, we would naturally expect performance for the first set to be near chance because there would be no information available to perform the identification task. In fact, an ideal machine that used all of the available stimulus information to do the same task would perform poorly on the first set, too. The same ideal machine would perform better on the second stimulus set, because there is more information available for the task. This example illustrates that without some measure of available stimulus information, it is unclear whether differences in performance are due to variations in the *ability to use* information or variations in the *availability* of information (or both).

1.2. Ideal observer analysis and efficiency

One solution to the problem caused by variations in the amount of available information across conditions is to compare human performance to that of an *ideal observer*. An ideal observer is a theoretical device whose performance is constrained only by the availability of stimulus information. By definition, the ideal observer uses a strategy that yields the best possible performance for a given task, and therefore provides an index of information available to perform that task (Barlow, 1978; Kersten, 1987; Geisler, 1989; Kersten, 1990; Braje et al., 1995; Tjan et al., 1995). *Efficiency* in any given

Table 1
A summary of experiments that have measured identification performance for various kinds of complex patterns[a]

Authors	Stimulus	Measure	SF Manipulation	Performance
Harmon (1973)	Faces	Accuracy	Pixelization	16×16 pixel quantization (8 c/fw) yielded threshold (50%) recognition accuracy
Ginsburg (1978)	Letters	Accuracy	LP filtering	Threshold percent correct obtained between 1.5 and 3 c/letter
Tieger and Ganz (1979)	Faces	Accuracy	Masking by two orthogonal component sine-wave gratings	17.6 cycles/fw produced greatest masking
Ginsburg (1980)	Faces	Accuracy	2-octave Gaussian filtering	Threshold percent correct obtained between 1 and 4 c/face
Fiorentini, Maffei and Sandini (1983)	Faces	Accuracy	HP and LP filtering	Fewer errors with HP = 5 cycles/fw than with LP = 5 c/fw
Ginsburg (1984)	Geometric forms	Accuracy	LP filtering	Threshold percent correct obtained at ~2.28 c/object
Rubin and Siegel (1984)	Faces and letters	Accuracy	LP filtering	Threshold percent correct obtained at.7 c/letter and 1 c/face
Legge, Pelli, Rubin and Schleske (1985)	Letters	Reading rate	LP filtering	Reading rate declined rapidly below ~2 c/letter
Hayes, Morrone and Burr (1986)	Faces	Accuracy	1.5-octave ideal filtering	Peak accuracy between 20 and 25 c/fw
Norman and Ehrlich (1987)	Toy tanks	Accuracy, RT	2-octave, HP, and LP pseudo-Gaussian filtering	Fewest errors and lowest RTs beyond 28.6 c/picture width
Schuchard and Rubin (1989)	Faces	Accuracy	1.5-octave filtering	Equal performance across frequency ranges
Bachmann (1991)	Faces	Accuracy	Pixelization	Steep decrease in accuracy between 15 and 18 pixels/fw (7.5–9 c/fw)
Peli, Lee, Trempe and Buzney (1994)	Faces	Accuracy	LP filtering by adding 1-octave exponential filters	Steep increase in accuracy above 4 c/face height
Costen, Parker and Craw (1994)	Faces	Accuracy, RT	Ideal LP filtering; pixelization; Gaussian blurring	Steep increase in accuracy between 5.5 and 10.5 c/fw
Solomon and Pelli (1994)	Letters and gratings	Accuracy, efficiency	HP and LP noise masking	Steep increase in threshold s/n below 3 c/ letter for LP and above 3 c/letter for HP
Costen, Parker and Craw (1996)	Faces	Accuracy, RT	Exponential HP and LP filtering; pixelization	Peak changes in accuracy between 8 and 16 c/fw
Cannon, Hoffmeister, and Fullenkamp (1997)	Aircraft	Accuracy	HP and LP noise masking	Steep increase in threshold s/n below ~6 c/letter for LP and above ~6 c/letter for HP
Chung and Legge (1997)	Letters	Accuracy	1-octave cosine-log filtering	Peak contrast sensitivity for recognition between 1.8 and 2.5 c/letter
Majaj, Palomares, Mouchraud, Kotlyarenko and Pelli (1998)	Fonts, letters, faces, objects, and gratings	Accuracy	HP and LP noise masking	Similar findings to Solomon and Pelli (1994) and Gold et al. (1998) for all patterns
Schwartz, Bayer, and Pelli (1998)	Facial expressions	Accuracy	HP and LP noise masking	Steep increase in threshold below 8 c/ face for LP and above 8 c/face for HP

[a] The studies are listed in chronological order.

task is defined as the ratio of ideal to human threshold energy. An efficiency of 100% implies that a human observer is using stimulus information optimally. An efficiency less than 100% implies that a human observer is using stimulus information sub-optimally; information is being lost somewhere between stimulus presentation and response measurement. Because comparison to the ideal observer controls for differences in available information, variations in efficiency across experimental conditions imply variations in the ability of the human observer to use available information. In contrast, constant efficiency implies that the ability to use available information is invariant, regardless of changes in performance.

1.3. Identification efficiency

No study has measured efficiency for face identification, although one study has looked at efficiency for recognizing facial expressions (Bayer, Schwartz & Pelli, 1998). Several studies have measured efficiency for identifying letters (Parish & Sperling, 1991; Beckman & Legge, 1994; Solomon & Pelli, 1994; Burns, Farrel, Moore & Pelli, 1995; Tjan et al., 1995) and 3-D shapes (Braje et al., 1995; Liu et al., 1995; Tjan et al., 1995). The stimuli, methods, and results of these studies are summarized in Table 2. Critical identification bands estimated for letters (Parish & Sperling, 1991) and low-pass filtered shapes (Braje et al., 1995) are similar. However, none of these studies directly compared identification efficiencies for different kinds of band-pass filtered patterns (e.g. letters versus shapes; letters versus faces) in the same observers. Therefore, we used a technique similar to Parish and Sperling (1991) to compare identification efficiencies for 1- and 2-octave wide band-pass filtered letters and faces. Again, the main purpose for comparing identification efficiencies for filtered letters and faces was to determine whether the efficiency with which human observers use information at different spatial scales depends on the kind of pattern being identified.

2. Method

2.1. Observers

Three female and three male members of the University of Toronto Vision Laboratory volunteered as observers (not all participants were tested in every condition). All had normal or corrected-to-normal vision. Participants ranged from 21 to 37 years of age, with a mean age of 28. Three of the observers were naive with respect to the experimental hypotheses. All but one of the participants had previous experience in psychophysical tasks.

2.2. Apparatus

Stimuli were displayed on a pair of AppleVision 1710 color monitors. One monitor showed the signal and the other showed dynamic Gaussian noise (see Section 2.3 below for details). Each monitor displayed 800×600 pixels, which subtended a visual angle of $15.8 \times 11.5°$ from the viewing distance of 100 cm, at a frame rate of 75 Hz (non-interlaced). A half-silvered mirror placed at a 45° angle relative to the monitors optically combined the signal and noise displays. Luminance calibrations were performed with a Hagner Optikon universal spot photometer, and the calibration data were used to build a 2025-element look-up table (Tyler, Chan, Liu, McBride & Kontsevich, 1992) for each display. The experiment was conducted in the MATLAB programming environment (version 4.2c1), using the extensions provided by the Psychophysics Toolbox (Brainard, 1997) and the Video Toolbox (Pelli, 1997). When constructing the stimuli used on each trial, the computer software selected appropriate luminance values from the calibrated look-up tables and stored them in the 8-bit look-up tables of each display. Luminance on the optically-combined display ranged between 0.5 and 77.9 cd/m^2, with an average luminance of 39 cd/m^2. Pixel contrast (as defined by Eq. (1) below) on the optically-combined display could be varied between -0.99 and 0.99.

2.3. Stimuli

The stimuli were digital images (256×256 pixels in size) of letters and faces that were constructed using Adobe Photoshop (version 3.0), MacPhase (version 2.0), and MATLAB. The images were generated prior to the experiment and stored on disk. The values in each image represented the contrast (c_{xy}) at pixel location (x, y), defined by Eq. (1):

$$c_{xy} = \frac{l_{xy} - L}{L} \tag{1}$$

where L is average luminance and l_{xy} is the pixel luminance. The values in each image file varied from -1 to 1, and were normalized so that contrast variance (i.e. the variance of the contrast values across the entire image) equaled 1. Prior to each experimental trial, an image file was read into memory, contrast variance was set to the desired value by multiplying the image data by an appropriate constant, and the contrast values were converted to luminance values. These luminances were used to construct a linear 8-bit look-up table for the display. Finally, the image luminance values were mapped onto the values stored in the look-up table.

Ten uppercase letters of the English alphabet (A, F, G, J, L, P, Q, R, T and Y) were used for the letter identification experiment. The Geneva font was chosen for its relative simplicity and comparability with previous studies (Solomon & Pelli, 1994; Tjan et al., 1995). The letters were equated for height (198 pixels), and ranged between 100 and 190 pixels in width (mean of 142 pixels, S.D. = 29 pixels). Each letter was centered within a uniform background of average luminance that was 256×256 pixels in size.

Table 2
A summary of experiments that have measured identification efficiency for various kinds of complex patterns[a]

Authors	Stimulus	Measure	SF Manipulation	Performance	Efficiency
Parish and Sperling (1991)	Letters	Accuracy, efficiency	2-octave Gaussian filtering	Peak sensitivity at 20.25 c/lh	Peak efficiency at 1.49 c/lh (42%)
Solomon and Pelli (1994)	Letters and gratings	Accuracy, efficiency	HP and LP noise masking	Steep increase in threshold s/n below 3 c/ letter for LP and above 3 c/letter for HP	Peak efficiency at $\sim$ 3 c/letter (13%)
Burns, Farrel, Moore and Pelli (1995)	Letters (of various alphabets)	Accuracy, efficiency	–	–	Efficiencies ranged from 4–16%
Braje, Tjan and Legge (1995)	Line drawings and silhouettes of 3-D shapes	Accuracy, efficiency	LP Gaussian filtering	Performance is relatively constant as bandwidth increases beyond 6 c/object	Steep decrease in efficiency below 6 c/object
Liu, Knill and Kersten (1995)	3-D thick wire shapes	Accuracy, efficiency	–	–	Efficiencies between 10 and 20%
Tjan, Braje, Legge and Kersten (1995)	Unfiltered letters and 3-D shapes	Accuracy, efficiency	–	–	Letters: 16.3% Shapes: 8%
Bayer, Schwartz and Pelli (1998)	Facial expressions	Accuracy, efficiency	Filtering by critical band, as determined by Schwartz et al. (1998)	–	Efficiency of 9% with critical band & critical area

[a] The studies are listed in chronological order.

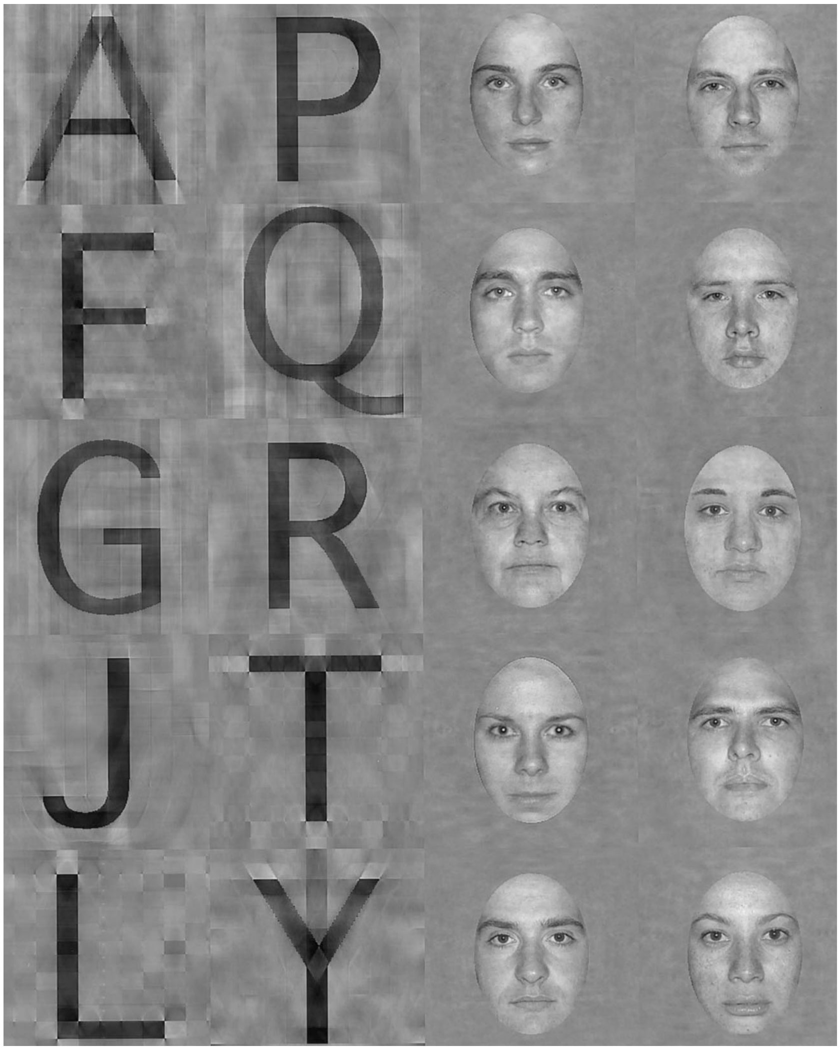


Fig. 1. The ten unfiltered letter and face stimuli used in the main experiment. All ten images in each set share the same amplitude spectrum (an average of the individual spectra; see text for details).

The ten letters have different amplitude spectra. Such amplitude differences could lead to differences in stimulus detectability, which might be used as a cue to identify some letters. To minimize this potential problem, differences in the amplitude spectra were removed in the following way. First, each letter was set to a contrast of -1 and the background was set to a contrast of zero. Next, the Fourier transform of each letter was computed, and the modulus at each spatial frequency and orientation was averaged across all letters. After averaging, the DC component was set to 0. Finally, the amplitude spectrum for each letter was replaced by the average amplitude spectrum, and the inverse Fourier transforms were computed. The result of this process was a set of letter stimuli that had identical amplitude spectra (Fig. 1). It is important to note that the amplitude spectra of the original letters differed only slightly from each other, and so the appearance of letters in Fig. 1 did not differ significantly from the original items.

Seven sets of 1-octave, band-pass filtered stimuli were constructed by filtering the average amplitude spectrum by 1-octave rectangular filters with center frequencies of 1.1, 2.2, 4.4, 8.8, 17.5, 35.0, and 70.0 c/object height (c/obj). From the viewing distance of 100 cm, the center frequencies were 0.2, 0.5, 1.0, 2.0, 4.0, 8.0, and 16.0 c/deg. Four sets of 2-octave, band-pass stimuli also were constructed by filtering the average amplitude spectrum with 2-octave rectangular filters with center frequencies of 1.5, 6.2, 24.8, and 49.5 c/obj (0.3, 1.4, 5.6, 11.2 c/deg). The two highest 2-octave bands overlapped by 1 octave within the region of 24.8–49.5 c/obj. The filtered and unfiltered amplitude spectra were combined with the phase spectrum of each letter, to produce 120 letter stimuli (11 filtered amplitude spectra $\times$ 10 phase spectra $+$ 1 unfiltered amplitude spectrum $\times$ 10 phase spectra). Examples of a 1- and 2-octave filtered letter are shown in Fig. 2.

The faces of ten Caucasian models (five male and five female) from the University of Toronto Psychology

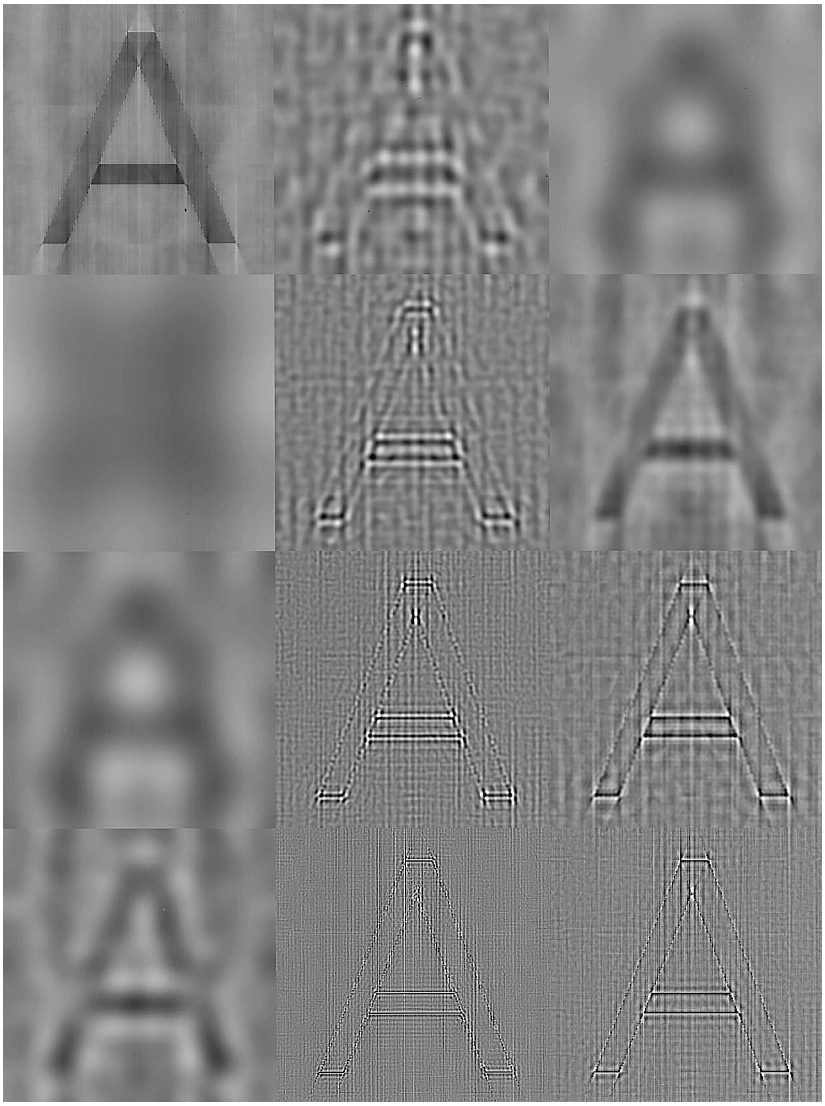


Fig. 2. Examples of filtered letter stimuli used in the main experiment. The top left image is an unfiltered 'A'. The corresponding 1-octave wide band-pass filtered versions of 'A' are shown below the unfiltered image and in the middle column. The center frequencies of the filters applied to the images are, from top to bottom in the left column: 1.0, 2.2, and 4.4 c/obj height; from top to bottom in the middle column: 8.8, 17.5, 35.0 and 70.0 c/obj height. The 2-octave wide band-pass filtered versions of the same letter are shown in the rightmost column. The center frequencies of the filters applied to the images are (from top to bottom): 1.5, 6.2, 24.8, and 49.5 c/obj height.

department were used in the face identification experiments. Models were highly familiar to all of the observers in the experiment, and none of the observers served as models. All faces were photographed in front of a uniform black field. Glasses, makeup, and any other non-facial cues were removed from models' faces before being photographed. Each model's hair was held back away from the face and forehead by a small head cap. None of the models had facial hair. All models were asked to look directly at the camera with a neutral facial expression. The film was developed directly to photographic CD-ROM, and each picture was digitally converted to grayscale and cropped to show only the inner portion of the face, eliminating non-facial cues such as hair and ears. The shape of the visible region of each face was elliptical, and the size and height:width ratio were constant across all stimuli (198 pixels:140 pixels). The faces were centered within a 256×256 pixel background of average luminance.

The contrast values for each face were first linearly transformed so that they ranged from -1 to 1 and the background was set to zero. Next, differences in the amplitude spectra of the faces were eliminated the same way as was done with letters. The faces thus shared a common amplitude spectrum (Fig. 1). The same 1- and 2-octave wide filters used in the creation of the letter stimuli were applied to the face average amplitude spectrum and combined with the phase spectrum for

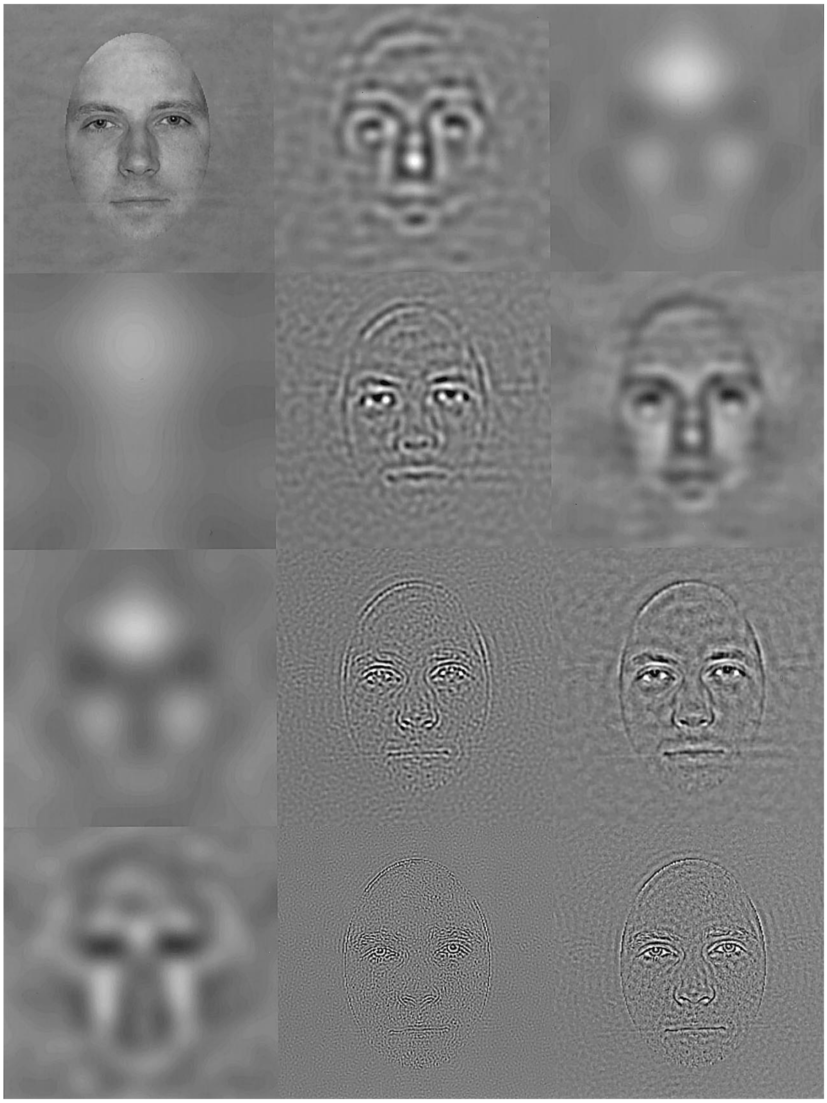


Fig. 3. Examples of filtered face stimuli used in the main experiment. The top left image is an unfiltered face. The corresponding 1-octave wide band-pass filtered versions of the face are shown below the unfiltered image and in the middle column. The 2-octave wide band-pass filtered versions of the same face are shown in the rightmost column. The images are arranged the same way as in Fig. 2.

each of the ten faces, producing a set of 120 face stimuli. Examples of a 1- and 2-octave filtered face are shown in Fig. 3.

The face and letter stimuli can be obtained over the internet[1].

2.4. Noise fields

Unfiltered dynamic 2-D Gaussian noise was produced from individual Gaussian noise fields that were 256×256 pixels in size. The values were taken from a Gaussian pseudo-random number generator with a mean of 0 and contrast variance of 0.2. As with the signal stimuli, each value within the matrix was treated as a contrast value. The variance chosen ensured that 95% of the values in the distribution would fall within the linear contrast range of the noise display. Values beyond ± 2 S.D. from the mean were truncated at the maximum and minimum contrast values. The spectral density of the noise (energy per unit bandwidth) was 1.15×10^{-6} deg^2 in all experimental conditions. A set of 80 static noise fields was created, and noise fields were chosen randomly from the set for every frame within the presentation duration of 500 ms (i.e. a total of 37 frames/trial).

2.5. Viewing conditions

At the viewing distance of 100 cm, the 256×256 pixel stimulus field subtended a visual angle of $5.25 \times 5.25°$. Each letter within the stimulus field subtended a

[1] The unfiltered face and letter images are available over the world wide web at http://www.psych.utoronto.ca/ ~ vislab.

vertical visual angle of 4.0°, with a horizontal visual angle ranging between 2.0 and 3.8°. Faces subtended a vertical visual angle of 4.0° and a horizontal visual angle of 2.9°. Viewing was binocular through natural pupils, and a head/chinrest stabilized the observer's head. The monitors supplied the only sources of illumination during the experiment.

2.6. Procedure

Identification thresholds were estimated using a single-interval, ten-choice identification task. Stimulus contrast variance was manipulated using the method of constant stimuli. For human observers, pilot studies identified six contrast variances that typically spanned the threshold range in each condition. During the experiment, the order of contrast variances was randomized, and stimuli were presented at each variance for a block of 15 trials. Thus, a minimum of 120 trials (15 trials × 6 contrast levels) were shown in each condition. The order of stimulus presentation within each 15 trial block was random, with the restriction that each of the ten stimuli was presented at least once. In some cases the contrast variances taken from the pilot experiments did not span an observer's threshold, so additional contrast levels were tested. In all cases where performance did not significantly exceed chance, stimulus contrast values included the maximum contrast possible on the display.

Observers were familiarized briefly with the unfiltered versions of each image before beginning the experiment. At the start of each trial, a fixation point appeared at the center of the signal screen, and a brief tone indicated a trial could commence with a mouse click. After an initial 67 ms of noise alone, the stimulus + noise combination appeared for 427 ms. Next, the displays were set to average luminance, and after a brief pause (100 ms), a set of ten thumbnail versions of the original unfiltered images (128 × 128 pixels in size) appeared on the screen surrounding the region where the signal and noise had been displayed. Observers identified the stimulus by clicking the mouse on the appropriate image. Once an image was chosen, the displays were cleared, and set to average luminance. Auditory feedback after each trial indicated whether the response was correct. Only one type of stimulus (either faces or letters) was used in each testing session. The orders of the tasks were counterbalanced for observers participating in both the face and letter identification tasks. The frequency bands were presented in a different random order for each observer.

Best-fitting (least squares) Weibull functions were fit to the data, and threshold was defined as the contrast variance yielding 67% correct responses, which corresponds to a d' of 2 in our conditions.

2.7. Ideal observer

For our stimuli and task, it can be shown that maximizing the cross-correlation between the stimulus (i.e. signal + noise) and each of the ten possible signal matrices (templates) is the strategy that yields optimal performance (Green & Swets, 1966; Tjan et al., 1995). Ideal observer thresholds were obtained in all conditions through Monte Carlo simulations, in which each template was compared to the filtered stimulus + noise combination at a range of contrast values. For each trial, the ideal observer simply chose the template that yielded the highest cross-correlation with the stimulus. Ideal thresholds were estimated from psychometric functions that were fit to the data from at least 600 simulated trials. Efficiency was defined as the ratio of ideal to human threshold energy in each condition.

3. Results

3.1. 1-Octave ranges

One-octave letter and face identification thresholds for two human subjects and the ideal observer are plotted in Fig. 4. The corresponding one-octave letter and face identification efficiencies for the human observers are plotted in Fig. 5.

3.1.1. Letters

Fig. 4 shows that the ideal observer was able to perform the letter identification task in all conditions, so information always was available to perform the task. The two human observers were able to perform the letter identification task in all but the highest frequency condition (70 c/obj), where the maximum displayable image contrast yielded less than 50% correct performance. For the remaining seven conditions in which identification was possible, thresholds were elevated in the lowest frequency condition (1.1 c/obj) and remained relatively constant in the remaining conditions (including the unfiltered condition). The corresponding letter identification efficiencies in Fig. 5 are also relatively constant across frequencies, except at the lowest frequencies where efficiency declined by 0.4 (JMG) and 0.7 (JMH) log units and at the highest frequencies where efficiency fell to 0. Peak efficiency for the filtered stimuli was about 0.6% for observer JMG and 0.5% for observer JMH, and was nearly identical to each observer's efficiency in the unfiltered condition.

3.1.2. Faces

Fig. 4 shows that the ideal observer was also able to perform the face identification task in all conditions, and therefore that stimulus information for identification was available in all conditions. However, both human observers were unable to perform the face identification task in most conditions. Both observers were able to identify the faces in the unfiltered and 17.5 c/obj conditions; observer JMG was also able to perform above chance in the 8.8 c/obj condition (although performance never exceeded about 75% correct in this condition). The corresponding face identification efficiencies in Fig. 5 show efficiency was 0 for both observers in most conditions. Peak efficiency for both observers occurred in the unfiltered condition (JMG $\approx$ 1.3%; JMH $\approx$ 0.4%).

3.1.3. Comparison of letters and faces

Probably the most striking aspect of the data is the fact that human observers were unable to identify 1-octave filtered faces at most spatial scales, yet were able to identify identically filtered letters at all but the highest frequencies. Despite the inability of both observers to perform the face task in most conditions, thresholds were obtained for the ideal observer in all conditions, indicating that information always was available to perform the face identification task. However, ideal face identification thresholds were 0.5–1 log unit higher than ideal letter identification thresholds for the unfiltered stimuli and for band-pass stimuli with center frequencies below 17.5 c/obj. This difference raises the possibility that human observers could perform both the letter and face tasks with equal efficiency, but that some of the thresholds in the face task simply were higher than the maximum contrast that could be displayed on our equipment.

We tested this idea in two ways. First, in each condition, we calculated the ratio of ideal face identification thresholds divided by ideal letter identification thresholds, and then multiplied that ratio by human letter identification thresholds. The resulting values are the predicted face identification thresholds for human observers if efficiency in each condition was the same for letters and faces. In all cases the predicted threshold was below the maximum displayable contrast. Second, we measured face identification performance for observer JMG with no noise and with the stimuli set to the maximum device contrast in all frequency ranges. Performance was near chance for all but the 8.8 c/obj,

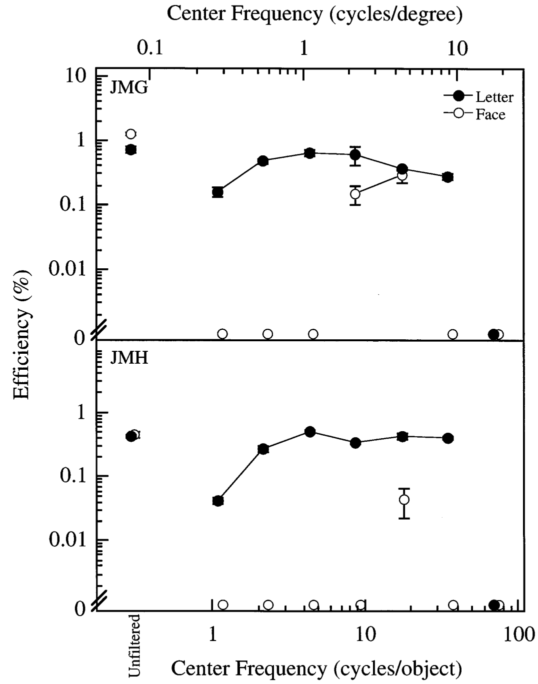


Fig. 5. One-octave wide filtered face and letter identification efficiencies for two human observers, plotted as a function of the center frequency of the filter. The upper abscissa corresponds to c/deg, and the lower abscissa c/obj. The left ordinate corresponds to efficiency, expressed as percent. Closed symbols show the performance in the letter conditions, open symbols the face conditions. The first data point after the origin along the abscissa is the unfiltered condition. Error bars on each symbol depict ± 1 S.E. of the threshold estimate. Conditions in which thresholds were not obtained are plotted at the bottom of each panel (efficiency = 0).

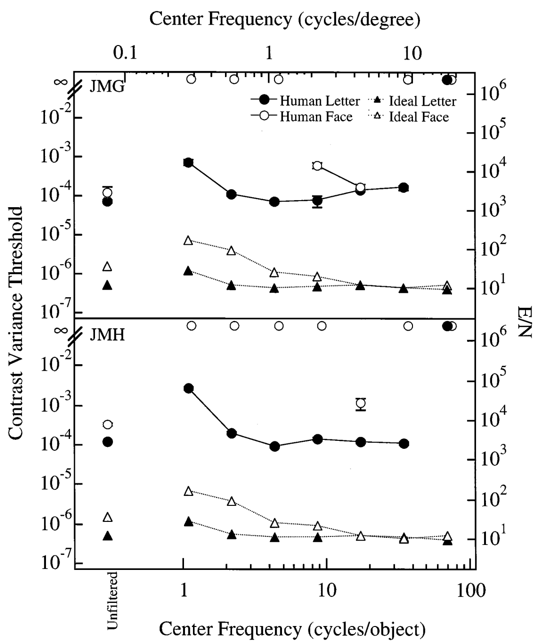


Fig. 4. One-octave wide filtered letter and face identification thresholds for two human observers and the ideal observer, plotted as a function of the center frequency of the filter applied to the images. The upper abscissa corresponds to c/deg, and the lower abscissa c/obj. The left ordinate corresponds to contrast variance threshold. The right ordinate corresponds to threshold signal-to-noise ratio, (E/N), where E equals the threshold contrast energy (i.e. the product of the contrast variance and the size of the stimulus, in ° of visual angle^2) and N is the noise spectral density. Closed symbols show the performance in the letter conditions, open symbols the face conditions. Circles represent human performance, triangles ideal performance. The first data point after the origin along the abscissa is the unfiltered condition. Error bars on each symbol depict ± 1 S.E. of the threshold estimate. Often, the error bars are smaller than the symbols. Conditions where thresholds were not obtained because the observer was unable to perform significantly above chance are plotted at the top of each panel (threshold = ∞).

17.5 c/obj, and unfiltered conditions. Both results suggest that the failure to identify faces was not solely due to the limited dynamic range of the displays. Instead, the failure to identify faces reflects a genuine reduction in efficiency.

3.2. 2-Octave ranges

Two-octave letter and face identification thresholds for three human subjects and the ideal observer are plotted in Fig. 6. The corresponding two-octave letter and face identification efficiencies for the human observers are plotted in Fig. 7.

3.2.1. Letters

Fig. 6 shows that human and ideal observers were able to perform the letter identification task in all conditions. As was found with 1-octave filtered letters, identification efficiency with 2-octave filtered letters varied only slightly over a 33-fold range of spatial scale,

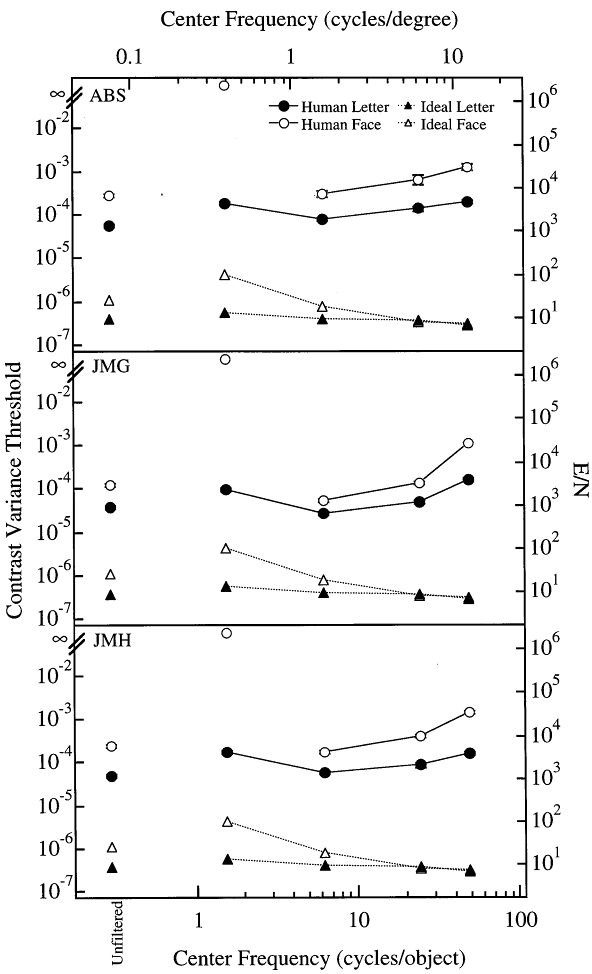


Fig. 6. Two-octave wide filtered letter and face identification thresholds for three human observers and the ideal observer, plotted as a function of the center frequency of the filter. Symbol conventions are the same as in Fig. 4.

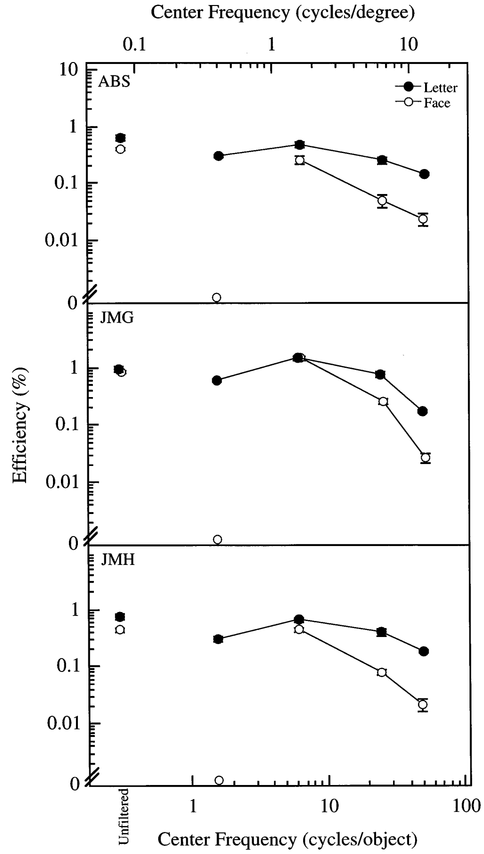


Fig. 7. Two-octave wide filtered letter and face identification efficiencies for three human observers, plotted as a function of the center frequency of the filter. Symbol conventions are the same as in Fig. 5.

but the 2-octave functions were somewhat more band-pass in shape (Fig. 7). For all three observers, peak efficiency for filtered stimuli (0.7–1.5%) occurred at 6.2 c/obj and was 0.25–0.5 log units lower in the 1.5 and 49.5 c/obj conditions. Each observer's peak efficiency was similar to that obtained with unfiltered stimuli.

3.2.2. Faces

Fig. 6 shows that the ideal observer was able to perform the face identification task in all conditions. Unlike what was found with 1-octave filtered faces, human observers were able to identify 2-octave filtered faces in all but the lowest frequency condition (1.5 c/obj). As with 2-octave filtered letters, efficiency for the filtered faces was greatest in the 6.2 c/face condition (Fig. 7). Peak efficiency (0.3–1.5%) was nearly identical to that obtained with unfiltered faces for all observers. Efficiency decreased by 0.5–1.2 log units as center frequency increased from 6.2 to 49.5 c/face. Efficiency was zero for all observers in the 1.5 c/face condition.

3.2.3. Comparison of faces and letters

Human efficiencies for both the filtered faces and letters peaked with a 2-octave band centered at 6.2

c/obj. Peak efficiencies were similar for faces and letters, and observers were as efficient in the 6.2 c/obj conditions as in the unfiltered conditions. Also, peak efficiencies for these stimuli were comparable to the peak efficiencies obtained in the 1-octave task. However, our data show a clear difference between letter and face identification as a function of spatial frequency: human observers were less able to use information available for face than for letter identification above and below the 2-octave range centered at 6.2 c/obj.

3.3. Possible factors contributing to low, broadband letter identification efficiency functions

Absolute letter identification efficiencies in both the 1- and 2-octave conditions were lower than previous estimates: Our highest identification efficiencies never exceeded about 1.5%, whereas Parish and Sperling (1991) reported a peak efficiency of 42%, and Solomon and Pelli (1994) and Tjan et al. (1995) obtained peak values between 12 and 16%. Also, our functions relating letter identification efficiency to spatial frequency were broadly tuned (especially in the 1-octave condition), whereas Parish and Sperling (1991) obtained much narrower functions. We conducted an additional series of experiments to examine these discrepancies. Specifically, we examined the effects of using: (i) response feedback; (ii) band-pass, rather than white, noise; and (iii) static, rather than dynamic, noise.

3.4. The use of feedback

Unlike the current experiments, Parish and Sperling (1991) did not include feedback in their procedure. We examined the effects of feedback by testing a new observer (PJB) without feedback in the 1-octave letter identification task. Thresholds for PJB with no feedback were virtually identical to those obtained previously with observers who received feedback. Thus, the use of feedback does not appear to alter the either the peak or the bandwidth of the letter identification efficiency functions.

3.5. Dynamic versus static noise

Parish and Sperling (1991) and Solomon and Pelli (1994) used static noise in their experiments, whereas the current experiment used dynamic noise. To examine whether this difference reduced efficiency in our task, we re-measured efficiencies for the same 2-octave filtered letters and faces used in the main experiment embedded in 2-D static Gaussian white noise. All other experimental conditions were the same, except that identification contrast variance thresholds for the ideal and human observers were estimated with the QUEST adaptive staircase procedure (Watson & Pelli, 1983). Stimuli remained blocked for frequency range, and, within a block, images were chosen randomly from the set of 10 possible choices. A block lasted until either the standard error of the threshold estimate reached 0.1 log units or 75 trials had elapsed. The final threshold estimates consisted of the average of at least two thresholds obtained in this fashion.

Efficiencies for both observers with static and dynamic noise are shown in Fig. 8. The ideal observer sums information across time with maximal efficiency, and therefore reducing the number of independent noise fields presented during each trial causes ideal letter and face identification thresholds to increase. Human letter identification thresholds also increased in the static noise condition, but less than ideal thresholds. Thus, letter identification efficiencies with static noise were three to nine times greater than to those obtained using dynamic noise. The highest letter identification efficiency with static noise was approximately 8%, a value that is closer to the letter identification efficiencies reported by Solomon and Pelli (1994) and Tjan et al. (1995). Thus, much of the difference in absolute efficiency between our initial letter identification data and those obtained previously can be attributed to the use of dynamic noise. However, the use of static noise did not influence absolute face identification efficiency: human and ideal face identification thresholds increased by similar amounts in all conditions, and therefore identification efficiencies for faces embedded in static noise were similar to those obtained using dynamic noise. These results suggest that temporal summation of stimulus information is more efficient for faces than for letters.

3.6. Unfiltered versus band-pass noise

The current experiments used unfiltered Gaussian noise, whereas Parish and Sperling (1991) used Gaussian noise that was filtered by the same band by which the stimuli had been filtered in each condition. We tested the possibility that the wide bandwidth of the noise across all conditions also contributed to our low absolute efficiencies by re-measuring the 2-octave letter and face thresholds for one human observer (JMG) with static Gaussian noise filtered in the same fashion as the stimuli in each condition. For this type of filtered stimulus and noise, cross-correlation is still the ideal decision rule. Under our conditions, ideal thresholds with unfiltered noise and noise filtered identically to the stimulus are the same[2].

[2] In our conditions, the signal-to-noise ratio at each frequency outside of the stimulus pass-band is zero for both unfiltered noise and noise filtered identically to the stimulus. Within the stimulus pass-band, the signal-to-noise ratio at each frequency is the same for both kinds of noise. Hence, the ideal rule (and ideal threshold) remains the same under our conditions because filtering the noise in the same fashion as the stimulus does not change the signal-to-noise ratio at any frequency.

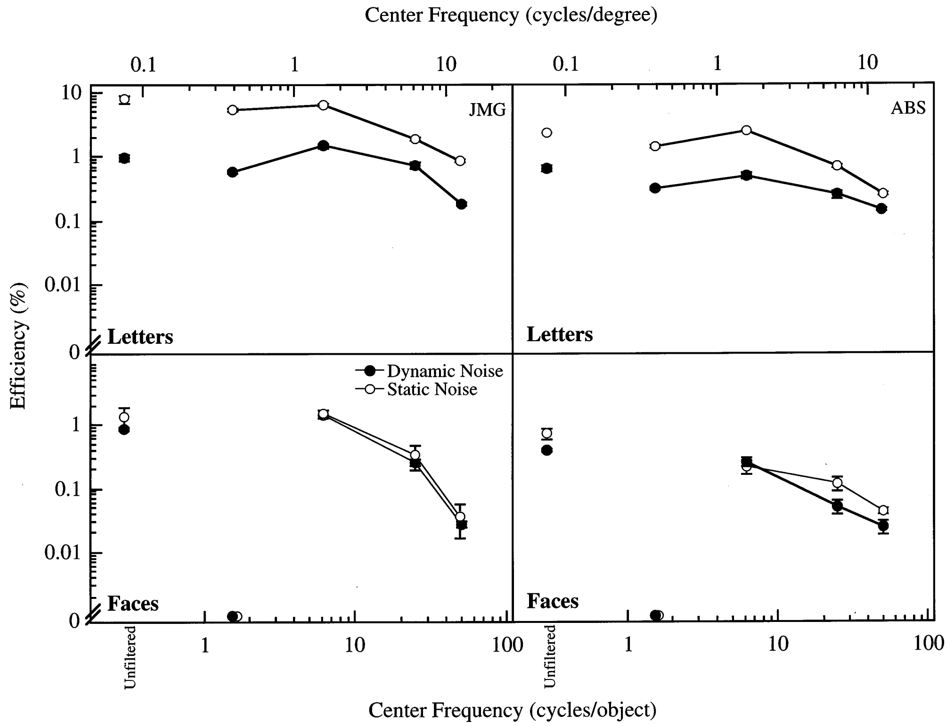


Fig. 8. Comparison of two-octave wide filtered letter (top panels) and face (bottom panels) identification efficiencies for two human observers using either dynamic (closed circles) or static (open circles) noise. Data are plotted as a function of the center frequency of the filter. Symbol conventions are the same as in Fig. 5.

The resulting letter identification efficiencies were nearly identical to those obtained with unfiltered noise. However, face identification efficiencies with band-pass noise increased by 0.2–0.4 log units in all but the unfiltered condition (where efficiency remained the same). Thus, the use of unfiltered noise in our original experiment appears to have contributed to low face identification efficiencies but not to low letter identification efficiencies. It is interesting to note that these results suggest that, in our conditions, human observers integrated information across spatial frequency differently for letters and faces. The fact that letter identification efficiency did not depend on the bandwidth of the noise suggests that human observers integrated information across a band of frequencies no wider than the stimulus bandwidth (i.e. 2 octaves). However, face identification efficiency was lower with unfiltered noise, suggesting that human observers integrated across a band that was wider than the stimulus bandwidth.

4. Discussion

The purpose of our study was to determine how human observers use information carried by different bands of spatial frequencies when identifying letters and faces. By comparing identification efficiencies for band-pass versions of two different kinds of complex patterns, we hoped to determine whether the way the visual system makes use of information across spatial scales depends upon the kind of pattern being identified.

The functions relating identification efficiency to spatial frequency for 1-octave band-pass filtered letters were relatively broadband in shape, declining by 0.4–0.7 log units below 2.2 c/letter and falling to 0 above 35.0 c/letter. In the middle range of frequencies, identification efficiency was relatively constant and approximately equal to the efficiency obtained with unfiltered stimuli (0.5–1%). Two-octave filtered letter identification efficiency functions were somewhat more band-pass in shape than the 1-octave functions, with efficiency peaking at 6.2 c/letter and gradually declining by 0.25–0.5 log units at the lowest and highest frequencies. For observers JMG and JMH, peak efficiencies for the 2–octave filtered letters were 0.15–0.3 log units higher than those obtained with the 1-octave filtered letters. We suspect that this small difference was due to practice effects because efficiencies with unfiltered stimuli, which were identical across experiments, also increased by similar amounts. Thus, peak identification efficiencies obtained with 1- and 2-octave filtered letters were similar to efficiencies obtained with unfiltered letters.

Identification efficiency for unfiltered faces was similar to that obtained with unfiltered letters (i.e. 0.5–

1.5%). However, face identification was impossible for both human observers in most of the 1-octave filtered conditions. Identification also was impossible in the lowest 2-octave filtered condition (i.e. a center frequency of 1.5 c/face). In the other 2-octave conditions, face identification efficiency peaked at a stimulus center frequency of 6.2 c/face and at a value that was similar to the one obtained with unfiltered faces, declining by 0.5–1.2 log units as the center frequency increased to 49.5 c/face.

Peak efficiency for letters and faces in the 2-octave condition occurred with band-pass stimuli centered at 6.2 c/obj. This spatial frequency falls within the critical letter identification band estimated by Solomon and Pelli (1994), borders the critical letter identification band estimated by Parish and Sperling (1991), and falls within several previous estimates of the critical face identification band (Harmon, 1973; Harmon & Julesz, 1973; Bachmann, 1991; Costen et al., 1994, 1996; Peli et al., 1994). However, absolute peak identification efficiencies were lower than values reported previously. Using dynamic noise appears to have contributed to low identification efficiencies for letters, but not for faces, in our experiment.

Our results thus reveal some similarities between letter and face identification efficiencies. First, the highest efficiencies obtained with filtered stimuli were similar for both kinds of patterns. Second, in the 2-octave tasks, peak efficiency occurred at the same frequency range (~ 6 c/obj). Finally, efficiencies were similar for unfiltered letters and faces. However, we also found several striking differences between letter and face identification efficiencies. Human observers were able to identify letters, but not faces, at the lowest spatial scale in the 2-octave filtered face task, and efficiency for 2-octave filtered faces fell off much more dramatically beyond 6 c/obj than for identically filtered letters. More dramatic differences were found in the 1-octave conditions, where human observers were able to identify 1-octave filtered letters, but not faces, at most spatial scales. Ideal observer analyses demonstrated that information always was available to perform the face identification task, and subsequent tests showed that neither the dynamic contrast range offered by our displays nor the use of external noise placed a ceiling upon performance in the conditions where faces were unidentifiable.

How do we interpret the differences found between letter and face efficiencies? One possibility is that higher-order mechanisms simply make different use of information across spatial frequencies for these two kinds of patterns. However, before accepting this conclusion, it is important to consider whether low-level constraints acting on both letters and faces affect discrimination information differently for the two classes of stimuli. If so, it may be unnecessary to posit different processes or mechanisms for letter and face identification.

In the following sections we describe three psychophysical experiments and a series of computer simulations that tested the effects of several low-level factors that we hypothesized could have contributed to inefficiencies in processing. Specifically, two of the experiments examined whether the differences between letter and face identification efficiencies were due to differences in either: (i) letter and face amplitude spectra; or (ii) letter and face detection efficiencies. A third experiment examined the possibility that observers had more difficulty learning to associate the filtered faces with their unfiltered counterparts in the response selection window. The simulations considered the impact of: (i) intrinsic position uncertainty; (ii) intrinsic size uncertainty; and (iii) the use of sub-ideal templates on cross-correlator performance. We hoped to address two related questions with these experiments and simulations. First, to what extent do these factors contribute to reduced identification efficiency? Second, can the effects of any of these factors account for the differences we found between face and letter identification efficiencies?

4.1. Differences in letter and face amplitude spectra

One difference between letters and faces is the manner in which energy is distributed across the various spatial frequencies and orientations: our set of letters had contrast power concentrated at the vertical and horizontal orientations, whereas contrast power was distributed more uniformly across orientations for faces. To determine whether these differences contributed to differences between letter and face identification efficiencies, we re-measured efficiencies with a set of hybrid images made from our original faces and letters. The hybrid stimuli were exactly the same as those used in the first experiment, except the average amplitude spectra were switched between the face and letter sets (recall that all faces shared a common averaged amplitude spectrum, as did the letters). Only the 2-octave wide filters were used, so a total of 100 new patterns were created (4 frequency bands $\times$ 20 amplitude and phase combinations $+$ 20 unfiltered images). All other testing conditions were the same as in the main experiment.

Ideal observer thresholds were re-calculated with the new hybrid stimuli and were almost identical to those obtained in the main experiment with the original images. Efficiencies for the hybrid and original stimuli for one observer (JMG) are shown in Fig. 9. Although there was a drop in absolute efficiency in all hybrid conditions, the effects of center spatial frequency on efficiency were similar to those obtained with the original images. Differences between face and letter amplitude spectra therefore cannot account for the differences between the shapes of the letter and face efficiency functions found in the main experiment.

4.2. Differences in detection efficiencies

Recent studies have suggested that, for some sets of stimuli, differences among identification efficiencies may be linked to differences among detection efficiencies (Braje et al., 1995; Tjan et al., 1995). Thus, we examined whether differences between letter and face identification efficiencies were related to differences between letter and face detection efficiencies. Detection efficiencies were measured for the original 2-octave wide band-pass filtered face and letter stimuli embedded in the same dynamic 2-D Gaussian noise used in the main experiment. Detection threshold was defined as 92% correct, which corresponds to the same value of d' (2) used in the identification tasks (Macmillan & Creelman, 1991). The procedure for human observers was the same as in the identification task, except for the following changes: (1) there were two 500 ms intervals separated by a 250 ms delay, during which the displays were set to average luminance and the fixation point reappeared in the center of the screen, and one interval was randomly selected to contain the signal + noise and the other noise alone; (2) The selection window consisted of the numbers '1' and '2' enclosed in separate boxes on either side of the location where the stimulus had previously appeared, and subjects were told to click the '1' box if the signal had appeared in the first interval and the '2' box if the signal has appeared in the second interval; (3) thresholds for both the human observers and the ideal observer were obtained using the QUEST procedure. At least two thresholds were averaged in each condition. The ideal observer's detection thresholds were estimated from 1000 simulated trials (see Appendix A for a description of the ideal decision rule).

Letter and face detection efficiencies were measured for three observers. Across all observers, the average difference between face and letter detection efficiencies in each condition was less than 0.1 log units and did not vary systematically across conditions. The failure to find consistent differences in detection efficiencies makes it unlikely that the differences between face and letter identification efficiencies were due solely to differences in stimulus detectability.

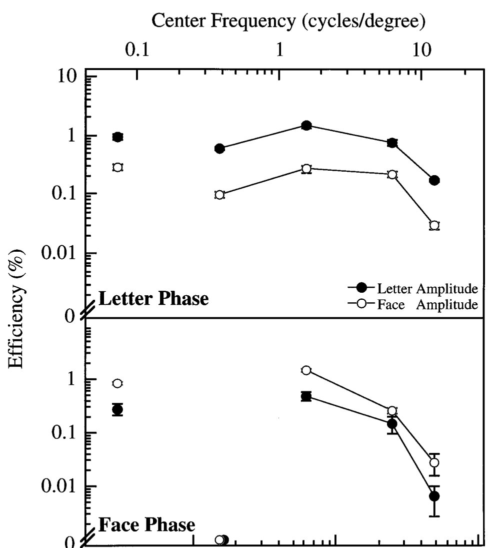


Fig. 9. Comparison of identification efficiencies for two-octave wide filtered hybrid and normal letters and faces for one human observer (JMG). The top panel illustrates efficiencies with the phase spectrum of letters and either the average amplitude spectrum of the faces (open symbols) or the average amplitude spectrum of the letters (closed symbols). The bottom panel illustrates efficiencies with the phase spectrum of faces and either the average amplitude spectrum of the faces (open symbols) or the average amplitude spectrum of the letters (closed symbols).

4.3. Differences in learning and similarity

Our procedure required observers to identify band-pass filtered stimuli by choosing one item from a set of unfiltered patterns. One obvious difference between letters and faces is that filtered and unfiltered letters look similar, whereas filtered and unfiltered faces look different (compare Figs. 2 and 3). This difference in similarity raises the possibility that it was easier for observers to learn the association between filtered and unfiltered letters than for filtered and unfiltered faces. In other words, differences between letter and face identification efficiency may reflect differences in response selection, rather than visual processing per se.

We tested this 'response selection hypothesis' by re-measuring 2-octave identification efficiencies for one observer (JMG) using a response window that contained thumbnail images that matched the filtered stimuli in each condition. We were unable to include the highest frequency condition because the thumbnail images in this condition were difficult to detect, even at the highest displayable contrast. However, in all other respects, the experiment was the same as the static noise experiment described above. Since the original experiment, observer JMG had viewed the letters and faces for many thousands of trials while participating in other experiments. Therefore, we re-measured efficiency in the original unfiltered condition to estimate the size of any practice effect. Letter identification efficiencies in the new and original experiments were essentially identical in all conditions: the differences ranged from 0.12 to -0.17 log units, and the mean difference was 0.0008 log units. Thus, there was no evidence of a practice effect for letter identification, and no evidence that using matched thumbnail images in the response selection window improved letter identification efficiency. Identification efficiency for unfiltered faces was 0.18 log units higher than in the original experiment. Efficiencies for the filtered faces also improved slightly: the differences ranged from 0.17 to 0.08 log units, and the mean difference was 0.13 log units. Thus, differences between efficiencies for 2-octave faces and letters seems not to be based on differences in response selection. The nearly uniform increase in efficiency in all conditions is consistent with the hypothesis that practice on the additional trials with faces produced a small, non-specific improvement in JMG's face identification efficiency. However, there was no evidence that using matched thumbnail images produced an additional improvement in performance in the band-pass filtered conditions. Similar results were found in a subsequent experiment with 1-octave filtered faces. These preliminary results suggest that it is unlikely that differences between letter and face identification efficiencies were due to observers having difficulty learning the associations between filtered and unfiltered versions of the patterns.

The fact that filtered letters look like unfiltered letters means that both the filtered and unfiltered stimuli would be familiar to the observers. Such is not the case with faces: in many of the conditions the filtered faces do not look like the unfiltered faces, so the filtered stimuli would not be familiar. Thus, one could argue that the reason identification efficiency is higher for filtered letters than for filtered faces is that the former are more familiar than the latter. We cannot rule out this possibility. However, it is worth noting that the greater similarity between filtered and unfiltered letters—and therefore the greater familiarity of the filtered letters—probably is caused by perceptual processes, and is not an artifact of our filtering procedures. In other words, despite the fact that filtered letters *look* more similar to the unfiltered letters, the cross-correlations[3] between 1- and 2-octave band-pass filtered stimuli and their unfiltered counterparts are nearly the same for faces and letters. In fact, the cross-correlations are slightly higher for faces in most conditions. Moreover, for both letters and faces the correlations are highest in the low spatial frequency conditions and decline monotonically with increasing frequency. Thus, the physical similarity between filtered and unfiltered items, at least as measured by cross-correlation, cannot account for the way identification efficiency varies across spatial scales, nor for the differences between faces and letters. Of course, the cross-correlation measure also does not account for the *perceived* similarity between filtered and unfiltered items.

4.4. Position and size noise

The first factors we considered using simulations were position and size noise. The ideal observer cross-correlates the stimulus with templates of appropriate position and size on every trial. If human observers are doing something like a template match, then it is reasonable to assume that the position and size of the templates humans use vary randomly across trials. In this section we examine the impact of these factors on human performance in our tasks.

We modeled the effect of each factor by using Monte Carlo simulations to measure cross-correlator thresholds in the 2-octave letter and face identification tasks with varying amounts of either position or size noise. Position noise was simulated by jittering the relative horizontal and vertical position of each template to the stimulus on each trial for each cross-correlation. The horizontal and vertical displacements were drawn from Gaussian distributions with S.D. of 1, 2, 3, and 5 pixels, and the size of the displacement was rounded to the nearest pixel. Size uncertainty was simulated by isotropically scaling (using nearest-neighbor interpolation) the relative size of each template to the stimulus on each trial for each cross-correlation. The magnitude of the size scaling was a random variable drawn from Gaussian distributions with S.D. of 2, 5, and 10% of the original template size.

Not surprisingly, the results of the simulations showed that position and size noise increased identification thresholds significantly at high spatial frequencies, but had very small effects at low frequencies. For example, position jitter of ± 1 pixel (i.e. ± 1.2 arcmin) had no effect on thresholds in all but the highest frequency condition, where efficiency fell by approximately 0.7 log units. Jitter of ± 5 pixels had no effect on thresholds in the 4.4 and 8.8 c/obj conditions, but made identification impossible at higher spatial scales. However, the changes in efficiency at higher frequencies were nearly identical for letters and faces. As with position jitter, size jitter of $\pm 2\%$ only increased thresholds in the three highest frequency conditions and jitter of ± 5 and $\pm 10\%$ increased thresholds in the four highest frequency conditions. Unlike position jitter, size jitter reduced efficiencies more dramatically for letters than for faces, the opposite of what we found with human observers. Thus, small amounts of position and size noise may contribute to the fall off in human efficiency at the highest spatial scales, but the simula-

[3] To calculate the correlation between a filtered item and its unfiltered counterpart, the contrast variance of both items was set to one, and the maximum value of the 2-D cross-correlation of the unfiltered image with the filtered image was computed. One consequence of averaging the amplitude spectra of the faces and letters was that the correlation between a filtered face and its unfiltered counterpart was constant across all faces in each condition. Likewise, the correlation was constant across all letters in each condition.

tions showed that the effects of position and size noise alone cannot account for the differences between letter and face identification efficiencies at either high or low spatial frequencies.

4.5. Sub-optimal templates

We next considered the idea that human observers are constrained by the kind of representation(s) they use to identify stimuli. For our task, the ideal observer cross-correlates stimuli with templates that exactly match the possible targets. Constraints on the kinds of templates that can be used by human observers should generally reduce efficiency. What is less clear is whether identical constraints on letter and face templates would introduce significant differences between letter and face identification efficiencies.

We first considered the possibility that human observers used band-pass filters that, unlike the ideal observer, were not matched precisely to the frequency content of the stimuli. For example, the bandwidths of the filters may have been too broad or too narrow, or they may not have been centered on the stimulus' center frequency. It is unlikely that this type of inefficiency can account for the differences that we observed between faces and letters. Using a filter that is broader or narrower than the ideal one, or is shifted to lower or higher frequencies, reduces the available identification information or introduces more noise and therefore elevates identification thresholds. In our conditions, the amount of available identification information (as indexed by the performance of the ideal observer) in 1- and 2-octave bands of spatial frequency changed slowly as a function of spatial frequency (see Figs. 4 and 6). Therefore, slight mismatches between stimulus and filter bandwidth and/or center frequency should decrease identification efficiency by nearly constant amounts at all spatial scales. We tested this idea by using Monte Carlo simulations to measure the performance of non-optimal cross-correlators in our experimental conditions. The cross-correlators used templates that were constructed by filtering the unfiltered versions of the stimuli with isotropic, zero phase shift, rectangular spatial frequency filters of various bandwidths and center frequencies. In all other aspects, the simulations were identical to the ones used to measure the performance of the ideal observer. As expected, mismatches between stimulus and template frequency spectra increased identification thresholds, but the effects were similar across experimental conditions when the mismatches were held constant on an octave scale. Thus, mismatches in bandwidth and center frequency on the order of 1–2 octaves simply reduced efficiency uniformly across spatial scale. Although such mismatches in filter shape might contribute to low overall efficiency in human observers, they still cannot account for the fall off in efficiency at very low spatial frequencies in the 2-octave conditions, or the dramatic differences between faces and letters in the 1-octave conditions.

We next considered the possibility that human observers used templates corresponding to the unfiltered stimuli in all stimulus conditions. Why might this be so? Recall that in our tasks the response window always contained icons of the unfiltered stimuli, and therefore human observers were asked to match unfiltered versions of the stimuli to the filtered images. This response arrangement may have biased observers to use representations of unfiltered items. One line of evidence against this idea comes from the experiment that found that identification efficiency for 2-octave filtered letters was the same with unfiltered noise and with 2-octave band-pass noise. That result suggests that human observers based their responses on signals restricted to the pass-band of the stimulus, rather than the entire bandwidth of unfiltered letters. Another piece of evidence that suggests that human observers did not use unfiltered templates is that, for both letters and faces, peak efficiency in the filtered stimuli conditions was the same as efficiency in the unfiltered condition. If human observers always used unfiltered templates, then efficiency should have been significantly greater in the unfiltered stimulus condition. Thus, it is unlikely that differences between letter and face identification efficiencies were caused by human observers using unfiltered templates to encode all of the stimuli.

Do observers use a fixed, band-pass channel to identify letters and faces? If a fixed channel has a bandwidth that is narrow relative to the stimulus bandwidth, then shifting the stimulus center frequency should have a large effect on the amount of information passed by the filter. Consequently, one would expect identification efficiency to vary dramatically with stimulus center frequency. If the channel bandwidth is significantly greater than the stimulus bandwidth, then shifting the stimulus center frequency should have a small effect on the amount of information passed by the filter. Therefore, an observer that uses a fixed, broad-band channel should have nearly constant efficiency across a sizable range of spatial scales. Observers using channels with intermediate bandwidths would have efficiencies that fall somewhere in-between these two extremes. For letters, efficiencies in the 1- and 2-octave conditions are relatively constant across a wide range of spatial scales, a result that is consistent with the hypothesis that observers use a fixed, broad-band channel. However, the results from the band-pass noise experiment showed that observers integrated information across a band of frequencies no wider than 2-octaves when identifying letters. Thus, observers must have used a relatively narrow-band channel whose center frequency was adjusted to match that of the stimulus to identify letters. For faces, efficiency in the 2-octave condition fell off as stimulus center frequency increased beyond 6.2 c/obj. Furthermore, the filtered noise experiment showed that observers were using a channel whose bandwidth was greater than the 2-octave stimulus bandwidth when identifying faces. Both results are qualitatively consistent with the hypothesis that observers use a fixed channel of intermediate bandwidth (i.e. greater than 2 octaves) to identify faces. However, the results with 1-octave filtered faces are inconsistent with this hypothesis, because observers could only identify filtered faces within a very narrow frequency range.

Monte Carlo simulations were used to further examine the possibility that face identification in human observers was based on the response of a fixed spatial frequency channel. Computer simulations calculated the performance of a fixed-channel observer that uses an ideal decision rule to identify faces that have been passed through a noisy, linear spatial frequency filter (Braje et al., 1995). In our simulations, the input stimulus, consisting of the face and external noise, was passed through a spatial frequency filter. Internal Gaussian white noise was added after the filtering operation[4], and the simulated observer then compared the filtered stimulus to a set of filtered templates using the ideal decision rule derived by Braje et al. (1995). The modulation transfer function of the filter was a Gaussian function of the logarithm of spatial frequency. Filters with center frequencies of 4.6, 9.3, and 18.6 c/obj and bandwidths (at half-height) of ± 0.5, ± 1, and ± 1.5 octaves were used in separate simulations. The contrast variance of the internal Gaussian noise was varied from 0.007 to 0.2 in separate simulations. When the filter bandwidth was ± 0.5 octaves, the functions relating efficiency and stimulus center frequency all had prominent peaks. For example, in the 1-octave conditions, when the filter was centered on 9.3 c/obj and the variance of the internal noise was 0.02, efficiency was 13, 64, and 34% for stimulus center frequencies of 4.4, 8.8, and 17.5 c/obj 1-octave conditions, respectively, and essentially 0 with other center frequencies. In the 2-octave conditions, efficiency was 55 and 16% for stimuli with center frequencies of 6.2 and 24.8 c/obj, respectively, and essentially 0 with other center frequencies. Efficiency in the unfiltered condition was 21%. Changing the amount of internal noise simply shifted efficiencies by approximately the same proportion in all conditions. Changing the filter's center frequency to 4.6 or 18.6 c/obj simply shifted the location of peak efficiency to 4.6 or 18.6 c/obj, and lowered efficiency in the unfiltered condition. Finally, increasing the filter's bandwidth to ± 1.5 octaves resulted in significantly flatter efficiency versus stimulus frequency curves: going from the lowest to the highest stimulus center frequencies, efficiency was 12, 30, 66, 75, 74, 46, and 10% in the 1-octave conditions and 1, 68, 50, and 35% in the 2-octave conditions. Efficiency in the unfiltered condition was 50%.

If human observers used a fixed band-pass filter to identify faces, and if all other constraints on face identification do not depend on the spatial frequency content of the stimulus, then human efficiencies should be qualitatively similar to those obtained with a simulated fixed-band observer. More specifically, we would expect human efficiencies to be a constant fraction of the fixed-band observer's efficiencies. The two human observers could identify 1-octave filtered faces only in the 8.8 c/obj condition, and one observer also was able to perform the task in the 17.5 c/obj condition. The fixed-band observer is most efficient in those conditions and performs very poorly in the other conditions when the filter has a bandwidth of ± 0.5 octaves and is centered on 9.3 c/obj. If we assume that human efficiency is a small fraction of the fixed-band observer's, then we could construct a model that correctly predicts that identification would be possible only in the 8.6 and 17.5 c/obj conditions. However, such a model makes some important incorrect predictions. First, it predicts that peak efficiency in the band-pass conditions should be higher than efficiency in the unfiltered condition. Second, it predicts that in the 2-octave conditions, human observers should be unable to identify faces in the highest frequency condition. To account for the results in the 2-octave conditions, the model would have to be changed by assuming that the internal filter has a broader bandwidth (e.g. ± 1.5 octaves). However, a model with a broader bandwidth incorrectly predicts that face identification should be possible in many of the 1-octave conditions. No single combination of internal noise, filter center frequency, or filter bandwidth mimicked human performance in all conditions. The simulations therefore suggest that either human observers did not base face identification on the responses of a fixed channel, or that the effects of other constraints that were not included in our simulated observer varied across conditions.

4.6. Spatial sampling

So far, we have considered how letter and face identification are affected by intrinsic uncertainty about stimulus position and size, and by changes in the spatial frequency tuning of the filters underlying identification. Our simulations indicate that these factors have similar effects on the amount of available identification information for faces and letters, and therefore cannot ac-

[4] For the Gaussian frequency filters used in our simulations, adding internal noise is necessary because filtering simply attenuates the signal and noise at each frequency component by equal amounts, and therefore does not change the signal-to-noise ratio. Hence, the performance of the narrow-band observer is equivalent to that of the ideal observer when the internal noise is zero.

count for the dramatic differences between face and letter identification efficiencies found in human observers. Thus, if these factors are the primary constraints on human performance, then it is necessary to assume that they differ significantly for letter and face stimuli. For example, one could construct a model in which spatial filters that differed in bandwidth, center frequency, and internal noise were used to identify letters and faces. Although such a scheme may account for the major features of the data, it is unsatisfying theoretically because frequency-selective filters generally are thought to be part of the initial, bottom-up processing of the retinal image, and the operation of such mechanisms is thought to be independent of stimulus type. Thus, we have searched for other factors that, when applied in the same way, constrain performance in face and letter identification in different ways. In this section we consider one such factor, namely incomplete spatial sampling of the stimulus.

The ideal observer uses information available at every pixel to identify stimuli. One way of modeling low identification efficiency by human observers is to assume that they use the available information only at a subset of pixels (Tjan et al., 1995). The effects of selective spatial sampling on identification were examined by programming a simulated observer that was identical to the ideal observer, except that most of the pixels in the ideal templates were set to 0. In other words, only a small subset of stimulus pixels affected the performance of this simulated observer.

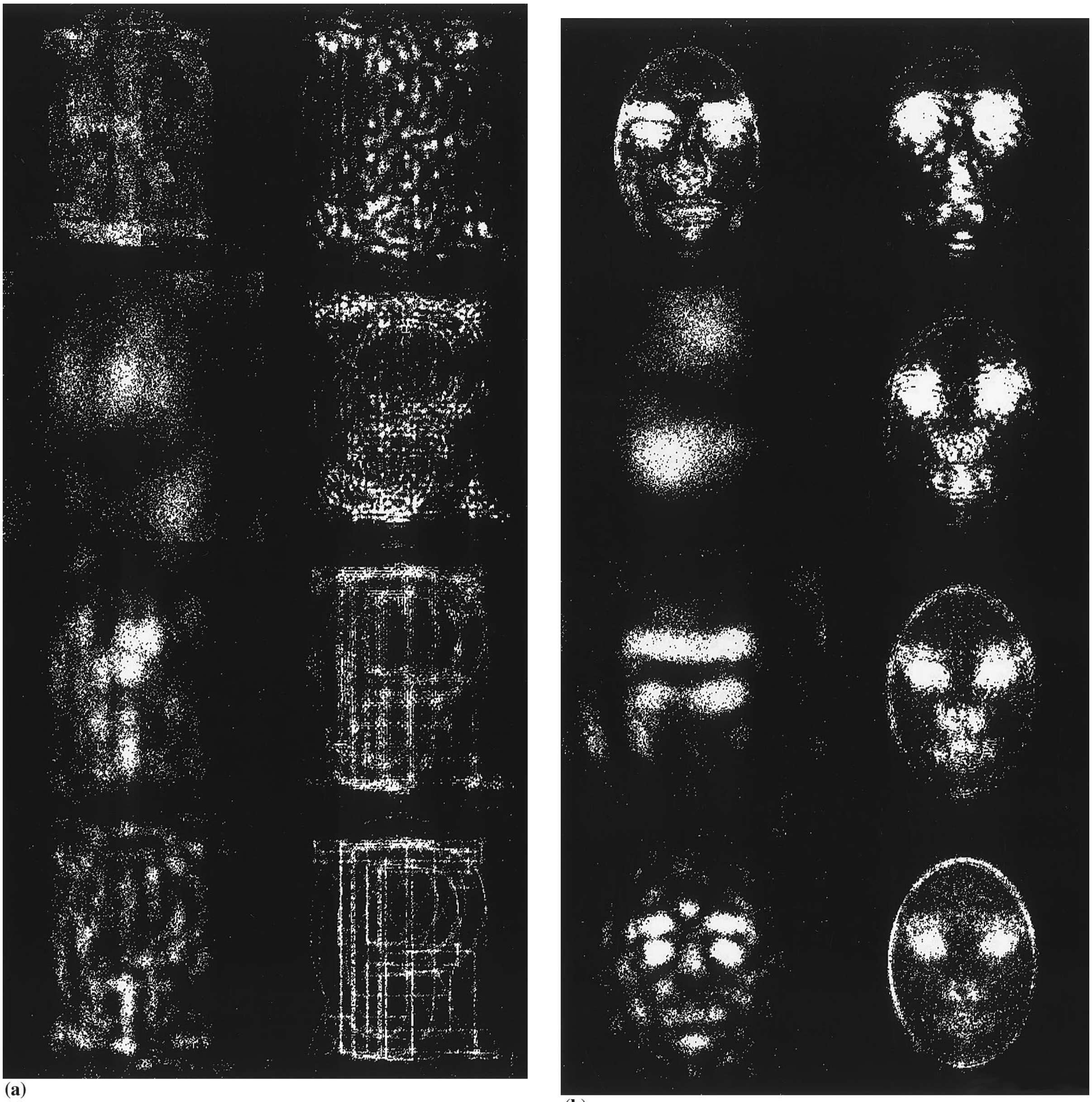
Our working hypothesis was that human observers would use the most informative pixels, and therefore it was necessary to estimate the amount of information carried by each pixel. The amount of information was estimated by calculating the performance of an ideal observer that identified a stimulus on the basis of the contrast value at a single pixel. In these single-pixel simulations, the noise was identical to the noise used in the ideal observer simulations, but stimulus contrast in each condition was set to 4 log units above the ideal observer's identification threshold. On each simulated trial, a target letter or face was selected randomly, noise was added to the contrast value at a single pixel for that stimulus, and the simulated observer had to identify the stimulus on the basis of that single contrast value. The pixel's coordinates and the signal-plus-noise distributions for all items were known to the observer, and the item that maximized the a posterior probability of receiving that contrast value was selected. Five hundred simulated trials were run at each pixel, and the proportion of correct responses, which ranged from approximately 0.1 to approximately 0.7, were used to rank order the pixels in terms of the information available for identification. Fig. 10 shows the top 10% most informative pixels in the 1-octave letter and face conditions: each pixel that was in the top 10% is set to 1 (i.e. white), and the remaining pixels are 0 (i.e. black). For faces, the 10% most informative pixels were clustered near the eyes, nose, and mouth. Although it is not obvious from Fig. 10, the pixels near the eyes carried the most information, followed by pixels near the nose and mouth[5].

To simulate the effects of selective spatial sampling, the ideal observer's templates for each condition were multiplied by matrices like those illustrated in Fig. 10. The result of this multiplication was a template that was 0 everywhere except at the most informative pixels, where the values were the same as in the ideal template. These templates then were used in simulations to estimate identification thresholds for the selective spatial sampling observer. Finally, estimates of efficiency were calculated from the ratio of thresholds for the ideal and spatial sampling observers. Efficiencies for both letters and faces increased monotonically with stimulus center frequency. However, the rate of increase was the same for both letters and faces. As the percentage of pixels used increased, identification efficiencies increased nearly uniformly in all conditions.

Clearly, the pattern of efficiencies for the spatial sampling observer differs from the one obtained from human observers. Therefore, if human observers do engage in selective spatial sampling, they do not use the most informative pixels in each condition. This conclusion may not be surprising, because the amount of experience that observers received with the band-pass letters and faces was quite limited (approximately 120 trials per threshold), at least relative to the experience they have had with unfiltered versions of those stimuli. Moreover, the testing conditions may not have been optimal for learning which pixels were most informative. Recall that our simulated single-pixel observer was presented with stimuli that were substantially above threshold. The elevated contrast was necessary because each pixel carries a very small fraction of the total stimulus information. Hence, when stimulus contrast was set to the ideal observer's threshold, performance on any single pixel, even the most informative one, was only slightly better than chance. When testing human observers, the stimulus contrast on most trials was near identification threshold, and so may not have been high enough for people to determine accurately which spatial locations carried the most information.

If human observers did not sample the most informative pixels, then which ones did they choose? One study has suggested that a single band-pass channel centered between 3 and 8 c/obj mediates identification of both unfiltered letters and faces (Majaj et al., 1998). Perhaps human observers have learned a spatial sampling strategy that would be optimal for that channel, and then

[5] Note that the set of n most informative pixels (as defined by our procedure) will not necessarily be the most informative set of n pixels.



(a) (b)

Fig. 10. The top 10% most informative pixels in each condition of the 1-octave (a) letter and (b) face identification tasks. For both figures, the top leftmost panel corresponds to the unfiltered condition. The center frequencies of the stimuli for the remaining panels are (from top to bottom): in the left column, 1.0, 2.2, and 4.4 c/obj; in the right column, 8.8, 17.5, 35.0 and 70.0 c/obj. See text for more details.

used that strategy in *all* of our conditions. Although speculative, this suggestion is plausible because: (i) human observers have considerable experience identifying letters and frontal views of faces; (ii) the most informative pixels in this range correspond to the eyes, nose, and mouth, and these features are thought to be important for face recognition; and (iii) as stated above, there was little opportunity to learn new sampling strategies. For these reasons, we estimated the performance of a selective-sampling observer that used only a fraction of the most informative pixels drawn from the 2-octave, 6.2 c/obj filtered letters and faces in all conditions. We chose to use the top 2% most informative pixels because these pixels were all clustered around the eyes in the face stimuli. We also included position jitter (± 1 pixel) to model the fall-off in efficiency found for human observers at the higher frequencies.

The estimated efficiencies for the 1- and 2-octave stimulus conditions are shown in Fig. 11. Several aspects of these results are noteworthy. First, the drop in efficiency at low frequencies was especially pronounced in faces. In fact, face identification efficiency was essentially 0 in the 1.1, 2.2, and 4.4 c/obj 1-octave conditions and in the 1.5 c/obj 2-octave condition. Second, efficiencies for both letters and faces fell off at the highest frequencies in both the 1-octave and 2-octave tasks. For letters, efficiency was 0 in the 70 c/obj 1-octave condition. Separate simulations showed that the fall off in efficiency at low spatial frequencies was due entirely to the effects of selective spatial sampling, whereas the fall off at high spatial frequencies was due entirely to the effects of position jitter. Third, efficiency in the unfiltered condition was within 0.1–0.2 log units of peak efficiency in the band-pass conditions for both letters and faces. Fourth, the efficiencies obtained with both letters and faces did not depend significantly on stimulus bandwidth: for both types of patterns, identification efficiencies were similar in the 1- and 2-octave conditions. All of these results are qualitatively consistent with those obtained from human observers. In particular, the selective spatial sampling observer, like human observers, exhibits large differences between letter and face efficiencies at low frequencies.

However, there also are discrepancies between the simulated and real efficiencies. The most obvious difference is that the selective spatial sampling observer is two to five times more efficient at identifying faces than letters, whereas human observers were about equally efficient with both types of stimuli. Another difference is that the selective sampling observer is capable of identifying faces even at the highest spatial scales, whereas humans were unable to identify letters and faces in the 70 c/obj 1-octave condition, and faces in the 35 c/obj 1-octave condition. Finally, in our simulations the high frequency fall-off in efficiency for the selective spatial-sampling observer in the 2-octave conditions was similar for letters and faces, whereas the decrease in efficiency at high frequencies for human observers was more dramatic for faces than for letters. Thus, selective spatial sampling provides a qualitative account for some, but not all, aspects of our data.

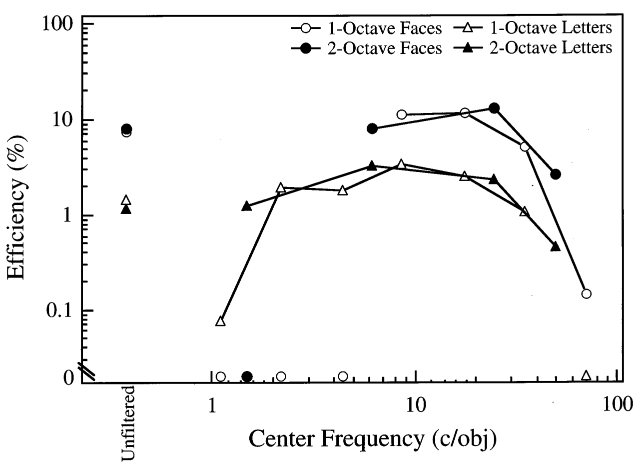


Fig. 11. One- and 2-octave wide filtered letter and face identification efficiencies for a simulated observer that used only the top 2% most informative pixels from the 2-octave, 6.2 c/obj condition across all conditions, plotted as a function of the center frequency of the filter. There was also ± 1 pixel of random position jitter. Open symbols show the performance in the 1-octave conditions, closed symbols the 2-octave conditions. Circles represent face efficiencies, triangles letter efficiencies. Plotting conventions are the same as in Fig. 5.

5. Conclusion

We found that the ability of human observers to make use of information across spatial scales differs for letters and faces. We ruled out several potential explanations for the differences, including: differences in global amplitude spectra; differences in detection efficiency; differences in the ability to learn the association between filtered and unfiltered patterns; the use of a single, fixed, band-pass channel; spatial uncertainty; and size uncertainty. However, selective spatial sampling did provide a qualitative account for many aspects of our data. Some of the quantitative discrepancies between spatial sampling and human observers might reflect non-optimal sampling strategies used by humans. For example, human observers may sample faces in a way that supports both identification of the face and recognition of the face's emotional expression. Recognition of emotional expression depends strongly on an area centered on the mouth (Bayer et al., 1998), but in our tasks the most informative pixels for identification were clustered near the eyes. Thus, a sampling strategy that diverted some samples from the eyes to the mouth would lower identification efficiency. In any case, the goal of these simulations was not to outline a quantitative model of letter and face identification. Instead, the goal was to highlight factors that ultimately might be incorporated into a quantitative model. Our results suggest that it would be fruitful to investigate which spatial sampling strategies are used by human observers.

Acknowledgements

We would like to thank Bosco Tjan and Richard Murray for help with the ideal observer analyses. We would also like to thank Denis Pelli and two anonymous reviewers for helpful comments on an earlier version of this manuscript. Parts of the above research were submitted to the University of Toronto in partial fulfillment of a M.A. degree in Psychology (Gold, 1996) and presented at the Association for Research in Vision and Ophthalmology (Gold, Sekuler & Bennett, 1997). This research was supported by NSERC grants OGP105494 and OGP0042133, the German Academic Exchange Council, and the Alexander von Humboldt Foundation.

Appendix A

The ideal rule for maximizing percent correct in a two-interval forced-choice detection task limited by white noise is to choose the interval that yields the highest a posteriori probability of having contained the signal (Green & Swets, 1966; Braje et al., 1995; Tjan et al., 1995). For our task, the ideal observer does not know which of the ten possible stimuli appeared within one of the intervals, and therefore must compute the a posteriori probabilities that each of the ten possible images appeared within each of the two intervals. It also must take into account the a posteriori probabilities that each interval contained noise alone. Formally, the decision rule may be expressed as:

$$P = \frac{\sum_{j=1}^{10} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(R_i - T_{ij})^2\right]\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(L_i^2)\right]}{\sum_{j=1}^{10} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(L_i - T_{ij})\right]\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(R_i^2)\right]}$$

where

R = the data in interval one, represented as contrast values,

L = the data in interval two, represented as contrast values,

T_{ij} = the value at pixel i in the jth template (image), represented as contrast values,

n = the number of pixels in each image (256×256),

σ = the standard deviation of the noise contrast.

On a given trial, the ideal observer chooses the first interval if P is greater than 1, and the second interval if P is less than 1. Because all of our stimuli share a common amplitude spectrum,

$$\sum_{i=1}^{n} T_{ij}^2$$

will be the same for all of the images in a set, and can be removed from the ratio. The ideal rule can then be simplified to:

$$P = \frac{\sum_{j=1}^{10} \exp\left[\sum_{i=1}^{n} R_i T_{ij}\right]}{\sum_{j=1}^{10} \exp\left[\sum_{i=1}^{n} L_i T_{ij}\right]}$$

Thus the ideal decision rule for our task and stimuli is to pick the interval that yields the highest sum of the exponentiated cross-correlations between the data and the templates.

References

Alexander, K. R., Xie, W., & Derlacki, D. J. (1994). Spatial frequency characteristics of letter identification. *Journal of the Optical Society of America A*, *11*(9), 2375–2382.

Bachmann, T. (1987). Different trends in perceptual pattern microgenesis as a function of the spatial range of local brightness averaging. Inaugural Meeting of the European Society for Cognitive Psychology, (1985 Nijmegen, Netherlands). *Psychological Research*, *49*(2–3), 107–111.

Bachmann, T. (1991). Identification of spatially quantised tachistoscopic images of faces: how many pixels does it take to carry identity? Special issue: face recognition. *European Journal of Cognitive Psychology*, *3*(1), 87–103.

Barlow, H. B. (1978). The efficiency of detecting changes of density in random dot patterns. *Vision Research*, *18*(6), 637–650.

Bayer, H., Schwartz, O., & Pelli, D. (1998). Recognizing facial expressions efficiently (ARVO supplement). *Investigative Ophthalmology and Visual Science*, *39*(4), 172.

Baylis, G. C., Rolls, E. T., & Leonard, C. M. (1985). Selectivity between faces in the responses of a population of neurons in the cortex in the superior temporal sulcus of the monkey. *Brain Research*, *342*(1), 91–102.

Beckman, P. J., & Legge, G. E. (1994). A sequential ideal–observer model of letter identification in central and peripheral vision (ARVO supplement). *Investigative Ophthalmology and Visual Science*, *35*(4), 3212.

Benton, A. L. (1980). The neuropsychology of facial recognition. *American Psychologist*, *35*(2), 176–186.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 443–446.

Braje, W. L., Tjan, B. S., & Legge, G. E. (1995). Human efficiency for recognizing and detecting low-pass filtered objects. *Vision Research*, *35*(21), 2955–2966.

Burns, C. W., Farrel, B., Moore, D. M., & Pelli, D. G. (1995). Alphabet complexity and letter identification efficiency (ARVO supplement). *Investigative Ophthalmology and Visual Science*, *36*(4), 4170.

Cannon, M. W., Hoffmeister, J. W., & Fullenkamp, S. C. (1997). The spatial frequency bandwidth for aircraft identification (ARVO supplement). *Investigative Ophthalmology and Visual Science*, *38*(4), 2997.

Chung, S. T. L., & Legge, G. E. (1997). Spatial-frequency dependence of letter recognition in central and peripheral vision (supplement). *Investigative Ophthalmology and Visual Science*, *38*(4), 2992.

Costen, N. P., Parker, D. M., & Craw, I. (1994). Spatial content and spatial quantisation effects in face recognition. *Perception*, *23*(2), 129–146.

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*(4), 602–612.

Damasio, A. R. (1985). Prosopagnosia. *Trends in Neurosciences*, *8*(3), 132–135.

Damasio, A. R., Tranel, D., & Damasio, H. (1990). Face agnosia and the neural substrates of memory. *Annual Review of Neuroscience*, *13*, 89–109.

de Haan, E. H., Young, A. W., & Newcombe, F. (1992). Neuropsychological impairment of face recognition units. *Quarterly Journal of Experimental Psychology Human Experimental Psychology*, *44A*(1), 141–175.

De Valois, K., & De Valois, R. (1988). *Spatial vision*. New York: Oxford Press.

Desimone, R. (1991). Face-selective cells in the temporal cortex of monkeys. Special issue: face perception. *Journal of Cognitive Neuroscience*, *3*(1), 1–8.

Desimone, R., Albright, T. D., Gross, C. G., & Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *Journal of Neuroscience*, *4*(8), 2051–2062.

Desimone, R., & Schein, S. J. (1987). Visual properties of neurons in area V4 of the macaque: sensitivity to stimulus form. *Journal of Neurophysiology*, *57*(3), 835–868.

Ellis, H. D., & Young, A. D. (1989). Are faces special? In *The handbook of research on face perception* (pp. 2–26). The Netherlands: Elsevier Science Publishers.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A*, *4*(12), 2379–2394.

Field, D. J. (1993). Scale-invariance and self-similar 'wavelet' transforms: an analysis of natural scenes and mammalian visual systems. In M. Farge, J. C. R. Hunt, & J. C. Vassilicos, *Wavelets, fractals and Fourier transforms* (pp. 151–193). Oxford: Oxford University Press.

Fiorentini, A., Maffei, L., & Sandini, G. (1983). The role of high spatial frequencies in face perception. *Perception*, *12*(2), 195–201.

Geisler, W. S. (1989). Sequential ideal-observer analysis of visual discriminations. *Psychological Review*, *96*(2), 267–314.

Ginsburg, A.P. (1978). *Visual information processing based on spatial filters constrained by biological data*. Ph.D. Thesis, Cambridge University.

Ginsburg, A. P. (1980). Specifying relevant spatial information for image evaluation and display design: an exploration of how we see certain objects. *Proceeding of the SID*, *21*, 219–227.

Ginsburg, A. P. (1984). The spatial bandwidth for the identification of geometric form (ARVO supplement). *Investigative Ophthalmology and Visual Science*, *25*(3), 97.

Gold, J. (1996). *The role of spatial frequency information in the identification of complex visual patterns*. MA Thesis, University of Toronto.

Gold, J., Sekuler, A. B., & Bennett, P. J. (1997). The effect of band-pass filtering on the efficiency of complex pattern identification (ARVO supplement). *Investigative Ophthalmology and Visual Science*, *38*(4), 2996.

Graham, N. (1989). *Visual pattern analyzers*. New York: Oxford University Press.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: John Wiley and Sons.

Gross, C. G., Rocha-Miranda, C. E., & Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the Macaque. *Journal of Neurophysiology*, *35*(1), 96–111.

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

Harmon, L. D., & Julesz, B. (1973). Masking in visual recognition: effects of two-dimensional filtered noise. *Science*, *180*(91), 1194–1197.

Hay, D. C., & Young, A. W. (1982). The human face. In A. W. Ellis, *Normality and pathology in cognitive functions*. Hillsdale, NJ: Erlbaum.

Hayes, T., Morrone, M. C., & Burr, D. C. (1986). Recognition of positive and negative band-pass filtered images. *Perception*, *15*(5), 595–602.

Heywood, C. A., & Cowey, A. (1992). The role of the 'face-cell' area in the discrimination and recognition of faces by monkeys. *Philosophical Transcript of the Royal Society of London B*, *335*, 31–38.

Kersten, D. (1987). Statistical efficiency for the detection of visual noise. *Vision Research*, *27*(6), 1029–1040.

Kersten, D. (1990). Statistical limits to image understanding. In C. Blakemore, *Vision: coding and efficiency* (pp. 32–44). Cambridge: Cambridge University Press.

Legge, G. E., Pelli, D. G., Rubin, G. S., & Schleske, M. M. (1985). Psychophysics of reading: I. Normal vision. *Vision Research*, *25*(2), 239–252.

Liu, Z., Knill, D. C., & Kersten, D. (1995). Object classification for human and ideal observers. *Vision Research*, *35*(4), 549–568.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: a user's guide* (1st ed.). Cambridge: Cambridge University Press.

Majaj, N., Palomares, M., Mouchraud, C., Kotlyarenko, D., & Pelli, D. (1998). Identifying letters—bottom up (ARVO supplement). *Investigative Ophthalmology and Visual Science*, *39*(4), 408.

Morrone, M. C., Burr, D. C., & Ross, J. A. (1983). Added noise restores recognizability of coarse quantized images. *Nature*, *305*(5931), 226–228.

Norman, J., & Ehrlich, S. (1987). Spatial frequency filtering and target identification. *Vision Research*, *27*(1), 87–96.

Parish, D. H., & Sperling, G. (1991). Object spatial frequencies, retinal spatial frequencies, noise, and the efficiency of letter discrimination. *Vision Research*, *31*(7–8), 1399–1415.

Parker, D. M., Lishman, J. R., & Hughes, J. (1992). Temporal integration of spatially filtered visual images. *Perception*, *21*(2), 147–160.

Parker, D. M., Lishman, J. R., & Hughes, J. (1996). Role of coarse and fine spatial information in face and object processing. *Journal of Experimental Psychology*: *Human Perception and Performance*, *22*(6), 1448–1466.

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*(7), 1929–1939.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, *10*, 437–442.

Perrett, D. I., Mistlin, A. J., & Chitty, A. J. (1987). Visual neurones responsive to faces. *Trends in Neurosciences*, *10*(9), 358–364.

Perrett, D. I., Mistlin, A. J., Chitty, A. J., & Smith, P. A. (1988). Specialized face processing and hemispheric asymmetry in man and monkey: evidence from single unit and reaction time studies. IBBS Workshop on: hemispheric specialization and interhemispheric communication (1986, Rotterdam, Netherlands). *Behavioural Brain Research*, *29*(3), 245–258.

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

Riley, D., & Costall, A. (1980). Comments on 'Recognition of faces in the presence of two-dimensional sinusoidal masks' by Tieger and Ganz. *Perception and Psychophysics*, *27*(4), 373–374.

Rubin, G. S., & Siegel, K. (1984). Recognition of low-pass faces and letters (ARVO supplement). *Investigative Ophthalmology and Visual Science*, *25*(3), 96.

Schuchard, R. A., & Rubin, G. S. (1989). Face identification of band-pass filtered faces by low vision observers (ARVO supplement). *Investigative Ophthalmology and Visual Science*, *30*(3), 396.

Schwartz, O., Bayer, H., & Pelli, D. (1998). Features, frequencies, and facial expressions (ARVO supplement). *Investigative Ophthalmology and Visual Science*, *39*(4), 173.

Schyns, P. G., & Olivia, A. (1994). From blobs to boundary edges: evidence for time- and spatial-scale-dependant scene recognition. *Psychological Science*, *5*(4), 195–200.

Sergent, J. (1989). Structural processing of faces. In *The handbook of research on face perception* (pp. 2–26). The Netherlands: Elsevier Science Publishers.

Sergent, J., Ohta, S., & MacDonald, B. (1992). Functional neuroanatomy of face and object processing. *Brain*, *115*, 15–36.

Solomon, J. A., & Pelli, D. G. (1994). The visual filter mediating letter identification. *Nature*, *369*(6479), 395–397.

Tieger, T., & Ganz, L. (1979). Recognition of faces in the presence of two-dimensional sinusoidal masks. *Perception and Psychophysics*, *26*(2), 163–167.

Tjan, B. S., Braje, W. L., Legge, G. E., & Kersten, D. (1995). Human efficiency for recognizing 3-D objects in luminance noise. *Vision Research*, *35*(21), 3053–3069.

Tyler, C. W., Chan, H., Liu, L., McBride, B., & Kontsevich, L. (1992). Bit-stealing: how to get 1786 or more grey levels from an 8-bit color monitor. In B. E. Rogowitz, *Human vision*, *visual processing*, *and digital display III*. *Proceedings SPIE*, *1666*, 351–364.

Watson, A. B., & Pelli, D. G. (1983). QUEST: a Bayesian adaptive psychometric method. *Perception and Psychophysics*, *33*(2), 113–120.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, *81*, 141–145.

Yin, R. K. (1970). Face recognition by brain-injured patients: a dissociable ability? *Neuropsychologia*, *8*(4), 395–402.