

## OBJECT SPATIAL FREQUENCIES, RETINAL SPATIAL FREQUENCIES, NOISE, AND THE EFFICIENCY OF LETTER DISCRIMINATION

DAVID H. PARISH and GEORGE SPERLING\*

Human Information Processing Laboratory, Department of Psychology and Center for Neural Sciences,  
New York University, NY 10003, U.S.A.

(Received 7 July 1988; in revised form 2 June 1990)

**Abstract**—To determine which spatial frequencies are most effective for letter identification, and whether this is because letters are objectively more discriminable in these frequency bands or because can utilize the information more efficiently, we studied the 26 upper-case letters of English. Six two-octave wide filters were used to produce spatially filtered letters with 2D-mean frequencies ranging from 0.4 to 20 cycles per letter height. Subjects attempted to identify filtered letters in the presence of identically filtered, added Gaussian noise. The percent of correct letter identifications vs  $s/n$  (the root-mean-square ratio of signal to noise power) was determined for each band at four viewing distances ranging over 32:1. Object spatial frequency band and  $s/n$  determine *presence of information* in the stimulus; viewing distance determines retinal spatial frequency, and affects only *ability to utilize*. Viewing distance had no effect upon letter discriminability: object spatial frequency, not retinal spatial frequency, determined discriminability. To determine discrimination efficiency, we compared human discrimination to an ideal discriminator. For our two-octave wide bands,  $s/n$  performance of humans and of the ideal detector improved with frequency mainly because linear bandwidth increased as a function of frequency. Relative to the ideal detector, human efficiency was 0 in the lowest frequency bands, reached a maximum of 0.42 at 1.5 cycles per object and dropped to about 0.104 in the highest band. Thus, our subjects best extract upper-case letter information from spatial frequencies of 1.5 cycles per object height, and they can extract it with equal efficiency over a 32:1 range of retinal frequencies, from 0.074 to more than 2.3 cycles per degree of visual angle.

Spatial filtering    Scale invariance    Psychophysics    Contrast sensitivity    Acuity

### INTRODUCTION

#### *Characterizing objects*

When we view objects, what range of spatial frequencies is critical for recognition, and how is our visual system adapted to perceive these frequencies? Ginsburg (1978, 1980) was among the first to investigate this problem by means of spatial bandpass filtered images of faces and lowpass filtered images of letters. He noted the lowest frequency band for faces and the cutoff frequency for letters at which the images seemed to him to be clearly recognizable. The cutoff frequency for letters was 1–2 cycles per letter width; faces were best recognized in a band centered at 4 cycles per face width. He also proposed that the perception of geometric visual illusions, such as the Müller-Lyer and Poggen-dorf, was mediated by low spatial frequencies (Ginsberg, 1971, 1978; Ginsberg & Evans, 1979).

An issue that is related to the lowest frequency band that suffices for recognition is the encoding economy of a band. For a filter with a bandwidth that is proportional to frequency (e.g. a two-octave-wide filter), the lower the frequency, the smaller the number of frequency components needed to encode the filtered image of a constant object. Combining these two notions, Ginsburg concluded that objects were best, or most efficiently, characterized by the lowest band of spatial frequencies that sufficed to discriminate them. Ginsburg (1980) went on to suggest that higher spatial frequencies were redundant for certain tasks, such as face or letter recognition.

Several investigators were quick to point out that objects can be well discriminated in various spatial frequency bands. Fiorentini, Maffei and Sandini (1983) observed that faces were well recognized in either high or in lowpass filtered bands. Norman and Erlich (1987) observed that high spatial frequencies were essential for discrimination between toy tanks in photographs.

\*To whom reprint requests should be addressed.

With respect to geometric illusions, both Janez (1984) and Carlson, Moeller and Anderson (1984) observed that the geometric illusions could be perceived for images that had been highpass filtered so that they contained no low spatial frequencies. This suggests that low and high spatial frequency bands may carry equivalently useful information for higher visual processes.

#### *Characterizing the visual system*

In the studies cited above, the discussion of spatial filtering focuses on *object* spatial frequencies, that is, frequencies that are defined in terms of some dimension of the object they describe (cycles per object). Most psychophysical research with spatial frequency bands has focused on *retinal* spatial frequencies, that is, frequencies defined in terms of retinal coordinates. For example, the spatial contrast sensitivity function (Davidson, 1968; Campbell & Robson, 1968) describes the threshold sensitivity of the visual system to sine wave gratings as a function of their *retinal* spatial frequency. Visual system sensitivity is greatest at 3–10 cycles per degree of visual angle (c/deg). How does visual system sensitivity relate to object spatial frequencies?

#### *Unconfounding retinal and object spatial frequencies*

Retinal spatial frequency and object spatial frequency can be varied independently to determine whether certain object frequencies are best perceived at particular retinal frequencies. Object frequency is manipulated by varying the frequency band of bandpass filtered images; retinal frequency is manipulated by varying the viewing distance.

The cutoff *object* spatial frequency of lowpass filters and the observer's viewing distance were varied independently by Legge, Pelli, Rubin and Schleske (1985) who studied reading rate of filtered text at viewing distances over a 133:1 range. Over about a 6:1 middle range of distances, reading rate was perfectly constant, and it was approximately constant over a 30:1 range. At the longest viewing distances, there was a sharp performance decrease (as the letters became indiscriminably small). At the shortest viewing distance, performance decreased slightly, perhaps due to large eye movements that the subjects would have to execute to bring relevant material towards their lines of

sight, and to the impossibility of peripherally previewing new text.

While viewing distance changed the overall level of performance in Legge et al., the cutoff *object* frequency of their low-pass filters at which performance asymptotized did not change. From this study, we learn that reading rate can be quite independent of retinal frequency over a fairly wide range, and that dependence on critical object frequency does not depend on viewing distance. Because the authors measured reading rate only in lowpass filtered images, we cannot infer reading performance in higher spatial frequency bands from their data.

#### *Unconfounding object statistics and visual system properties*

Human visual performance is the result of the combined effects of the objectively available information in the stimulus, and the ability of humans to utilize the information. In studying visual performance with differently filtered images, it is critical to separate availability from ability to utilize. For example, narrow-band images can be completely described in terms of a small number of parameters—Fourier coefficients or any other independent descriptors—than wide-band images. Poor human performance with narrow-band images may reflect the impoverished image rather than an intrinsically human characteristic—an ideal observer would exhibit a similar loss.

The problem of assessing the utility of stimulus information becomes acute in comparing human performance in high and in low frequency bandpass filtered images. Typically, filters are constructed to have a bandwidth proportional to frequency (constant bandwidth in terms of octaves). For example, Ginsburg (1980) used faces filtered into 2-octave-wide bands; while Norman and Ehrlich (1987) also used 2-octave bands for their filtered tank pictures. With such filters, high spatial frequency images contain more independent frequencies than low frequency images.

Although linear bandwidth represents perhaps the important difference between images filtered in octave bands at different frequencies, the informational content of the various bands also depends critically on the nature of the specific class of objects, such as faces or letter. Obviously, determining the information content of images is a difficult problem. When it is not solved, the amount of stimulus information available within a frequency band is confounded

with the ability of human observers to use the information. Direct comparisons of performance between differently filtered objects are inappropriate. This distinction between objectively available stimulus information and the human ability to use it has not been adequately posed in the context of spatial bandpass filtering.

### Efficiency

In the present context, physically available information is best characterized by the performance of an ideal observer. If there were no noise in the stimulus, the ideal observer would invariably respond perfectly. To compare the performance of an observer, human or ideal, noise of root-mean-square (r.m.s.) amplitude  $n$  is progressively added to the signal of r.m.s. amplitude  $s$  until the performance is reduced to some criterion, such as 50% correct in a letter identification task. This defines the signal to noise ratio,  $(s/n)_c$ , for a criterion  $c$ . Efficiency  $eff$  of human performance is defined by:

$$eff = \left( \frac{s_i}{n_i} \right)_c^2 / \left( \frac{s_h}{n_h} \right)_c^2$$

where  $h$  and  $i$  indicate *human* and *ideal* observers, and  $s$  and  $n$  are r.m.s. signal and noise amplitudes (Tanner & Birdsall, 1958). In a pure, quantally limited system, efficiency actually represents the fraction of quanta absorbed (utilization efficiency). In the context of signal detection theory, efficiency is given by a  $d'$  ratio:

$$eff = (d'_h/d'_i)^2.$$

### Overview

For an object that contains a broad spectrum of spatial frequencies, object spatial frequency is determined by the center frequency of a spatial bandpass filtered image. Retinal spatial frequency is determined by the viewing distance at which the stimulus is viewed. Stimulus information is determined jointly by the signal-to-noise ratio, by the spatial filtering, and by the characteristics of the set of signals; these three informational components are combined in the efficiency computation. Letters are a convenient stimulus to study because they are highly over-learned so that human performance can be expected to be reasonably efficient, and because much is already known about the visibility of letters in the presence of internal noise (letter acuity) and about the visual processing of letters.

Specifically, to determine the roles of object and retinal spatial frequencies, letters are filtered into various frequency bands. Noise is added, and the psychometric function for correct identification is determined as a function of  $s/n$ . Accuracy depends only on  $s/n$  and not on overall contrast, for a wide range of contrasts (Pavel, Sperling, Riedl & Vanderbeck, 1987). This determination is repeated for every combination of object frequency band and viewing distance. Thereby, retinal spatial frequency and object spatial frequency are unconfounded, enabling us to determine whether a particular object frequency band is better discriminated in one visual channel (retinal frequency) than any other (Parish & Sperling, 1987a, b). Moreover, by computing an ideal observer for the identification task, we obtain an objective measure of the information that is present in each of the frequency bands. Finally, the comparison of human performance with the performance of the ideal observer gives us a precise measure of the ability of our subjects to utilize the information in the stimulus. Having untangled these factors, we can determine which spatial frequencies most efficiently characterize letters for identification.

## METHOD

Two experiments were conducted using similar stimuli and procedures.

### Stimuli

*Letters (signals) and noise.* The original, unfiltered letters were selected from a simple  $5 \times 7$  upper-case font commonly used on CRT terminals. Since this is an experiment in pattern recognition, we felt that the simplest letter pattern might be the most general; indeed, this font has been widely used in letter discrimination studies. For the purpose of subsequent spatial filtering, the letters were redefined on a pixel grid that measured 45 (vertical height)  $\times$  35 (maximum horizontal extent of letters M and W). The letters had value 1 (white); the background had value 0 (black). To avoid edge effects in filtering, the background was extended to  $128 \times 128$  pixels for all computations. However, only the center  $90 \times 90$  pixels of the stimulus were displayed, as these contained effectively all the usable stimulus information, even for low spatial-frequency stimuli. Letters for presentation were chosen pseudo-randomly from the set of 26 upper-case English letters. Noise

Table 1. Parameters of the bandpass filters: lower and upper half-amplitude frequencies, peak, and 2D mean frequencies in cycles/letter height

| Band | Lower | Peak     | Upper | Mean <sup>a</sup> |
|------|-------|----------|-------|-------------------|
| 0    | 0     | Lowpass  | 0.53  | 0.39              |
| 1    | 0.26  | 0.53     | 1.05  | 0.74              |
| 2    | 0.53  | 1.05     | 2.11  | 1.49              |
| 3    | 1.05  | 2.11     | 4.22  | 2.92              |
| 4    | 2.11  | 4.22     | 8.44  | 5.77              |
| 5    | 6.33  | Highpass | 22.5  | 20.25             |

<sup>a</sup>Frequencies are weighted according to their squared amplitude (power) in computing the mean.

fields were defined on a  $128 \times 128$  array by choosing independent Gaussian noise samples for each pixel, with the mean equal to zero and a variance  $\sigma^2$  as required by the condition. (As with the letters, only the central  $90 \times 90$  pixels were displayed.) Forty different noise fields were created.

**Filters.** Each stimulus consisted of a filtered letter added to an identically filtered noise field. Six spatial filters were available, corresponding to six successive levels of a Laplacian pyramid (Burt & Adelson, 1983). The zero-frequency component was added to the images so that they could be viewed. The object-relative filter characteristics, upper and lower half-amplitude cutoff and 2D mean frequency (cycles per letter height), appear in Table 1. The 2D mean frequency  $\bar{f}$  for a given band is:

$$\bar{f} = \frac{\sum_{x=0}^{127} \sum_{y=0}^{127} f_{x,y} a_{x,y}^2}{\sum_{x=0}^{127} \sum_{y=0}^{127} a_{x,y}^2}$$

where  $f_{x,y}$  is the 2D frequency and  $a_{x,y}$  is its amplitude. Cycles per object height is used rather than the more usual cycles per object width because the height of our upper-case letters remained constant across the entire set, whereas the width varied between letters.

The transfer functions (spectra) of the filters are displayed in Fig. 1. Approximately, filters are separated in spatial frequency by an octave (factor of 2) and have a bandwidth at half-amplitude of two octaves. The small mound in the lower right corner of Fig. 1 is a negligible imperfection in filter 4. For convenience, the limited range of spatial frequencies passed by each of the filters will be referred to as the *band* of that filter; a specific band is  $b_i$  ( $i = 0, 1, 2, 3, 4, 5$ ), where  $b_0$  is the lowest set of frequencies and  $b_5$  is the highest.

The filter spectra (shown in Fig. 1) are approximately symmetrical in log frequency coordinates, a symmetrical spectrum in log coordinates is highly skewed to the right in linear frequency coordinates, resulting in a mean that

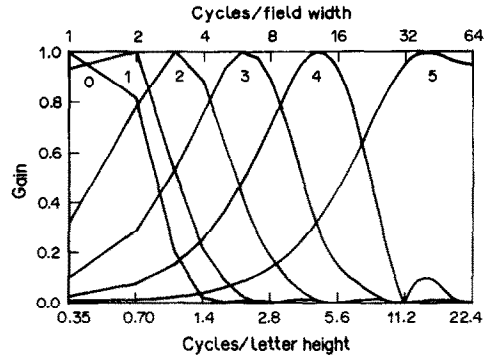


Fig. 1. Filter characteristics for the filters used in the experiments. There are two abscissas, both on a log scale. The top abscissa is the frequency in cycles per unwrapped field width (128 pixels); the bottom abscissa is in cycles per letter height (45 pixels). The ordinate is the normalized gain. The parameter  $i$  indicates the filter designation  $b_i$  in the text.

is much greater than the mode. In a 2D (vs 1D) filter, the rightward shift is accentuated. For example, band 2 has a peak frequency of 1.05 c/object but a 2D mean frequency of 1.49 c/object. The single most informative characterization of such a skewed bandpass spectrum depends somewhat on the context; usually use the mean rather than the peak.

Figure 2 (top) shows the letter G, filtered in bands 1-5 without noise; the bottom shows the same signals plus noise,  $s/n = 0.5$ . The full  $128 \times 128$  array (extended by reflection beyond its edges) was passed through the filter so that the effect of the picture boundary did not intrude into the critical part of the display.

**Signal to noise ratio,  $s/n$ .** A filtered letter is a *signal*. Let  $i, j$  index a particular pixel in the  $x, y$  coordinate space of the stimulus. The signal contrast  $c_s(i, j)$  of pixel  $i, j$  is:

$$c_s(i, j) = \frac{l(i, j) - l_0}{l_0} \quad (1)$$

where  $l_{i,j}$  is the luminance of pixel  $i, j$  and  $l_0$  is the mean signal luminance over the  $90 \times 90$  array. Signal power per pixel,  $s$ , is defined as mean contrast power averaged over the  $90 \times 90$  pixel array:

$$s = (IJ)^{-1} \sum_i \sum_j c_s(i, j)^2 \quad (2)$$

where  $c_{i,j}$  is the contrast of pixel  $i, j$  and  $I = J = 90$ .

Noise contrast  $c_n(i, j)$  is the value of the  $i, j$ th noise sample divided by the mean luminance. Analogously to signal power (equation 2), noise contrast power per pixel,  $n$ , is equal to  $(\sigma/l_0)^2$ . The signal to noise ratio is simply  $s/n$ .

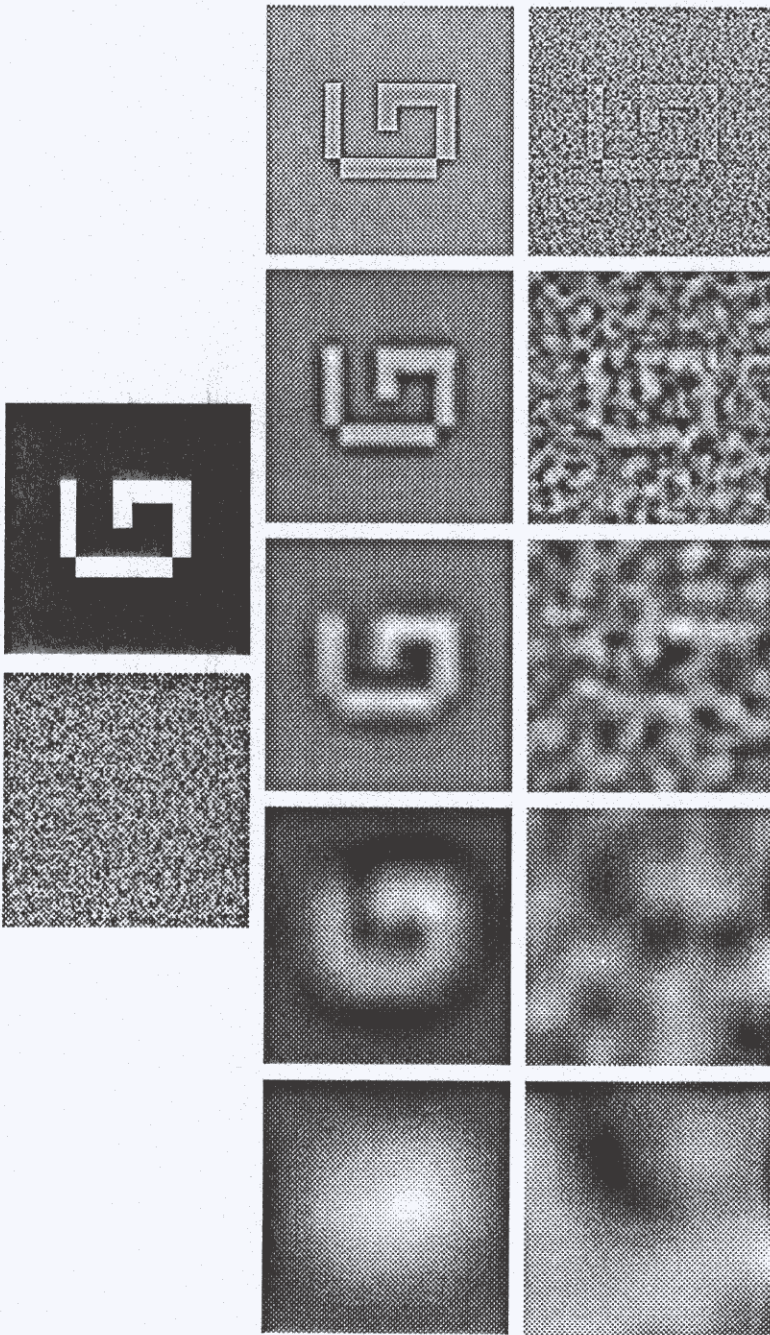


Fig. 2. Top: unfiltered noise and unfiltered letter G. Middle: the letter G filtered in spatial frequency bands 1-5 with only quantization noise. Bottom: filtered letter G plus filter noise in the same bands with a signal-to-noise ratio of 0.50 in all panels. The effective  $s/n$  in the reproduction is somewhat lower (from Parish & Sperling, 1987a). The first row of numerals indicates the number by which the filter band is referred to in the text; the bottom row indicates the *mean* frequency of the bands in cycles per letter height.

**Quantization.** Our display system produced 256 discrete luminance levels. Level 128 was used as the mean luminance  $l_0$ ;  $l_0$  was 47.5 cd/m<sup>2</sup>. To produce a visual display of a given letter, band, and  $s/n$ , signal power  $s$  and noise power  $n$  were normalized so that the luminance of every one of the 8100 displayed pixels fell within the range of the display system; there was no truncation of the tails of the Gaussian noise. (Although the relationship between input gray-level and output luminance was not quite linear at the extreme intensity values, it was determined that more than 90% of the pixels fell within the linear intensity range.) Intensity normalization was applied separately to each stimulus (combination of signal plus noise). By normalizing the total stimulus  $s + n$ , the actual value of  $s$  displayed to the subject diminished as  $n$  increased; i.e. the actual value of  $s$  was not known by the subject. Indeed, even stimuli with precisely the same letter in the same band and with the same  $s/n$  might be produced with slightly different  $s$  and  $n$  depending on the extreme values of the noise fields.

Seven values of  $s/n$  were available for each band, chosen in a pilot study to insure that the data yielded the entire psychometric function (chance to best performance). The same pilot study showed that subjects never performed above chance when confronted with noise-free letters from  $b_0$ ; this band was omitted from the present study.

#### Procedure: experiment 1

Four of the experimental variables—letter identity, noise field, frequency band, and  $s/n$ —were randomized within each session. A fifth variable, viewing distance, was held constant within each session and was varied between sessions. Four viewing distances were used: 0.121, 0.38, 1.21 and 3.84 m. A chin rest was used to stabilize the subject's head for viewing at the shortest distance. At the four distances, the 90 × 90 pixel stimulus subtended 31.6, 10, 3.16 and 1.0 deg of visual angle respectively. The

upper and lower half-amplitude cut-off retinal frequencies for the upper six filters, with respect to the four viewing distances used in this experiment, and for a fifth distance used in the second experiment, appear in Table 2. Subjects participated in four 1-hr sessions at each viewing distance. Each session consisted of 315 trials, nine trials at each of seven  $s/n$ 's for each of the five frequency bands.

Prior to the first session, subjects were shown noise-free examples of the unfiltered letters. They were told that each stimulus presentation consisted of a letter and a certain amount of noise, and that the letter may appear degraded in some way. They were informed that at no time would a letter be shifted in orientation or from its central location in the stimulus field. Finally, they were instructed to view each stimulus for as long as they desired before making their best guess as to which letter had been presented. A response (letter identity) was required on every trial. Subjects typed the response on a keyboard connected to the host computer (Vax 11/750); subsequently, typing a carriage return erased the video screen and initiated the next trial in a few seconds. The room illumination was very dim; the response keyboard was lighted by stray light from its associated CRT terminal. No feedback was offered to the subjects.

#### Observers

Three subjects, two male and one female, between the ages of 20 and 27 participated in the experiment. All subjects had normal or corrected-to-normal vision. One of the subjects was a paid participant in the study.

#### Procedure: experiment 2

This experiment was run before expt 1. It is reported here because it offers additional data with two new and one old subject at a fifth viewing distance. Except as noted, the procedures are similar to expt 1. The screen was viewed through a darkened hood at a distance

Table 2. Lower and upper half-power frequency and 2D mean frequency (in c/deg of visual angle) for all bands and viewing distances used in both experiments

| Band         | Viewing distance (m) |                  |                    |                    |                  |
|--------------|----------------------|------------------|--------------------|--------------------|------------------|
|              | 0.12                 | 0.38             | 1.21               | 3.84               | 0.48             |
| 0 (lowpass)  | 0.00–0.04 (0.03)     | 0.00–0.12 (0.09) | 0.00–0.37 (0.27)   | 0.00–1.18 (0.87)   | 0.00–0.15 (0.11) |
| 1            | 0.02–0.07 (0.05)     | 0.06–0.23 (0.16) | 0.18–0.74 (0.52)   | 0.58–2.34 (1.65)   | 0.07–0.29 (0.21) |
| 2            | 0.04–0.15 (0.10)     | 0.12–0.47 (0.33) | 0.37–1.48 (1.04)   | 1.18–4.70 (3.30)   | 0.15–0.59 (0.41) |
| 3            | 0.07–0.30 (0.20)     | 0.23–0.94 (0.64) | 0.74–2.97 (2.04)   | 2.34–9.40 (6.48)   | 0.29–1.18 (0.81) |
| 4            | 0.15–0.59 (0.40)     | 0.47–1.88 (1.27) | 1.48–5.94 (4.04)   | 4.70–18.80 (12.82) | 0.59–2.36 (1.60) |
| 5 (highpass) | 0.30–2.25 (1.41)     | 0.94–7.13 (4.45) | 2.97–22.53 (14.19) | 9.40–71.27 (45.00) | 1.77–8.96 (5.63) |

of 0.48 m. At this distance, the  $90 \times 90$  stimuli subtended 7.15 deg of visual angle. The half-amplitude cut-off frequencies and the mean frequencies of the six spatial filters are given in the rightmost column of Table 2. Three male subjects between the ages of 20 and 27 participated in the experiment. All subjects had normal or corrected-to-normal vision. Two of the subjects were paid for their participation, and one, DHP, also participated in expt 1. Five sessions of 315 trials were run for each subject.

## RESULTS

### *Psychometric functions: $\hat{p}$ vs $\log_{10} s/n$*

The measure of performance is the observed probability  $\hat{p}$  of a correct letter identification.

The complete psychometric functions are displayed in Figs 3 (expt 1) and 4 (expt 2). A separate psychometric function is shown for each subject, viewing distance and frequency band. In band  $b_1$ , for all subjects, performance asymptotes (for noiseless stimuli) at  $\hat{p} \approx 0.5$ . In all other bands, performance improves from near-chance (1/26) to near perfect as the value of  $s/n$  increases.

### *Noise resistance as a function of frequency band*

An obvious aspect of the data of both experiments is that the data move to the left of the figure panels as band spatial frequency increases. This means that high spatial frequency stimuli (bands  $b_4, b_5$ ) are identifiable at smaller

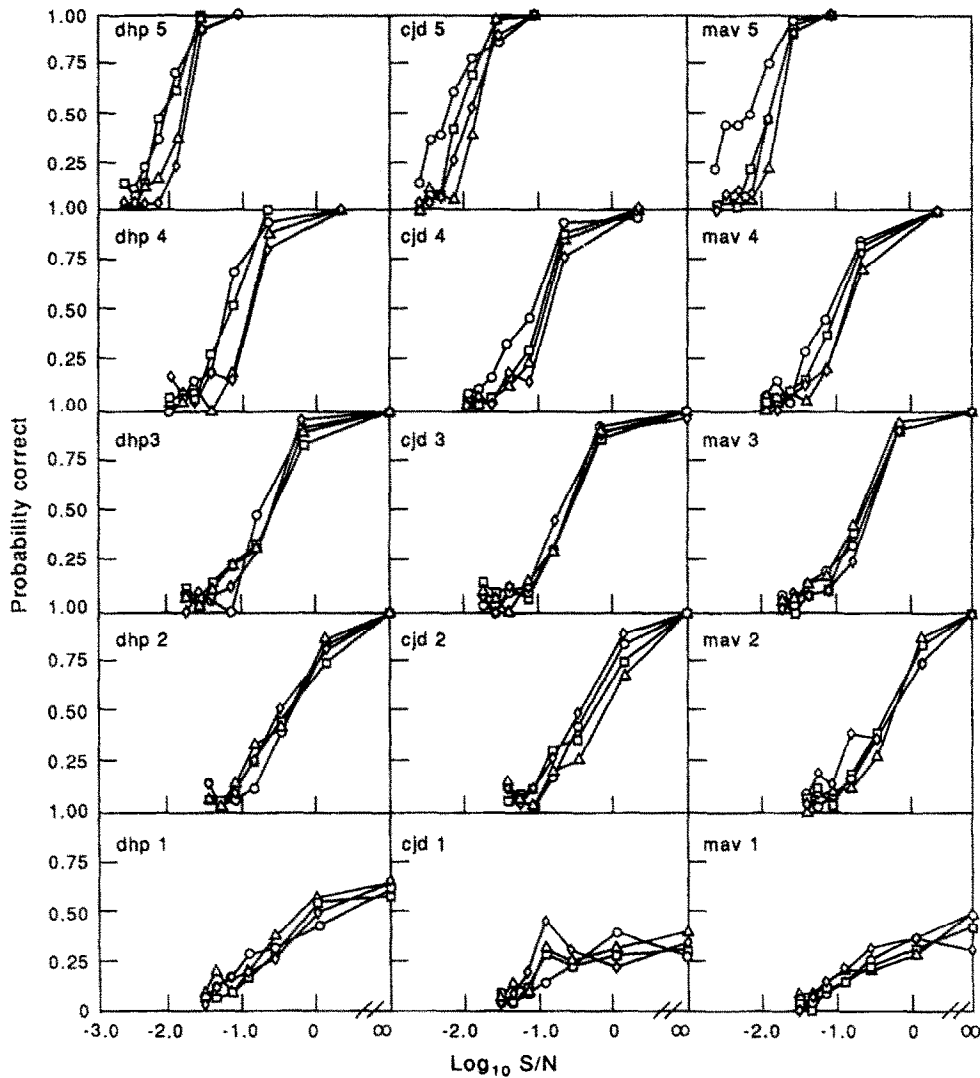


Fig. 3. Psychometric functions from expt 1. Each graph displays performance as a function of  $\log_{10} s/n$ , within a frequency band. The parameter is viewing distance. Subjects are arranged in columns and frequency band is arranged in rows, progressing from the highest frequency band at the top to the lowest band at the bottom. The four viewing distances are 3.84 (○), 1.21 (△), 0.38 (□), and 0.121 (◇) m.

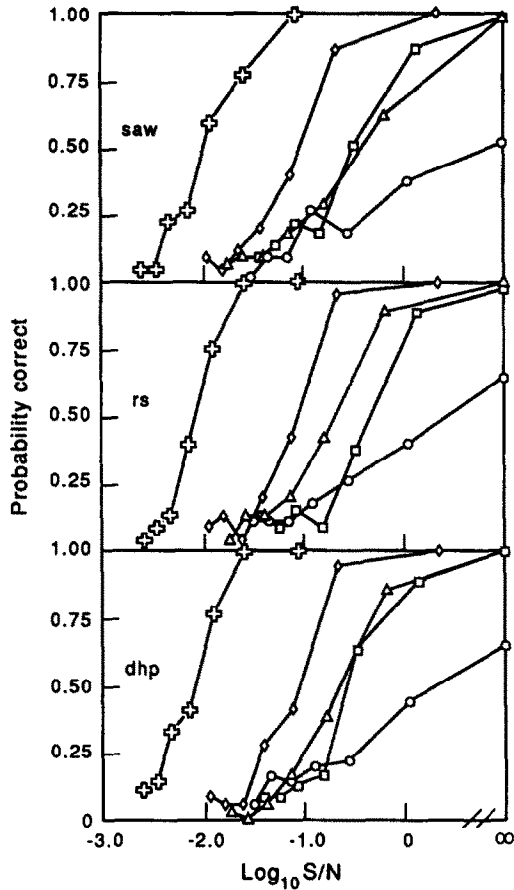


Fig. 4. Psychometric functions for each subject and frequency band in expt 2. Viewing distance was 0.48 m. The five frequency bands,  $b_1$ - $b_5$ , are indicated, respectively, by O, □, Δ, ◇ and +. The probability of a correct response is plotted as a function of  $\log_{10} s/n$ .

$s/n$  than stimuli in bands  $b_1$  and  $b_2$ ; resistance to noise increases with spatial frequency band. To enable comparisons of noise sensitivity as a function of band, the  $s/n$  at which  $\hat{p} = 50\%$  was estimated for each subject and frequency band from expt 1 by means of inverse interpolation from the best fitting logistic function. As viewing distance had no effect, all estimates were made using the data collected when viewing distance was equal to 0.38 m. A graph of these  $(s/n)_{50\%}$  points as a function of the mean object frequency of the band is plotted in Fig. 5 (O). For comparison, the expected rate of improvement in  $(s/n)_{50\%}$ , based on the increasing number of frequency components as one moves from low to high frequency bands, is plotted as a series of parallel lines in Fig. 5. Performance improves [ $(s/n)_{50\%}$  decreases] somewhat faster than  $1/f$  (the slope of the parallel lines). These results, and Fig. 5, will be analyzed in detail in the Discussion section.

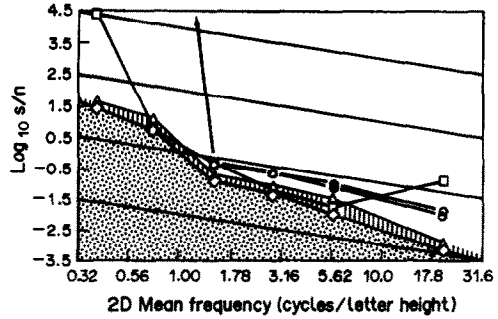


Fig. 5. Performance of human subjects and various computational discriminators. The abscissa indicates  $\log_{10}$  of the mean frequency of each bandpass stimulus. The ordinate indicates the (interpolated)  $s/n$  ratio at which a probability of a correct response  $p = 0.5$  is achieved. Circles indicate each of the three subjects in expt 1 at the intermediate viewing distance of 1.21 m. In band  $b_1$ , 2 of 3 human subjects fail to achieve 50% correct ( $eff = 0$ ); these points lie outside the graph. ( $\Delta$ ) indicates sub-ideal and ( $\diamond$ ) indicates super-ideal performances of discriminators that brackets the ideal discriminator. The shaded area below the super-ideal discriminator indicates theoretically unachievable performance. Squares indicate performance of a spatial correlator-discriminator. The oblique parallel lines have slope  $-1$  that represents the improvement in expected performance (decrease in  $s/n$ ) as function of the number of frequency components in each band when filter bandwidth is proportional to frequency.

*The non-effect of viewing distance*

Another property of the data is that, in most conditions, viewing distance has no effect on performance. Analysis of variance, carried out individually for each subject, shows that there is no significant effect of distance in any band for subject dhp and a significant effect of distance in bands  $b_4$  and  $b_5$  for the other two subjects. Further analysis by a Tukey test (Winer, 1971) in bands  $b_4$  and  $b_5$  for these subjects shows that the only significant effect of distance is that visibility at the longest viewing distance is *better* than at the other three distances. For subject CJD, the improvement is equivalent to a gain in  $s/n$  of 0.19 and 0.28  $\log_{10}$  (for bands  $b_4$  and  $b_5$ , respectively); for MAV, the corresponding gains were 0.21 and 0.40.

Improved performance at long viewing distances is almost certainly due to the square configuration of individual pixels, which produces a high frequency spatial pixel noise that is attenuated by viewing from sufficiently far away (Harmon & Julesz, 1973). In low frequency bands, pixel-boundary noise is not a problem because the spatial filtering insures that adjacent pixels vary only slightly in intensity. We explored the hypothesis of pixel-boundary noise with subject CJD, who showed a distance effect



in band 5. At an intermediate viewing distance of 1.21 m, CJD squinted her eyes while viewing stimuli from band 5. By blurring the retinal image of the display in this way, performance improved approximately to the level of the furthest viewing distance.

To summarize, the only significant effect of distance that we observed was a lowering of performance at near viewing distances relative to the furthest distance. This impairment occurred primarily in bands 4 and 5. In these bands, the spatial quantization of the display ( $90 \times 90$  square-shaped pixels) produces artifactual high spatial frequencies that mask the target. These artifactually produced spatial frequencies can be attenuated by deliberate blurring (squinting), or by producing displays with higher spatial resolution, or by increasing the viewing distance to the point where the pixel boundaries are attenuated by the optics of the eye and neural components of the visual modulation transfer function. In all cases, blurring improves performance and eliminates the slightly deleterious effect of a too small viewing distance. Thus, for correctly constructed stimuli, in the frequency ranges studied, there would be no significant effect of viewing distance on performance. This finding is in agreement with the results of Legge et al. (1985), who examined reading rate rather than letter recognition. It is in stark disagreement with the results of sinewave detection experiments in which retinal frequency is critical—see Sperling (1989) for an explanation.

## DISCUSSION

A comparison of performance in different frequency bands shows that subjects perform better the higher the frequency band; and subjects require the smallest signal-to-noise ratio in the highest frequency band. To determine whether performance in high frequency bands is good because humans are more efficient in utilizing high-frequency information, or because there is objectively more information in the high-frequency images, or both, requires an investigation of the performance of an ideal observer. The performance of the ideal observer is the measure of the objective presence of information. Human performance results from the joint effect of the objective presence of information and the ability of humans to utilize that information. Human efficiency is the ratio of human performance to ideal performance.

## *Ideal discriminator*

*Definition.* An ideal discriminator makes the best possible decision given the available data and the interpretation of “best.” The performance of the ideal discriminator defines the objective utility of the information in the stimulus. We prefer the name *ideal discriminator*, rather than *ideal observer*, because it indicates the critical aspect of performance under consideration, but we occasionally use *ideal observer* to emphasize the relations to a large, relevant literature on this subject. Our purposes in this section are first, to derive an ideal discriminator for the letter identification task, second, to develop a practical working approximation to this discriminator, and third, to compare the performance of the human with the ideal discriminator.

Although ideal observers have recently come into greater use in vision research, the applications have focused primarily on determining the limits of performance for relatively low-level visual phenomena. For example, Barlow (1978, 1980), and Barlow and Reeves (1979) investigated the perception of density and of mirror symmetry; Geisler (1984) investigated the limits of acuity and hyperacuity; Legge, Kersten and Burgess (1987) examined the pedestal effect; Kersten (1984) studied the detection of noise patterns; and Pelli (1981) detailed the roles of internal visual noise. Geisler (1989) provides an overview of efficiency computations in early vision. Our application differs from these in that we expand the techniques and apply them to a higher perceptual/cognitive function, letter recognition.

For the letter identification task, the ideal discriminator is conceptually easy to define. A particular observed stimulus,  $x$ , representing an unknown letter plus noise, consists of an intensity value (one of 256 possible values) at each of  $90 \times 90$  locations. The discriminator's task is to make the correct choice as frequently as possible from among the 26 alternative letters.

The likelihood of observing stimulus  $x$ , given each of the 26 possible signal alternatives, can be computed when the probability density function of the added noise is known exactly. The optimal decision chooses the letter that has the highest likelihood of yielding  $x$ . The expected performance of the ideal discriminator is computed by summing its probability of a correct response over the  $256^{8100}$  possible stimuli (256 gray levels,  $90 \times 90$  pixels). Unfortunately,

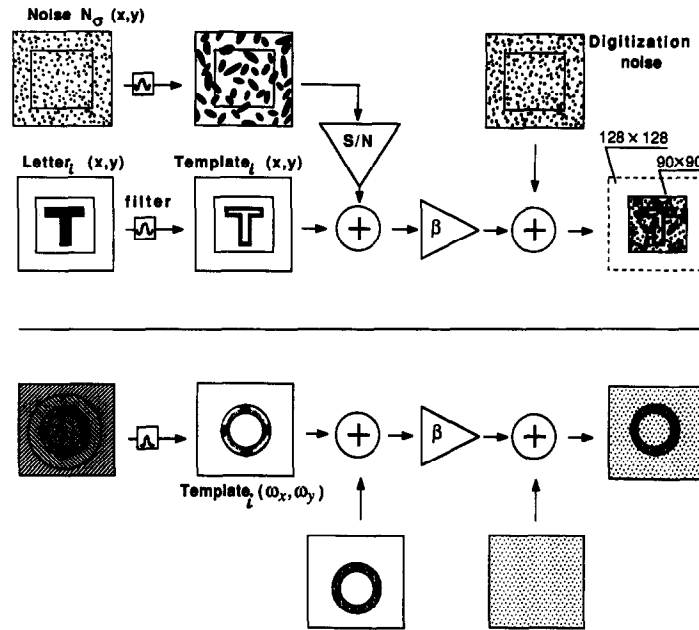


Fig. 6. Flow chart of the experimental procedures that are modelled by the ideal discriminator analysis. Upper half indicates space-domain operations; lower half indicates the corresponding operations in the frequency domain. Computations are carried out on  $128 \times 128$  arrays; the subject sees only the center  $90 \times 90$  pixels. A random letter and a random noise field are each filtered by the same filter ( $b$ ); the noise is amplified to provide the desired signal-to-noise ratio; the letter and noise are added, the output is scaled and quantized (represented by the addition of digitization noise), and the result is shown to the subject. In the frequency domain  $\omega_x, \omega_y$ , the bandpass filter selects an annulus, whereas the quantization noise is uniform over  $\omega_x, \omega_y$ .

when there is both bandpass filtered and intensity quantization, the usual simplifications that make this enormous computation tractable are not applicable.

As an alternative to computing the expected performance of the ideal discriminator, one can compute its performance with a particular subset of the possible stimuli—the stimuli that the subject actually viewed or, preferably, a larger set of stimuli for more reliable estimation. This Monte Carlo simulation of the performance of the ideal discriminator is a tractable computation that yields an estimate of expected performance.

*Derivation.* Stimulus construction is diagrammed in Fig. 6 which shows the equivalent operations in the space and the frequency domains. To derive an ideal discriminator, we need to carefully review the processes of stimulus construction. We use uppercase letters to represent quantities in the frequency domain and lowercase letters to represent quantities in the space domain. A letter is defined by a  $90 \times 90$  array that takes the value 1 at the letter locations and 0 at the background locations. When this array is spatially filtered in band  $b$ , it defines the letter template  $t_{i,b}(x, y)$ , where  $i$

indicates the particular letter,  $b$  the frequency band, and  $x, y$  the pixel location. We write  $T_{i,b}(\omega_x, \omega_y)$  for the Fourier series coefficient of  $t_{i,b}$  indexed by frequency.

An unknown stimulus  $u_{i,b}(x, y)$  to be viewed by a subject is produced by adding filtered  $n_b(x, y)$  with post-filtering variance  $\sigma_N^2$ , to the template  $t_{i,b}(x, y)$ , where letter identity  $i$  is unknown to the subject. The stimulus is scaled and digitized (quantized) to 256 levels prior to presentation, contributing an additional source of noise  $q_{i,b}(x, y)$ , called digitization noise. Finally, a d.c. component ( $dc$ ) is added to  $u_{i,b}$  to bring the mean luminance level to 128. These steps are diagrammed in Fig. 6 which shows both the space-domain and the corresponding frequency-domain operations. The space-domain computation is encapsulated in equations (3):

$$u_{i,b}(x, y) = \beta_{i,b}[t_{i,b}(x, y) + n_b(x, y)] \quad (3a)$$

$$u_{i,b}(x, y) = \beta_{i,b}[t_{i,b}(x, y) + n_b(x, y) + q_{i,b}(x, y) + dc] \quad (3b)$$

The scaling constant  $\beta_{i,b}$ , limits the range of real values for each pixel, prior to quantization, to  $[-0.5, 255.5]$ . The degree of scaling is determined by the maximum and minimum values in

the function  $t_{i,b} + n_b$ . Note that the extreme values in the image are determined by  $\sigma_{N_2}$  which is adjusted to yield the appropriate  $s/n$  for each condition; the values of  $t_{i,b}$  are fixed prior to scaling. Specifically:

$$\beta_{i,b} = \frac{256}{\max(t_{i,b} + n_b) - \min(t_{i,b} + n_b)}. \quad (4)$$

As a result of bandpass filtering, the noise samples in adjacent pixels are strongly dependent on each other. Therefore, the discriminator problem is best approached in the Fourier domain, where the random variables  $\{N_b(\omega_x, \omega_y)\}$  are jointly independent because the filtering operations simply scale the different frequency components without introducing any correlations (van Tress, 1968). The task of the ideal discriminator is to pick the template  $t_{i,b}$  that maximizes the likelihood of  $u_{i,b}$  with *a priori* knowledge of: (i) the fixed functions  $t_{i,b}$ , and their probabilities; and (ii) the densities of the jointly independent random variables  $\{N_b(\omega_x, \omega_y)\}$ . As is clear,  $\beta_{i,b}$ ,  $\sigma_N^2$ ,  $\{Q_{i,b}(\omega_x, \omega_y)\}$ , and  $\{N_{i,b}(\omega_x, \omega_y)\}$  are all jointly distributed random variables characterized by some density  $f$ . To compute the likelihood of  $u_{i,b}$  the ideal discriminator must integrate  $f$  over all possible values that may be assumed by the set of jointly distributed random variables, whose values are constrained only in that they result in a possible stimulus  $u_{i,b}$ . Unfortunately, no closed-form solution to this problem is available, forcing us to look for an alternative approach.

**Bracketing.** To estimate the performance of the ideal discriminator, we look for a tractable super-ideal discriminator that is better than the ideal but which is solvable. Similarly, we look for a tractable sub-ideal discriminator that is worse than the ideal. The ideal discriminator must lie between these two discriminators; that is, we bracket its performance between that of a "super-ideal" and a "sub-ideal" discriminator. The more similar the performance of the super- and sub-ideal discriminators, the more constrained is the ideal performance which lies between them.

Our super-ideal discriminator is told, *a priori*, the exact values for  $\beta_{i,b}$  and  $\sigma_N^2$  for each stimulus presentation. Therefore, it is expected to perform slightly better than the ideal discriminator which must estimate these values from the data. The sub-ideal discriminator estimates these same parameters from the presented stimulus in a simple but nonideal way. There-

fore, it is expected to perform slightly worse than the ideal discriminator. The computational forms used to compute  $\beta_{i,b}$  and  $\sigma_N^2$  for the sub-ideal discriminator are presented in the Appendix, along with the derivation of the likelihood estimator used by both discriminators. A complete discussion of these derivations and the problems associated with the formulation of an ideal discriminator for such complex stimuli is presented in Chubb, Sperling and Parish (1987).

**Performance of the bracketed discriminator.** The super- and sub-ideal discriminators were tested in a Monte Carlo series of trials, in which they each were confronted with 90 stimuli in each of the frequency bands at each of seven  $s/n$  values chosen to best estimate their 50% performance point. The  $s/n$  necessary for 50% correct discriminations was estimated by an inverse interpolation of the best fitting logistic function. The derived  $(s/n)_{50\%}$  is the measure of performance of a discriminator. The mean ratio, across frequency bands, of

$$(s/n)_{50\%} \text{ sub-ideal} / (s/n)_{50\%} \text{ super-ideal}$$

is about 2 (approx.  $0.3 \log_{10}$  units). The ratio does not depend on the criterion of performance.

#### *Efficiency of human discrimination*

In all conditions, human subjects perform worse than the sub-ideal discriminator. Notably, with no added luminance noise, the subideal (and, of course, the ideal) discriminator function perfectly, even in  $b_0$  where subject performance is at chance, and in  $b_1$  where subjects reached asymptote at about 50% correct.

Data from the subjects are plotted with the  $(s/n)_{50\%}$  sub-ideal and  $(s/n)_{50\%}$  super-ideal in Fig. 5. For comparison, Fig. 5 also shows the performance of a correlator discriminator which chooses the letter template that correlates most highly with the stimulus in the space domain. In the coordinates of Fig. 5 ( $\log_{10} s/n$  vs  $\log_{10} f$  where  $f$  represents the mean 2D spatial frequency of the band), the vertical distance  $d$  from the human data  $\log(s/n)_{50\%}$ , *human* down to the bracketed discriminator  $\log(s/n)_{50\%}$ , *ideal* represents the  $\log_{10}$  of the factor by which the bracketed discriminator outperforms the human observer at that value of  $f$ . For the purpose of specifying efficiency, we assume the ideal discriminator lies at the mid-point of the sub and super-ideal discriminators in Fig. 5. The

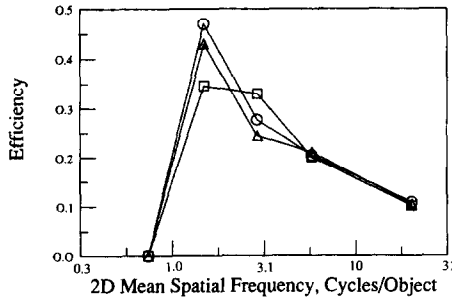


Fig. 7. Discrimination efficiency as a function of the mean frequency of a 2-octave band (in cycles per letter height) indicated on a logarithmic scale. Data are shown for three observers:  $\Delta$  = SAW,  $\square$  = RS,  $\circ$  = DHP. The viewing distance is 2.21 m, which is representative of all viewing distances tested.

efficiency  $eff$  of human discrimination relative to the bracketed discriminator is  $eff = 10^{-2d}$ , where:

$$d = \log(s/n)_{50\%, human} - \log(s/n)_{50\%, ideal}$$

The values of  $eff$  in each object frequency band are shown in Fig. 7. In band 0,  $eff$  is zero because human performance never reaches 50%; indeed, it never rises significantly above 4% (chance). In band 1, human performance asymptotically climbs close to 50% as  $s/n$  approaches infinity;  $eff \approx 0$ . In band 2,  $eff$  reaches its maximum of 35–47% (depending on the subject), and it declines rapidly with increasing frequency ( $b_3$ – $b_5$ ).

The 42% average efficiency in band 2 is similar in magnitude to the highest efficiencies observed in comparable studies. For example, efficiency has been determined for detecting various kinds of patterns in arrays of random dots (Barlow, 1978, 1980; van Meeteren & Barlow, 1981), tasks which, like ours, may require significantly cognitive processing. In a wide range of conditions, the highest efficiencies observed were about 50%, and frequently lower. Van Meeteren and Barlow (1981) also found that efficiency was perfectly correlated with object spatial frequency and was independent of retinal spatial frequency.

**Spatial correlator discriminator.** A correlator discriminator cross-correlates the presented stimulus with its memory templates and chooses the template with the highest correlation. Correlation can be carried out in the space or in the frequency domain. Correlation is an efficient strategy when noise in adjacent pixels is independent and when members of the set of signals have the same energy; both of these conditions

are violated by our stimuli. However, when sufficient prior information is available to subjects, they do appear to employ a cross-correlation strategy (Burgess, 1985).

It is interesting to note that the performance of the spatial correlator discriminator over the middle range of spatial frequencies is quite close to the performance of the sub-ideal discriminator. At high spatial frequencies, correlator performance degenerates, due to its inability to focus spatially on those pixel locations that contain the most information. A spatial correlator that optimally weighted spatial locations, could overcome the spatial focusing problem at high frequencies. (Spatial focusing is treated in the next section.)

At all frequencies, the spatial correlator is nonideal because noise at spatial adjacent pixels is not independent. At low spatial frequencies, the nonindependence of adjacent locations becomes extreme and the correlator fails miserably. This points out that, for our stimuli, correlation detection is better carried out in the frequency domain because there the noise at different frequencies is independent. The qualitative similarity between the correlator discriminator and the subjects' data suggests that the subjects might be employing a spatial correlation strategy, augmented by location weighting at high frequencies.

**Lowest spatial frequencies sufficient for letter discrimination.** Band 2 corresponds to a 2-octave band with a peak frequency of 1.05 c/object (vertical height of letters) and a 2D mean frequency of 1.49 c/object. At the four viewing distances, 1.05 c/object corresponds to retinal frequencies of 0.074, 0.234, 0.739 and 2.34 c/deg of visual angle. We observe perfect scale invariance: all of these retinal frequencies, and hence the visual channels that process this information, are equally effective in achieving the high efficiency of discrimination.

The finding that  $b_2$  with a center frequency of 1.05 c/object and a  $\frac{1}{2}$  amplitude cutoff at 2.1 c/object is critical for letter discrimination is in good agreement with previous findings of both Ginsburg (1978) for letter recognition and Legge et al. (1985) for reading rate. Legge et al. used low-pass filtered stimuli, which included not only spatial frequencies within an octave of 1 c/object ( $b_2$ ) but also included all lower frequencies. From the present study, we expect human performance with low-pass and with band-pass spatial filtering to be quite similar up to 1 c/object because the lowest frequency

bands, when presented in isolation, are perceptually useless (at least when presented alone).

It is an important fact that our subjects actually performed better, in the sense of achieving criterion performance at a lower  $s/n$  ratio, at higher frequency bands than  $b_2$ . This is explained by the increase in stimulus information in higher frequency stimuli. Increased information more than compensates for the subjects' loss in efficiency as spatial frequency increases.

#### *Components of discrimination performance*

Though the performance of the bracketed ideal discriminator is useful in quantifying the informational utility of the various bands, it is instructive to consider the changing physical structure of the stimuli as well. What components of the stimuli actually lead to a gain in information with increasing frequency? According to Shannon's theorem (Shannon & Weaver, 1949), an absolutely bandlimited 1-D signal can be represented by a number of samples  $m$  that is proportional to its bandwidth. When the signal-to-noise ratio in each sample  $s_i/n_i$  is the same, the overall signal-to-noise ratio  $s/n$  grows as  $\sqrt{m}$ . In the space domain, our filters were constructed (approximately) to differ only in scale but not in the shape of their impulse responses. Therefore, when the mean frequency of a filter band increased by a factor of 2, the bandwidth also increased by 2. Since the stimuli are 2D, the effective number of samples increases with the square of frequency, and the increase in effective  $s/n$  ratio is proportional to  $m$ . This expected improvement with frequency, based simply on the increase in effective number of samples, is indicated by the oblique parallel lines of Fig. 5 with slope of  $-1$ . The expected improvement in threshold  $s/n$  due simply to the linearly increasing bandwidth of the bands does a reasonable job of accounting for the improvement in performance for both human and bracketed discriminators between  $b_2$  and  $b_5$ .

Performance of all discriminators improves faster with frequency between 0.39 and 1.5 c/object and between 5.8 and 22 c/object than is predicted from the bandwidths of the images. A slope steeper than  $-1$  means that there is more information for discriminating letters in higher frequency bands even when the number of independent samples is kept the same in each band. Once sampling density is controlled, just how much information letters happen to contain in each frequency band is an ecological property of upper-case letters.

*Increasing spatial localization with increasing frequency band.* From the human observer's point of view, the letter information in low-pass filtered images is spread out over a large portion of the total image array. In high spatial-frequency images, the letter information is concentrated in a small proportion of the total number of pixels. In high spatial-frequency images, a human observer who knows which pixels to attend will experience an effective  $s/n$  that is higher than an observer who attends equally to all pixels. In this respect, humans differ from an ideal discriminator. The ideal discriminator has unlimited memory and processing resources, does not explicitly incorporate any selective mechanism into its decision, and uses the same algorithm in all frequency bands. Information from irrelevant pixels is enmeshed in the computation but cancels out perfectly in the letter-decision process. To understand human performance, however, it is useful to examine how, with our size-scaled spatial filters, letter information comes to be occupy a smaller and smaller fraction of the image array as spatial frequency increases.

Here we consider three formulations of the change in the internal structure of the images with increasing spatial frequency: (1) spatial localization; (2) correlation between signals; and (3) nearest neighbor analysis. We have already noted that, in our images, the information-rich pixels become a smaller fraction of the total pixels as frequency band increases. Indeed, this reduction can be estimated by computing the information transmitted at any particular pixel location or, more appropriately for estimating noise resistance, by computing the variance of intensity (at that pixel location) over the set of 26 alternative signals.

To demonstrate the degree of increasing localization with increasing frequency, the variance (over the set of 26 letter templates) was computed at each pixel location  $(x, y)$ . *Total power*, the total variance, is obtained by summing over pixel locations. The number of pixel locations needed to achieve a specific fraction of the total power is given in Fig. 8, with frequency band as a parameter. These curves describe the spatial distribution of information in the latter templates. If all pixels were equally informative, exactly half of the total number of pixels would be needed to account for 50% of the total power. The solid curves in Fig. 8 show that the number of pixels needed to convey any percentage of total signal power, decreases as the

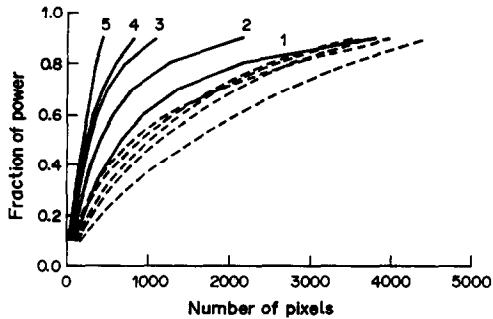


Fig. 8. Fraction of total power contained in the  $n$  most extreme-valued pixels as a function of  $n$  (out of 8100). Solid lines indicate the power fractions for signals; the curve parameter indicates the filter band. Dashed lines indicate power fractions for filtered noise fields. Although power fractions from successive bands of noise are too close to label, they generally fall in the same left-right 5-0 order as those for signal bands.

frequency band increases. These information distribution curves are an ecological property of our set of letter stimuli; different curves would be needed describe other stimulus sets.

The dashed curves in Fig. 8 were derived from random noise filtered in each of the six frequency bands ( $b_0$ - $b_5$ ). The distribution of noise power is very similar between the various bands, enormously more so than the distribution of signal power. For our letter stimuli, stimulus information coalesces to a smaller number of spatial locations as spatial frequency increases.

*Correlation between signals.* A more abstract way of describing the change of information with bandwidth is to note that letters become less confusable with each other in the higher frequency bands. A good measure of confusibility is the average pairwise correlation between the 26 letter templates in each frequency band (Table 3). The average correlation between letter templates diminishes from 0.94 in band 0 to 0.31 in band 5. In a band in which templates have a pairwise correlation over 0.9, the overwhelming amount of intensity variation ("information") is useless for discrimination. Small wonder that subjects fail completely in this band. Overall, performance of the ideal discriminator and of observers improves as the correlation decreases, but there is no obvious way to use the pairwise correlation between templates to predict performance.

*Nearest neighbors.* The analysis of nearest neighbors is a useful technique for predicting accuracy by the analysis of the possible causes of errors. We can regard a filtered image  $t_i$  of letter  $i$  as a vector in a space of dimensionality 8100 ( $90 \times 90$  pixels). When noise is added, the

Table 3. Average pairwise correlations and nearest neighbors (Euclidean distance  $\times 10^{-5}$ )

| Band | Correlations | Nearest neighbor |
|------|--------------|------------------|
| 0    | 0.94         | 0.01             |
| 1    | 0.91         | 0.30             |
| 2    | 0.58         | 1.2              |
| 3    | 0.38         | 2.3              |
| 4    | 0.33         | 3.1              |
| 5    | 0.31         | 4.1              |

possible positions of  $t_i$  are described by a cloud whose dimensions are determined by the  $s/n$  ratio. A neighboring letter  $k$  may be confused with letter  $i$  when the cloud around  $t_i$  envelopes  $t_k$ . The closer the neighbor, the greater the opportunity for error. Table 3 gives the average normalized distance to the nearest neighbor in each of the bands. The increase in distance to the nearest neighbor reflects the improvement in the representation of signals as spatial frequency increases.

We consider possible causes of lower efficiency of discrimination in bands below  $b_2$ . The letters in these bands have high pair-wise correlations and the mean band frequency is less than the object frequency. This means that letters differ only in subtle differences of shading, a feature that we usually do not think of as shape. Observers would need to be able to utilize small intensity differences to distinguish between letters. To eliminate an alternative explanation (the smaller number of frequency components in the low-frequency bands), we conducted an informal experiment with a lower fundamental frequency. The fundamental frequency, which is outside the band, nevertheless determines the spacing of frequency components within the band. Reducing the fundamental frequency of the letter by one-half increases the number of frequency components in the band by a factor of 4. (A  $256 \times 256$  sampling grid was used rather than  $128 \times 128$ .) These  $4 \times$  more highly sampled stimuli were not more discriminable than the original stimuli. This suggests that the internal letter representation (template) that subjects bring with them to the experiment cannot utilize low-frequency information, even when it is abundantly available. Whether, with sufficient training, subjects could learn to use low spatial frequencies to make letter discriminations is an open question.

## SUMMARY AND CONCLUSIONS

1. Visual discrimination of letters in noise, spatially filtered in 2-octave wide bands, is

independent of viewing distance (retinal frequency) but improves as spatial frequency increases.

2. The improvement in performance with increasing spatial frequency results mainly from an increase in the objective amount of information transmitted by the filters with increasing frequency (because filter bandwidth was proportional to center frequency) which is manifested as objectively less confusable stimuli in the higher bands.

3. The comparison of human performance with that of an estimated ideal discriminator demonstrates that humans achieve optimal discrimination (a remarkable 42% efficiency) when letters are defined by a 2-octave band of spatial frequencies centered at 1 cycle per letter height (mean frequency 1.5 c/letter). This high efficiency of discrimination is maintained over a 32:1 range of viewing distances.

4. Detection efficiency was invariant over a range of retinal spatial frequencies in which the contrast threshold for detection of sine gratings (the modulation transfer function, MTF) varies enormously. The independence of detection performance and retinal size held for all frequency bands.

5. A part of the loss of human efficiency in discrimination as spatial frequency exceeded 1 c/object height may have been due to the subjects' inability to identify, to selectively attend, and to utilize the smaller fraction of information-rich pixels in the higher frequency images.

6. Finally, it is important to note that without the comparison to the ideal observer, we would not have been able to understand the components of human performance in the different frequency bands.

*Acknowledgements*—We acknowledge the large contribution of Charles Chubb to the formulation and solution of the ideal discriminator. We thank Michael S. Landy for helpful comments and Robert Picardi for skillful technical assistance. The project was supported by USAF, Life Sciences Directorate, Visual Information Processing Program, grants 85-0364 and 88-0140.

#### REFERENCES

- Barlow, H. B. (1978). The efficiency of detecting changes of density in random dot patterns. *Vision Research*, 18, 637–650.
- Barlow, H. M. (1980). The absolute efficiency of perceptual decisions. *Philosophical Transactions of the Royal Society, London B*, 290, 71–82.
- Barlow, H. B. & Reeves, B. C. (1979). The versatility and absolute efficiency of detecting mirror symmetry in random dot displays. *Vision Research*, 19, 783–793.
- Burgess, A. (1985). Visual signal detection—III. On Bayesian use of prior knowledge and cross correlation. *Journal of the Optical Society of America A*, 2(9), 1498–1507.
- Burgess, A. (1986). Induced internal noise in visual decision tasks. *Journal of the Optical Society of America A*, 3, 93.
- Burt, P. J. & Adelson, E. H. (1983). The Laplacian pyramid as a compact image code. *IEEE Transactions on Communications, Com-34*(4), 532–540.
- Campbell, F. W. & Robson, J. G. (1968). Application of Fourier analysis to the visibility of gratings. *Journal of Physiology, London* 197, 551–566.
- Carlson, C. R., Moeller, J. R. & Anderson, C. H. (1984). Visual illusions without low spatial frequencies. *Vision Research*, 24, 1407–1413.
- Chubb, C., Sperling, G. & Parish, D. H. (1987). Designing psychophysical discrimination tasks for which ideal performance is computationally tractable. Unpublished manuscript, New York University, Human Information Processing Laboratory.
- Davidson, M. L. (1968). Perturbation approach to spatial brightness interaction in human vision. *Journal of the Optical Society of America A*, 58, 1300–1309.
- Fiorentini, A., Maffei, L. & Sandini, G. (1983). The role of high spatial frequencies in face perception. *Perception*, 12, 195–201.
- Geisler, W. S. (1984). Physical limits of acuity and hyperacuity. *Journal of the Optical Society of America A*, 1, 775–782.
- Geisler, W. S. (1989). Sequential ideal-observer analysis of visual discriminations. *Psychological Review*, 21, 267–314.
- Ginsburg, A. P. (1971). Psychological correlates of a model of the human visual system. In *Proceedings of the National Aerospace Electronics Conference (NAECON)* (pp. 283–290). Ohio: IEEE Trans. Aerospace Electronic Systems.
- Ginsburg, A. P. (1978). Visual information processing based on spatial filters constrained by biological data. *Aerospace Medical Research Laboratory*, 1(2), Dayton, Ohio.
- Ginsburg, A. P. (1980). Specifying relevant spatial information for image evaluation and display designs: An explanation of how we see certain objects. *Proceedings of SID*, 21, 219–227.
- Ginsberg, A. P. & Evans, P. W. (1979). Predicting visual illusions from filtered imaged based on biological data. *Journal of the Optical Society of America A*, 69, 1443.
- Harmon, L. D. & Julesz, B. (1973). Masking in visual recognition: Effects of two-dimensional filtered noise. *Science*, 180, 1194–1197.
- Janez, L. (1984). Visual grouping without low spatial frequencies. *Vision Research*, 24, 271–274.
- Kersten, D. (1984). Spatial summation in visual noise. *Vision Research*, 24, 1977–1990.
- 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.
- Legge, G. E., Kersten, D. & Burgess, A. E. (1987). Contrast discrimination in noise. *Journal of the Optical Society of America A*, 4(2), 391–404.
- van Meeteren, A. & Barlow, H. B. (1981). The statistical efficiency for detecting sinusoidal modulation of average dot density in random figures. *Vision Research*, 21, 765–777.
- van Nes, F. L. & Bouman, M. A. (1967). Spatial modulation transfer in the human eye. *Journal of the Optical Society of America*, 57, 401–406.

- Norman, J. & Ehrlich, S. (1987). Spatial frequency filtering and target identification. *Vision Research*, 27(1), 97–96.
- Parish, D. H. & Sperling, G. (1987a). Object spatial frequencies, retinal spatial frequencies, and the efficiency of letter discrimination. *Mathematical Studies in Perception and Cognition*, 87–8. New York University, Department of Psychology.
- Parish, D. H. & Sperling, G. (1987b). Object spatial frequency, not retinal spatial frequency, determines identification efficiency. *Investigative Ophthalmology and Visual Science (ARVO Suppl.)*, 28(3), 359.
- Pavel, M., Sperling, G., Riedl, T. & Vanderbeek, A. (1987). The limits of visual communication: The effect of signal-to-noise ratio on the intelligibility of American sign language. *Journal of the Optical Society of America A*, 4, 2355–2365.
- Pelli, D. G. (1981). Effects of visual noise. Ph.D. dissertation, University of Cambridge, England.
- Shannon, C. E. & Weaver, W. (1949). *The mathematical theory of communication*. Urbana: University of Illinois Press.
- Sperling, G. (1989). Three stages and two systems of visual processing. *Spatial Vision*, 4 (Prazdny Memorial Issue), 183–207.
- Sperling, G. & Parish, D. H. (1985). Forest-in-the-Trees illusions. *Investigative Ophthalmology and Visual Science (ARVO Suppl.)*, 26, 285.
- Tanner, W. P. & Birdsall, T. G. (1958). Definitions of  $d'$  and  $n$  as psychophysical measures. *Journal of the Acoustical Society of America*, 30, 922–928.
- van Tress, H. L. (1968). *Detection, estimation and modulation theory*. New York: Wiley.
- Winer, B. J. (1971). *Statistical principles in experimental psychology*. New York: McGraw-Hill.

## APPENDIX

Both sub-ideal and super-ideal discriminators must compute estimates of the likelihood that the stimulus  $u_{k,b}$  was produced with template  $t_{i,b}$  and noise  $n_b$ , where  $k$  is the letter used to generate the stimulus,  $i$  is an arbitrary letter, and  $b$  indexes spatial frequency band. Let  $x$  be an index on the pixels of the image:  $1 \leq x \leq 8100$ , for the  $90 \times 90$  images of the experiments.

For the Monte Carlo simulations of the super-ideal discriminator, the unknown stimulus parameters,  $\alpha_{i,b}$  and  $\sigma_N^2$  are computed during stimulus construction, and their exact values are supplied to the discriminator *a priori*. The sub-ideal discriminator, however, must estimate these parameters from the data as follows.

### Sub-Ideal Parameter Estimation

Recall that stimulus contrast is modulated for any pixel  $x$  in the image:

$$u_{k,b}[x] = \beta_{i,b}[t_{i,b}(x) + n_b(x)] + q_{i,b}(x). \quad (\text{A1})$$

The scaling constant  $\beta_{i,b}$  limits range of real values for each pixel, prior to quantization, to the open interval  $(-0.5, 255.5)$ ; the addition of  $q_{i,b}[x]$ , called quantization noise, rounds off pixel values to integers.

For each bandpass filtered template  $t_{i,b}$ , we first compute the correlation  $\rho_{k,i}$  of the template to the stimulus  $u_{k,b}$ :

$$\rho_{k,i} = \frac{\sum_x u_{k,b}(x) t_{i,b}(x)}{\left\{ \sum_x [u_{k,b}(x)]^2 \right\}^{1/2} \left\{ \sum_x [t_{i,b}(x)]^2 \right\}^{1/2}}. \quad (\text{A2})$$

To compute the likelihood estimates for each template  $t_{i,b}$ , we must be able to reverse the effect of  $\beta_{i,b}$ . Thus we define  $\alpha_{i,b} = 1/\beta_{i,b}$  and choose  $\alpha_{i,b}$  so as to minimize the expression:

$$\sum_x [\alpha_{i,b} u_{k,b}(x)]^2 = \sum_x [\rho_{k,i} t_{i,b}(x)]^2. \quad (\text{A3})$$

Solving for  $\alpha_{i,b}$  gives us:

$$\alpha_{i,b} = \rho_{k,i} \left\{ \frac{\sum_x [t_{i,b}(x)]^2}{\sum_x [u_{k,b}(x)]^2} \right\}^{1/2}. \quad (\text{A4})$$

Finally we set:

$$\sigma_N^2 = \frac{1}{X} \sum_{x=1}^X [\alpha_{i,b} u_{k,b}(x) - t_{i,b}(x)]^2 \quad (\text{A5})$$

where  $X = 8100$ , the number of pixels in the image.

### Likelihood Estimation

With estimates of  $\sigma_N^2$  and  $\alpha_{i,b}$  for the sub-ideal discriminator, and the *a priori* values for the super-ideal discriminator, we can formulate a maximum likelihood estimator. By rearranging terms of equation (A1) and dividing both sides by  $\beta$  yields:

$$\frac{u_{k,b}(x)}{\beta} - t_{i,b}(x) = n_b(x) + \frac{q_{i,b}(x)}{\beta}. \quad (\text{A6})$$

Substituting  $\alpha_{i,b}$  for  $1/\beta$ , and by transposing into the frequency domain, denoted by upper-case letters and indexed by  $\omega$ , we have:

$$\alpha_{i,b} U_{k,b}(\omega) - T_{i,b}(\omega) = N_b(\omega) + \alpha_{i,b} Q_{i,b}(\omega). \quad (\text{A7})$$

Note that the left side of equation (A7) is simply a difference image between the stimulus  $U_{k,b}(\omega)$  and the template  $T_{i,b}(\omega)$ . This difference is exactly equal to the sum of the luminance and quantization noise only when the correct template is chosen ( $i = k$ ). When the incorrect template is chosen ( $i \neq k$ ) the right hand side of equation (A7) is equal to the sum of the noise sources plus some residue that is equal to  $T_{k,b}(\omega) - T_{i,b}(\omega)$ . Under the assumption that quantization noise can be modeled as independent additive noise in the frequency domain, the density  $\Lambda$  of the joint realization of the right-hand side of equation (A7) is given by:

$$\Lambda = \prod_{\omega} \frac{X}{\pi [\sigma_Q^2 \alpha_{i,b}^2 + \sigma_N^2 |F_b(\omega)|]^2} \times \exp \left[ \frac{-X |\alpha_{i,b} U_{k,b}(\omega) - T_{i,b}(\omega)|^2}{\alpha_{i,b}^2 \sigma_Q^2 + \sigma_N^2 |F_b(\omega)|^2} \right] \quad (\text{A8})$$

where  $F_b(\omega)$  is simply the kernel of filter  $b$ , in the frequency domain. Dropping the multiplicative term in equation (A8), which does not depend on the template  $T$ , and taking logs, the ideal discriminator chooses the template that minimizes:

$$\sum_{\omega} \frac{X |\alpha_{i,b} U_{k,b}(\omega) - T_{i,b}(\omega)|^2}{\alpha_{i,b}^2 \sigma_Q^2 + \sigma_N^2 |F_b(\omega)|^2}. \quad (\text{A9})$$

Finally, it is more convenient to compute the power of the quantization noise in the space domain ( $\sigma_q^2$ ) than in the frequency domain ( $\sigma_Q^2$ );  $\sigma_q^2 = \sigma_Q^2$ . Spatial quantization noise,  $q_{i,b}(x)$ , is uniformly distributed on the interval  $[-0.5, 0.5)$ , so that  $\sigma_q^2$  is computed as:

$$\int_{-0.5}^{0.5} x^2 dx \quad (\text{A10})$$

and is equal to  $1/12$ .