Classification images for detection and position discrimination in the fovea and parafovea

Dennis M. Levi  
School of Optometry, University of California, Berkeley, CA, USA

Stanley A. Klein  
School of Optometry, University of California, Berkeley, CA, USA

Classification images provide an important new method for learning about which parts of the stimulus are used to make perceptual decisions and provide a new tool for measuring the template an observer uses to accomplish a task. Here we introduce a new method using one-dimensional sums of sinusoids as both test stimuli (discrete frequency patterns [DFP]) and as noise. We use this method to study and compare the templates used to detect a target and to discriminate the target’s position in central and parafoveal vision. Our results show that, unsurprisingly, the classification images for detection in both foveal and parafoveal vision resemble the DFP test stimulus, but are considerably broader in spatial frequency tuning than the ideal observer. In contrast, the classification images for foveal position discrimination are not ideal, and depend on the size of the position offset. Over a range of offsets from close to threshold to about 90 arc sec, our observers appear to use a peak strategy (responding to the location of the peak of the luminance profile of the target plus noise). Position acuity is much less acute in the parafovea, and this is reflected in the reduced root efficiency (i.e., square root of efficiency) and the coarse classification images for peripheral position discrimination. The peripheral position template is a low spatial frequency template.

Keywords: classification images; spatial vision; detection; position discrimination; fovea and parafovea

Introduction

The human fovea is remarkably adept at judging small offsets in position. In contrast, in the parafovea, position discrimination is considerably less acute (e.g., Westheimer, 1982; Levi, Klein, & Aitsebaomo, 1985; Beard, Levi, & Klein, 1997; Levi, McGraw, & Klein, 2000). The loss of position sensitivity in peripheral vision is, in part, due to reduced peripheral contrast sensitivity and a related change in the scale of spatial analysis. However, neither the loss of contrast sensitivity nor the spatial scale shift are sufficient to explain the loss of position sensitivity in peripheral vision (Levi & Waugh, 1994). For many psychophysical tasks, performance is limited by how well the observer's template matches the stimulus. A poorly matched template results in reduced efficiency. One of our goals is to learn what strategy the human observer actually uses in performing simple visual tasks.

We recently measured and modeled the template for a Vernier acuity task using a masking paradigm. Our results showed that a template model is able to account for many features of foveal Vernier acuity, including orientation, spatial frequency, and length-tuning results that cannot be easily accounted for by standard multiscale filter models (Levi, Klein, & Carney, 2000). Interestingly, the peripheral template for Vernier acuity is not as well matched to the stimulus (in two-dimensional spatial frequency space) as the foveal template (Levi, McGraw, & Klein, 2000).

Classification images provide an important new method for learning about which parts of the stimulus are used to make perceptual decisions and provide a new tool for measuring the template an observer uses to accomplish a task. Classification images are obtained by measuring visual performance in noise, and computing the correlation between the noise and the observer's response. The result (the classification image) is a map or spatial profile, which shows which image locations influence the observer's performance. Thus, the classification image may be thought of as a behavioral receptive field (Gold, Murray, Bennett, & Sekuler, 2000). Classification images have been derived (in normal foveal vision) for detection (Ahumada & Beard, 1999), Vernier acuity (Beard & Ahumada, 1997), illusory contours (Gold et al., 2000), as well as for less well traveled visual stimuli including imagined faces and letters (Gosselin, Bonnar, Paul, & Schyns, 2001).

Here we used the powerful technique of classification images to study and compare the templates used to detect a target and to discriminate the target’s position in central and parafoveal vision.
Methods

Figure 1. Left panel. Discrete frequency pattern (DFP) test pattern. Center and right panels: DFP pattern in sum-of-sinusoid noise (the center and right panels show different noise samples). For the detection experiments, the DFP pattern was always presented in the same position, aligned with the short bright reference line on the right side of the screen, at one of four contrast levels (including 0). For the position experiments, a fixed contrast (typically 28%) DFP pattern was presented in one of three positions (aligned with the reference line, with a fixed offset above it or below it).

Stimuli

The test pattern is a discrete frequency pattern (DFP), a barlike pattern (Figure 1) composed of 11 harmonics (from 1 to 11 c/degree) all added in phase.

\[ T(y) = c \cos^{10} (\pi y) \cos(2\pi 6y) \]

where the expansion of Equation 1 gives

\[ coef = [1, 10, 47, 120, 210, 252, 210, 120, 47, 10, 1] \]

and

\[ a_m = \frac{coef_m}{\sum_n coef_n} \]

for \( m \) ranging from 1 to 11. As seen in Equation 1, the test contrast, \( c \), is defined as the peak contrast at the center of the spatial pattern, \( y = 0 \). The normalization in Equation 3 assures that Equation 2 has the same definition.

The term \( \cos(2\pi 6y) \) is the carrier and the term \( \cos(\pi y)^{10} \) is the envelope. The envelope peaks at unity, falls to 0.5 at \( y = \pm 0.117 \) degrees and is zero at \( y = \pm 0.5 \) degrees. In the frequency domain, the envelope has components ranging from 0 to 5 c/degree in 1 c/degree steps. Equation 3 gives the spectrum of components of the full stimulus. The target (Equation 1) has the advantage of being localized in both space and spatial frequency, but with a well characterized discrete frequency spectrum (neglecting the truncation outside the displayed region). One cycle of the fundamental was shown. The fundamental was 1 c/degree so that the test and noise patterns subtended 1 degree vertically. The gratings were also 1 degree horizontally so that the stimulus was square.

The noise is a one-dimensional grating consisting of the same 11 harmonics with phases and amplitudes randomized.

\[ N(x) = n \sum_m (b_m \cos(2\pi mx) + d_m \sin(2\pi mx)) \]

where \( n \) is the root mean squared (RMS) contrast of each component in Equation 4 averaged over many stimuli, and \( b_m \) and \( d_m \) are zero mean, unit variance Gaussian random numbers. In our experiments, \( n \) was 4%.

Because the test and noise patterns are matched on average in their spectral characteristics, the noise provides a very potent mask. Discrete component noise has several advantages over noise with continuous spectra. (1) Discrete component noise strength can be specified in contrast units rather than in energy density units. (2) Ideal observer predictions can be computed in a straightforward manner, as will be discussed. (3) Because the noise can be specified by a small number of coefficients, linear regression rather than reverse correlation can be used to obtain the classification image with a reduction in the number of trials needed for a given image quality (Klein & Levi, 2002). In this study, we obtained the coefficients for each run, and then averaged them.

The noise is specified by 22 Gaussian random numbers (11 for the cosine and 11 for the sine phases) with new numbers for each trial. We assume the observer uses a template to view the signal plus noise, and the observer's judgment is directly related to the template.
output. The goal of reverse-correlation is to invert the process by taking the observer’s response and knowledge of the noise and calculate the template. For the detection task, linear regression was used to predict the response based on the 11 cosine amplitudes for each of the four stimulus levels giving a total of 44 coefficients (4*11). For the position task with 3 offset levels, the 11 sine phase noise amplitudes were used for the linear regression giving 33 coefficients (3*11 frequencies). The reduction from 22 to 11 components is based on prior information. The unshifted stimulus was in cosine phase. These coefficients are unbiased, and averaging them across the four or three stimulus conditions results in 11 coefficients for both tasks. We also examined whether the classification image depends on the stimulus phase. These coefficients are unbiased, and averaging them across the four or three stimulus conditions results in 11 coefficients for both tasks. We also examined whether the classification image depends on the stimulus level.

The target and noise were presented for 0.75 sec, in an approximately 1.7-degree square field with a mean luminance of approximately 42 cd/m² with a dark surround. The stimuli were presented on a monitor using MatVis™ software (Neurometrics Institute, Berkeley, CA).

**Ideal Observer and Template**

**Observer for Detection**

Because each Fourier component of the noise has a variance of \( n^2 \) (Equation 4), the performance of the ideal observer for the detection task is easy to calculate. The ideal \( d' \) for the \( m \)th component is given by the signal strength of the \( m \)th component, \( c a_m \) (Equation 2), divided by the RMS noise strength, \( n \) (Equation 4). This ratio is \( c a_m/n \) (from Equations 2-4). The total \( d'^2 \) is given by the sum of squares of the individual \( d' \)’s:

\[
(d’_{ideal})^2 = \sum_m (c a_m/n)^2
\]

or

\[
d’_{ideal} = \frac{c}{n} \sqrt{\sum_m a_m^2}
\]

(5)

The real observer does not use an ideal template so his or her \( d' \) will be reduced compared to the ideal observer’s. The classification image that we will estimate is the template that the observer uses for the detection task. In this section we calculate \( d’_{template} \) based on using a general template (the template observer) in which the 11 coefficients have weightings, \( w_m \). The \( d' \) value is the template response to the test pattern divided by the standard deviation of the template response to noise:

\[
d’_{template} = \sum_m c a_m w_m / \sqrt{\sum_m (n w_m^2)}
\]

or

\[
d’_{template} = \frac{c}{n} \sum_m a_m w_m / \sqrt{\sum_m w_m^2}
\]

(6)

The Pythagorean sum is present in the denominator because the various noise components are uncorrelated. In the numerator, the test components add linearly because they are phase coherent. The overall magnitude of the template has no effect in Equation 6 because the same magnitude is present in the numerator and denominator. For an ideal observer, Equation 6 becomes

\[
d’_{ideal} = \frac{c}{n} \sum_m a_m^2 / \sqrt{\sum_m a_m^2}
\]

which is identical to Equation 5, as expected. The Pythagorean sum in Equation 5 can be calculated from Equation 3 to be \( \sqrt{\sum_m a_m^2} = 0.419 \). Thus for our experiments with \( n = 4 \%), the ideal observer’s threshold \( (d' = 1) \) is given by Equation 5 to be \( c_{ideal} = 4%/0.419 = 9.56\% \). Alternatively, the ideal observer would have \( d’ \) values of 1.24, 2.49, and 3.73 for our three test contrasts of 12%, 24% and 36%.

The ratio of the template observer’s \( d’ \) to that of the ideal observer is

\[
\frac{d’_{template}}{d’_{ideal}} = \frac{\sum_m a_m w_m}{\sqrt{\sum_m w_m^2} \sqrt{\sum_m a_m^2}}
\]

(7)

The quantity in Equation 7 is precisely the correlation, \( r \), between the test pattern coefficients, \( a_m \), and the template, \( w_m \). Correlation coefficients are always between -1 and +1. In foveal vision of our normal observers, the correlation coefficients are typically between 0.7 and 0.8.

**Ideal Observer for Position**

The ideal observer’s template for the position task is the difference between the pattern with a rightward and a leftward offset:

\[
Template(x) = c \sum_m a_m \{ \cos(2\pi m (x - offset)) - \cos(2\pi m (x + offset)) \}
\]

\[
= 2c \sum_m a_m \{ \sin(2\pi moffset) \sin(2\pi m x) \}
\]

\[
= 2c \sum_m c_m \sin(2\pi m x)
\]

with

\[
c_m = a_m \sin(2\pi m offset)
\]

(8)

The ideal \( d' \) for a given component is \( c a_m \sin(2\pi m offset)/n \), where the factor of 2 has been removed because we are interested in the \( d' \) versus the stimulus with no offset rather than comparing opposite offsets. The calculation of \( d' \) for the ideal observer and the template observer is identical to what we did for the case of...
detection except that $c_m$ replaces $a_m$. The $d'$ for the ideal observer is

$$d'_{\text{ideal}} = \frac{c}{n} \sqrt{\sum_{m} c_m^2} = \frac{c}{n} \sqrt{\sum_{m} \left\{ a_m \sin(2\pi m \text{offset}) \right\}^2}$$

The $d'$ of the template observer is equal to $d'_{\text{ideal}}$ times the correlation between the template $w_m$ and the coefficients $c_m$ of Equation 8.

**Psychophysical Methods**

For the detection experiments, on each trial the test pattern was presented with one of four peak contrast levels (0%, 12%, 24%, and 36%) chosen at random. The test pattern was always presented in noise, with each of the 22 noise components having an RMS contrast of 4%. The central bar of the stimulus to be detected was always aligned with the approximately 10-min long by 3-min wide bright reference line at the center of the right edge of the screen.

The observer's task was to rate the visibility of the test pattern by giving integer ratings from 1 (no test pattern) to 4 (most visible test pattern) using the computer keyboard, and the computer provided verbal feedback about the test pattern contrast. Based on a rating scale signal detection analysis (e.g., Levi, Klein, & Carney, 2000), the three non-zero contrast patterns (in noise) could be discriminated from the blank (0 contrast) with $d'$ values of, on average, about 0.85, 1.7, and 2.5. A detection run consisted of 200 trials, and classification images are based on 4 to 8 runs.

For the position experiments, on each trial the test pattern was presented with a fixed suprathreshold contrast (28% for the fovea unless otherwise stated) in one of three positions: aligned with the reference line, or one step above or below it. The test pattern was always presented in noise. The observer's task was to rate the position of the test pattern relative to the reference line by giving integer numbers from −2 (below) to 2 (above), including 0 (aligned), and verbal feedback (below, aligned, or above) was provided following each trial. To achieve a range of performance, we varied the offset step between blocks of trials (from 360 arc sec to 4.5 arc sec in the fovea and 360 arc sec to 90 arc sec in the parafovea) to provide a range of $d'$ values for discriminating the direction of offset (in noise) from near 0 to about 2. A position run consisted of 200 trials, and classification images are based on 4 to 6 runs.

Three trained observers (one of the authors and two observers who were naive about the aims of the experiment) participated. Viewing was monocular, with the untested eye occluded with a black patch, under dim room illumination. For parafoveal experiments, the stimuli were presented in the lower visual field at either 1.25 or 2.5 degrees.

**Regression Method**

The classification image was obtained by a linear regression method. The observer’s internal response is assumed to be given by the linear relationship

$$r_{k,s} = \sum_i n_{k,s,i} w_{i,s} + q_{k,s} + f_s$$

where $r_{k,s}$ is the internal response on trial $k$ of a given stimulus level, $s$; $n_{k,s,i}$ is the external noise amplitudes, where the subscript $i$ goes from 1 to 11 for the 11 spatial frequencies, the subscript $s$ goes from 0 to 3 (detection task) or −1 to +1 (position task), and $q_{k,s}$ is the internal noise plus the truncation noise that is needed to make $r_{k,s}$ an integer. Equation 10 is based on the assumption that higher-order nonlinearities are negligible. We intend to investigate this assumption in future studies. The term $f_s$ in Equation 10 is a constant that depends on the stimulus level. Because it is a constant, it will cancel when the response is cross-correlated with the zero mean noise. The subscript $k$ indicates the trial number for a given level, and goes from 1 to about 50 (200/4) for the detection task and about 67 (200/3) for the position task. As will be discussed in “Results,” we separately analyzed each stimulus level, $s$, to minimize bias. The coefficients $w_{i,s}$ are the regression coefficients that correspond to the template weighting used by the observer. These coefficients are the classification image.

In order to calculate the classification image, we make the approximation that for each stimulus level, $s$, the internal response, $r_{k,s}$, is linearly related to the observer's response. This assumption is equivalent to an assumption that the criteria were uniformly spaced. This assumption seems reasonable because the observers were encouraged to distribute their responses uniformly. The subscript $s$ enables the constant of proportionality to be included in the coefficient $w_{i,s}$ so that $r_{k,s}$ can be taken as the observer's response. How the constant of proportionality depends on the placement of criteria is considered elsewhere (Klein & Levi, 2002). The standard method to obtain the coefficients $w_{i,s}$ is to cross-correlate the responses with the external noise

$$R_{j,s} = \frac{\sum_k r_{k,s} n_{k,s,j}}{n_{\text{trials}}} = \sum_i w_{i,s} (N_{s,i,j} + Q_{s,j})$$

where $n_{\text{trials}}$ is the number of trials at a given stimulus level, and from Equations 10 and 11,

$$N_{s,i,j} = \frac{\sum_k n_{k,s,i} n_{k,s,j}}{n_{\text{trials}}}$$

$$Q_{s,j} = \frac{\sum_k (q_{k,s} + f_s)n_{k,s,j}}{n_{\text{trials}}}$$

Equation 11 can be solved by multiplying both sides by the inverse of the square matrix $N$.
The second term is noise that is of order \( n_{\text{trials}}^{0.5} \) and will be neglected in the present analysis. \( \mathbf{N} \), the noise variance-covariance, is approximately a diagonal matrix with the diagonal elements being close to \( n^2 \). In that case, Equation 14 is approximately

\[
 w_{i,s} = \sum_j R_{j,i} \left( \mathbf{N}^{-1} \right)_{i,j} - \sum_j Q_{j,i} \left( \mathbf{N}^{-1} \right)_{i,j} \tag{14}
\]

Equation 15 gives the estimate obtained by the cross-correlation method that is the most common method for estimating the classification image. All of our results will use the linear regression method of Equation 14 that provides estimates with variance lower than the cross-correlation method. Our forthcoming plots of \( w_{i,s} \) will have an ordinate with units. It is useful to consider the meaning of the magnitude of the classification image averaged across all four levels. However, it is of some interest to know how the classification image depends on contrast. The blue (corresponding to \( w_{i,0} + w_{i,1} \) for the two below threshold contrast levels 0 and 0.12) and green (for the two above threshold contrast levels 0.24 and 0.36) lines show that in the fovea there is actually very little influence of contrast. The relative independence of these classification images with contrast reflects the relatively low transducer exponents for detecting the DFP test pattern in noise. We calculated the transducer exponents from our rating scale data in two ways: by fitting a power function to \( d' \) versus contrast and fitting a power function up to \( d' = 1 \) and then a straight line constrained to have the same slope as the power function at \( d' = 1 \). These two methods gave similar exponents of 0.92 and 0.89, respectively, much lower than the exponent of 1.5 to 2 typically obtained in detection experiments and consistent with Legge, Kersten, and Burgess (1987). A linear transducer function would imply that the sensitivity to small changes is independent of test level, as indicated by the regression coefficients being independent of contrast. The foveal detection classification images are reasonably similar in the three observers, and they also appear to be reasonably well matched to the ideal observer template (dotted black line), although the humans’ secondary peaks appear to be slightly narrower than the ideal’s. Note that in these, and all the subsequent classification image figures, the ordinate has arbitrary units.

The lower panels of Figure 2 show the regression coefficients for the foveal detection experiments, corresponding to the space plots above. The meaning of the ordinate values was discussed following Equation 15. The coefficients increase reasonably linearly up to about 6 to 8 c/degree, suggesting that our observers show a bias toward higher spatial frequencies. Thus, our observers use the 11th harmonic much more than they use the 1st harmonic. There are probably two reasons. (1) There is an asymmetry in which high spatial frequencies of the noise mask low spatial frequencies more than low spatial frequencies mask high even when plotted on logarithmic axes. This asymmetry shows up in previous adaptation studies (e.g., Blakemore & Campbell, 1969; see Stromeyer & Klein, 1974 for a discussion of the asymmetry). (2) Foveal attention could also contribute. That is, to detect the low-frequency components efficiently, one would need to attend to a large region of the field; however, we speculate that foveal attention might operate more effectively over a much smaller region of the field when noise is present. The ideal coefficients (black dotted lines) are considerably more narrowly tuned than those of the human observers. The ideal coefficients have an approximately Gaussian spatial frequency tuning curve, centered at 6 c/degree, with a full width at half height of one octave (from about 4 to 8 c/degree). Our

\[
 w_{i,s} = \frac{R_{i,s}}{n^2} \tag{15}
\]

where the ordinate has units of response times noise and the denominator has units of noise squared. Thus \( w_{i,s} \) has units of response divided by noise. Because the noise is \( n = 0.04 \), \( w_{i,s} \) is 25 times the response variability. Consider, for example, \( w_{6,s} \) in Figure 2, whose value is \( w_{6,s} = 5 \). That means the 6 c/degree component of the noise contributes a variation of 5/25 = 0.2 to the response \( n_{i,s} \). A larger value of \( w_{i,s} \) means a greater variability of responses, which would produce a lower \( d' \). Thus we have the counterintuitive result that a larger classification image is correlated with reduced \( d' \) (see discussion preceding Figure 9).

Klein & Levi (2002) provide further details on the meaning of the magnitude of the classification components, \( w_{i,s} \), including a redefinition of \( w_{i,s} \) that removes the response variance so that \( w_{i,s} \) becomes the correlation between the stimulus and response.

In “Results,” we will be plotting the 11 coefficients \( w_{i,s} \) versus the 11 spatial frequencies. We will also plot the classification images given by

\[
 C_s(y) = \sum_i w_{i,s} \cos(2\pi iy) \tag{16}
\]

for the detection task, and

\[
 C_s(y) = \sum_i w_{i,s} \sin(2\pi iy) \tag{17}
\]

for the position task.

**Classification Images for Detection**

The top panels of Figure 2 show the detection templates or classification images as a one-dimensional space plot for each observer. The black dotted line is the ideal template (Equations 1-3). We can construct classification images in a number of different ways. For example, in our detection experiment, the test pattern was presented at one of four contrast levels (see “Methods”). The solid red line shows the classification image averaged across all four levels. However, it is of

Downloaded from jov.arvojournals.org on 05/05/2019
speculation is that human observers look for a bright bar surrounded by dark bars centered at the fixation point and do not attend to the bright disinhibitory side lobes. This strategy, not unlike a peak detector, results in the broad tuning of the human coefficients.

Figure 2. Top. The foveal classification images for detecting the discrete frequency pattern are presented as a one-dimensional space plot for each observer. The solid red line shows the classification image averaged across all four contrast levels. The blue line is averaged across the two lowest contrast levels (0 and 0.12) and the green line for the two highest contrast levels (0.24 and 0.36). The black dotted line shows the classification image of the ideal observer (this is the luminance profile of the stimulus). In this figure, and in the following figures, the relative height of the measured classification image has been scaled to be roughly comparable to the ideal observer image. Bottom. The regression coefficients for each observer, corresponding to the space plots above, are plotted as a function of spatial frequency. Color coding is as above. The black dotted line shows the ideal observer’s regression coefficients and is given by the coefficients, $a$, in Equation 3. The ordinate in this and subsequent figures is for the regression coefficient. The height of the ordinate has not been rescaled and is discussed in the text.

Figure 3. Classification image (left) and coefficients (right) for detection, averaged across the three observers, for each of the four contrast levels (rather than grouped as above).
It is of some interest to look at the classification image for the zero contrast condition, corresponding to \( w_0 \). Figure 3 shows the classification image (left) and coefficients (right) for detection, averaged across the three observers, for each of the four contrast levels (rather than grouped as above). The three non-zero contrast stimuli give nearly identical responses and coefficients. The zero contrast condition (blue) gives a lower response (and coefficients), that is, \( w_{i,0} \) is less than \( w_{i,s} \) for \( s > 0 \). One possibility is that the noise that observers are trying to classify might be below threshold on some trials. That is, even though the overall transducer exponent appears to be near 1, at very low contrasts there may still be some acceleration. The classification method we are using may be very sensitive to the shape of the transducer function near zero contrast. A second explanation is that the placement of criteria at the low response categories might have been chosen to be widely spread apart, producing less response variability for the blank stimulus, thus causing a smaller classification image for blanks. Additional factors that affect the magnitude of the coefficients are discussed following Figures 4 and 9.

The classification images for detection in the parafovea (Figure 4, top panels), like those of the fovea, are similar to the ideal observer, except that the negative lobes are slightly weaker than either the ideal or the human fovea, and there is considerably more dependence on contrast (consistent with the higher transducer exponents, which were on average \( \approx 1.3 \)). It is also interesting to note that the coefficient plots (Figure 4, lower panels) are reasonably flat, showing significant coefficients up to at least 8 c/degree.

The foveal classification image in Figure 3a shows that the peak of the classification image for the zero contrast stimulus averaged across observers is about 75% of the peak for the positive contrast stimuli. However, inspection of the responses for individual contrast levels (not shown) reveals that in the parafovea, the average classification peak for zero contrast is only about 20% of the peak for positive test contrasts. It is unlikely that unequal placement of criteria could account for this difference (see discussion preceding Equation 11). Our hypothesis is that in parafoveal vision, the observer has difficulty properly placing the template. This uncertainty
would degrade the amplitude of the classification image. In the presence of non-zero test contrast pedestals, the uncertainty would be reduced. However, for the zero contrast condition, there were minimal cues for test location, so a reduced classification image would be expected. This hypothesis could be implemented mathematically in terms of an accelerating \( d' \) function.

The coefficients obtained by linear regression (Equation 14) are expected to be correlated to the stimulus strength according to the signal detection function that specifies the signal/noise (\( d' \)) as a function of stimulus strength. As an extreme case, suppose the \( d' \) function has a dead zone near threshold:

\[
d' = \begin{cases} 
0 & \text{for } c < c_0 \\
kc - c_0 & \text{for } c \geq c_0 
\end{cases}
\] (18)

As mentioned above, this dead zone could be produced because the template position is unstable in parafoveal vision. For small stimulus values, including the external noise, the dead zone would produce decreased internal response and the template could be reduced.

**Detection Efficiency**

Because our experiments involve detection in noise with discrete components, it is straightforward to calculate the observer's root efficiency, the square root of efficiency, defined as the ratio actual and ideal \( d' \):

\[
\text{Root Efficiency} = \frac{d'_{\text{actual}}}{d'_{\text{ideal}}} \quad (19)
\]

Alternatively, because we found that \( d' \) is approximately inversely related to contrast threshold, root efficiency is given by the ratio of ideal to actual thresholds. The ideal threshold was calculated following Equation 5 as \( c_{\text{ideal}} = 9.56\% \). Thus root efficiency is

\[
\text{Root Efficiency} = \frac{c_{\text{ideal}}}{c_{\text{actual}}} = 9.56/c_{\text{actual}} \quad (20)
\]

Figure 5A shows plots of \( d' \) versus eccentricity for the ideal observer (Equation 5), the template observer (Equation 6), and the mean of our actual observers for our detection task for a test stimulus of 12% contrast (our lowest non-zero contrast level). The template \( d' \) is about 80% of the ideal \( d' \), at all eccentricities (note that the ideal is independent of eccentricity, and the calculated template \( d' \) shows insignificant variation). However, as expected, the human \( d' \) for a fixed contrast target falls off with eccentricity.

Figure 5B shows our observer's root efficiency plotted as a function of eccentricity. This figure makes two points. (1) The mean foveal root efficiency is approximately 0.55. This means that the ideal observer's contrast threshold is about 0.55 times that of the human observers' threshold, within the range of other studies (e.g., root efficiencies between 55 and 70%, Burgess [1985]; Burgess, Wagner, Jennings, & Barlow [1981]). (2) Not surprisingly, the human observer's efficiency falls off with eccentricity. The interesting point, however, is that there is no significant change in the template. So what accounts for the loss of efficiency? Our speculation is that although on average the template for detection does not change in the parafovea, it is more variable, perhaps as a consequence of positional uncertainty (as will be evident below).
Classification Images for Position

The foveal classification image for position depends strongly on the size of the offset (Figures 6-8, left panels). For offsets of 90 arc sec and smaller (including offsets in the hyperacuity range [Westheimer, 1975]), the template is similar, and it makes little difference whether the classification image is computed across all offsets (red lines), non-zero offsets (blue lines), or zero offset (green lines). The classification image is not simply a picture of the stimulus, rather it is a map of the spatial information that is useful for the task. Interestingly, for runs with large position offsets (e.g., 180 and 360 arc sec), the classification images are qualitatively different, and their form depends on whether they are computed from the no-offset or offset trials. The classification image from the no-offset trials is substantially smaller than the image from the offset trials. The offset trials seem to act as a pedestal, making the offset more visible. Further insight into the classification images occurs when we consider the spatial frequency tuning of the regression coefficients (Figures 6-8, right panels). The agreement of the three observers is striking. For small offsets the classification image is directly proportional to spatial frequency, a characteristic of a dipole template, as will be discussed.

The shape of the foveal classification image for position is not strongly dependent on the target contrast (Figure 9), at least for contrast levels ranging from near the detection threshold to about 5 times threshold. Over this range, only the response amplitude changes systematically in an inverse relationship between target contrast and response amplitude. This inverse relationship can be understood in terms of our template observer model for Vernier acuity (Hu, Klein, & Carney, 1993; Levi, Klein, & Wang, 1994; Carney & Klein, 1997). The template observer approach says that for a fixed offset, d’ is proportional to contrast (see Figure 15 for the human d’ data). Observers tend to spread out their responses across all five response levels, independent of d’. For a high d’ (high contrast test pattern), the variability of responses must be low because that is what produces the high d’. The low variability produces classification coefficients that are small. Based on this logic, the size of the classification image is expected to be inversely proportional to the d’.

One might wonder why the same logic does not apply to Figures 6 to 8 where the template observer and the human observer (see Figure 15) show that d’ is proportional to offset for offsets less than 180 arc sec. The answer is that for small offsets, the d’ is so small that the observer’s responses extend over all the five response categories. That is, the stimulus offset information is small compared to the noise so the template is not much affected by the test offset. Although the template has a weak response to the test offset, it responds well to the noise, producing a template with small standard errors.

Figure 6. Left. The foveal classification images for position for a range of offsets for observer J.P. The classification images for each offset have been offset vertically for ease of viewing. The classification images were computed by averaging across all offsets (red lines), positive versus negative offsets (blue dotted lines), or zero offset (green dashed lines). The classification images have been shifted vertically for clarity. Right. The regression coefficients corresponding to the space plots on the left are plotted as a function of spatial frequency. Color coding is the same as on the left. In this and subsequent figures, regression coefficients for the smallest offset are plotted at their actual values, and each larger offset has been shifted vertically by 10 units for clarity.
Figure 7. Foveal classification images (left) and regression coefficients (right) for position for a range of offsets for observer D.L. Details as in Figure 6.

Figure 8. Foveal classification images (left) and regression coefficients (right) for position for a range of offsets for observer E.N. Details as in Figure 6.
Figure 9. Foveal classification images (left) and regression coefficients (right) for position for a fixed offset (90 arc sec) at three test pattern contrast levels. Top. Observer J.P. Bottom. Observer D.L. Details as in Figure 6.
In order to assess the human observers’ strategies for the position task, we compared the performance of our three observers (red lines) with two simulated observers: an ideal observer and a peak observer (Figure 10). For offsets less than 1/4 cycle of the 6 c/degree dominant frequency (150 sec), the ideal observer prediction (green dotted line) is approximately the derivative of the stimulus. For all offsets, the ideal observer prediction is the difference between the stimulus shifted to the right and to the left as given by Equations 7 and 8. The peak observer (gray dotted line) simply responds based on the location of the peak of the luminance profile on each trial. For small offsets, \( \sin(2\pi f \text{offset}) \) is proportional to \( f \), the Fourier spectrum of a dipole. The “Appendix” presents a derivation that the spatial pattern for small offsets is approximately

\[
dipole(y) = \frac{\sin(gy) - gy \cos(gy)}{y^2},
\]

where \( g = 11.5 \text{ c/degree} \) is the cutoff spatial frequency. This pattern is similar to the \( \text{sinc}(y) \) function that is the band-limited pattern of a line. The oscillations in the template are due to the sharp cutoff at 11.5 c/degree. The implications of having a smoother cutoff will be shown in Figure 11.

The coefficients for the ideal observer are given by Equation 8. The coefficients for the peak detector were determined by modifying the data-gathering software to produce a response based on the peak detector, so the peak observer predictions are based on limited noise samples and are, therefore, not smooth.

For offsets of 90 arc sec and smaller, the human observers’ performance is a close match to the peak observer’s, but is not a very good match to the ideal. This is particularly evident in the high frequency coefficients shown in Figure 10 for small offsets. Note that although our modeling is based on the peak observer, other localized models might give similar results. For example, we have examined a centroid model that computes the center of mass of a windowed profile (e.g., Watt & Morgan, 1984; see “Appendix” for details). Figure 11A shows the coefficients predicted for our task using Gaussian windows of different sizes, including 0 min, corresponding to the peak observer. Note that only the smallest windows (SD = 0 [the peak observer] and 0.5 min) result in the coefficients increasing linearly like the peak detector (and human observers). As the window size...
increases, the coefficients begin to fall at progressively lower spatial frequencies. For comparison, the ideal spatial template is also shown. Although the human visual system may use the centroid of a wider distribution for some tasks (e.g., Hess & Holliday, 1996; Akutsu, McGraw, & Levi, 1999), for our task, the computation of position appears to very localized. Figure 11B and 11C show the space plots of the centroid detector and ideal template, respectively. As the window size increases, the ripples in the ideal template decrease (Figure 11C); however, the dominant peak and trough change hardly at all until the Gaussian window has a standard deviation of 2 arc min. Thus, the coefficients provide a much more sensitive picture of the effect of Gaussian windowing. These plots also illustrate a limitation of our approach: although the ideal observer and our classification images have only the frequencies in the stimulus, and are, therefore, band-limited, the human observer, like the centroid model, almost certainly attends to a broader range of frequencies.

Position acuity is notoriously poor in the parafovea (e.g., Westheimer, 1982; Levi et al., 1985), and that is evident in the coarse classification images (Figures 12 and 13). Unlike the fovea, the parafoveal classification images are coarse over the range of offsets over which the observers could perform. This is not simply because of reduced visibility of the target, because (1) we increased the contrast of the parafoveal target to match it in visibility to the foveal target (i.e., both were 2-2.5 times detection threshold), and (2) as shown in Figure 8, the shape of the classification image is not strongly influenced by target contrast. Note that we have not done any size scaling, because we are interested in comparing the classification images for stimuli that are identical. Clearly, this is not an efficient template for position acuity. Inspection of the coefficient plots (the right hand panels of Figures 12 and 13) shows that position analysis in peripheral vision is a low spatial frequency analysis. In the parafovea, even near threshold, the coefficients peak at about 3 c/degree and fall rapidly, whereas in the fovea, they increase more or less linearly with spatial frequency, with a slight decrease at the highest frequency.

Even at an eccentricity of just 1.25 degrees, the position template is inefficient. Figure 14 compares the classification images for a 90 arc sec offset at the fovea, 1.25, and 2.5 degrees, and it is clear that the template becomes systematically coarser as the eccentricity increases. Peripheral position judgments may be based on locating the centroid of a broad window (sigma > 2 min; see Figure 11).

Figure 11. A. Coefficients predicted for our task using Gaussian windows of different sizes. B. Space plot of the (nonband-limited) centroid detector, for different Gaussian window sizes. C. Space plot of the (band-limited) ideal template for different Gaussian window sizes.
Figure 12. Parafoveal (2.5-degree lower field) classification images (left) and regression coefficients (right) for position for a range of offsets for observer J.P. All details as in Figure 6

Figure 13. Parafoveal (2.5-degree lower field) classification images (left) and regression coefficients (right) for position for a range of offsets for observer D.L. All details as in Figure 6
Position Precision and Efficiency

Because our experiments involve judging relative position in noise, we are able to calculate the observers’ root efficiency for our position task using Equation 9. In Figure 15 (left panels), our human observers’ performance (d’) is compared with the ideal observer’s performance as a function of position offset (top) and contrast (bottom) for foveal viewing. Both human and ideal performances increase approximately linearly with offset up to about 100 arc sec, and then decline slightly. The two curves are approximately parallel, resulting in a nearly constant root efficiency (i.e., human d’/ideal d’; Figure 15, top right), of, on average, about 15% (dotted line). Similarly, both human and ideal performances (d’) increase approximately linearly with target contrast (lower left, slope on log-log coordinates of ≈ 1), resulting in a root efficiency that is essentially independent of target contrast (Figure 15, lower right). We showed that d’ should be proportional to the product of the offset and the contrast. The linearity as a function of offset is shown in the top panels of Figure 15. The linearity as a function of contrast is shown in the bottom panels of Figure 15.

Figure 16 shows that in parafovea, performance (left) and root efficiency (right) are similar to the fovea at the largest offset. However, both d’ and root efficiency fall off approximately linearly at smaller offsets (note that the ideal performance, d’, is slightly higher in Figure 16 than in Figure 15 because the fovea was tested with a higher target contrast). Interestingly, much of the parafoveal loss of efficiency at small offsets reflects a poorly matched template for position. This can be seen by the low ratio of template/ideal root efficiency (gray symbols) in the range of 15% to 50%. At offsets of 180 and 360 sec, the human d’ is very close to the template d’. Thus the position task for large offsets is mainly limited by the template precision. In contrast, the ratio of template/ideal root efficiency for peripheral detection is much higher and the detection task in peripheral vision is limited not by the template shape but possibly by the template stability.
Figure 15. In the left panels, human performance (d') is compared with the ideal observer's performance as a function of position offset (top) and contrast (bottom) for foveal viewing. The right panels show root efficiency (i.e., human d'/ideal d', top right), of, on average, about 15% (dotted line) independent of target offset (lower left) or contrast (lower right).

Figure 16. Left. Parafoveal and ideal performance (d') as a function of offset. For comparison, the gray circles show the foveal performance data. The ideal performance is higher than in Figure 14 because the parafovea was tested at a higher contrast (32%). Right. Root efficiency of the parafovea versus offset. Unlike the fovea (open circles), root efficiency falls off markedly at smaller offsets.
Discussion

Classification images provide a powerful new method for assessing the spatial information that human observers use in making psychophysical judgments. Here we introduce a new method, using one-dimensional sums of sinusoids as both test stimuli (DFP) and as noise. The small number of coefficients allows us to use a linear regression technique that provides an improvement in the variance of coefficient estimates over the cross-correlation method for determining the classification image. One caveat is that the position classification images were calculated using only odd components, because even components will tend to cancel out. In future work, we plan to look at the contribution of the wrong polarity components.

Classification Images in Foveal Vision

Because the Fourier noise components have equal variance and are uncorrelated, the ideal observer's classification image for detection is matched to the stimulus. Our results show that the real observer's classification images for detecting the DFP pattern in both foveal and parafoveal vision are broader than the ideal observer's image and with a shift toward higher spatial frequencies in the fovea. The broader tuning would be expected if the observer focused attention on the central excitatory zone and the inhibitory flanks, and didn't pay much attention to the weak disinhibition of the secondary peak (see left panel of Figure 1). In addition, the foveal classification image is rather independent of target contrast, reflecting the low transducer exponent for detection in noise.

A more surprising finding is that the classification image for foveal position discrimination is not ideal, and depends on the size of the position offset. Over a range of offsets from close to threshold to perhaps 90 arc sec, our observers appear to use a peak strategy—responding on the basis of the peak of the luminance profile of the target plus noise. This strategy is equivalent to a localized centroid or a local luminance slope mechanism. Under these conditions, there is surprisingly little difference between the classification images computed across all three stimulus levels (plus and minus offsets and no offset). The humans' behavior for small offsets is remarkably similar to the analytic equation for a band-limited dipole (a dipole sinc function) given in Equation 21 (see “Appendix”). Actually the human might have some attenuation at the highest frequencies, so the cutoff would be less than 11 c/degree. At large offsets (180 and 360 arc sec), the classification image is qualitatively different; it is neither ideal nor peak, and it depends on whether an offset is present. It has been previously suggested that large offsets are treated categorically rather than quantitatively or metrically (e.g., Kosslyn, 1987; Cowin & Hellige, 1994), and the classification images are consistent with a qualitatively different strategy for large versus small position offsets. By “categorical processing,” Kosslyn means that position is categorized as above or below (or touching versus not touching), and does not require the precise computation of distance. For large offsets, the 0-offset condition is quite flat (see Figures 6-8). We do not have a firm explanation for the results at these large offsets. It might be easy to forget the precise location of the peak, or the observer might be using a template that underestimates the shift. Interestingly, for both human and ideal performances, d’s worsen when the offset exceeds about 100 arc sec.

Classification Images in Parafoveal Vision

It is well known that position acuity deteriorates rapidly when the target falls outside the fovea (Westheimer, 1982; Levi et al., 1985; Beard et al., 1997; Levi, McGraw, & Klein, 2000), and this is reflected in the coarse classification images for peripheral position discrimination (but not detection). The peripheral position template is a low spatial frequency template, even at near threshold offsets. This can be seen by comparing the coefficients in foveal and parafoveal vision (compare the right panels of Figures 6 and 7 with those of Figures 11 and 12). In the fovea, for offsets of 90 arc sec and smaller, the coefficients increase more or less linearly with spatial frequency. In contrast, in the parafovea, the coefficients decrease rapidly above about 3 c/degree, and are quite similar at 90 and at 360 arc sec. This low-pass position template contrasts with the more nearly flat parafoveal detection template (Figure 4, lower panels), consistent with previous studies showing that peripheral position judgments are more impaired than detection. Our parafoveal results are also consistent with those of Beard and Ahumada (2000). Using the cross-correlation method, they found that in the parafovea, observers showed a broader spatial spread in the vertical (offset) direction.

Using rather different stimuli (vertical Vernier ribbons) and masks (gratings), we measured and modeled the template for ribbon Vernier acuity (Levi, Klein, & Carney, 2000; Levi, McGraw, & Klein, 2000). We found that for short ribbons, the foveal and peripheral templates were qualitatively different. In both the fovea and periphery, the strongest threshold elevations occurred at a vertical spatial frequency corresponding to the ribbon spatial frequency. However, in the periphery, they occurred at a lower horizontal spatial frequency than in the fovea. We argued that the strong foveal threshold elevation due to masking might have a similar basis as the masking of Lincoln’s face by quantization (Harmon & Julesz, 1973), and the masking of faces in Chuck Close’s paintings by “blocking” (Pelli, 1999). In these faces, the
coherent high spatial frequency masks render the faces invisible, and the observer evidently cannot access the low spatial frequency content of the face. On the other hand, in peripheral vision, we suggested that the high horizontal spatial frequencies in the mask are much less visible so that the observer can still use lower frequency filters to perform the task. The present results are consistent with the asymmetry of high frequencies masking low frequencies. They are also consistent with a very low spatial frequency analysis in the parafovea, resulting in coarse and inefficient classification images, and this may explain the very low root efficiency of the periphery with small offsets (Figure 16).

Appendix A

Peak and Centroid Position Observer

This “Appendix” provides the mathematical details for our centroid and peak observer computations for the position task. The centroid represents the center of gravity of a limited spatial distribution. In the centroid calculation, we assume Gaussian windows of several sizes. The location of the centroid of a windowed luminance pattern P(x) is given by

\[
\text{centroid} = K \int_{-\infty}^{\infty} dx \, P(x) \exp \left( -\frac{x^2}{2\sigma^2} \right)
\]  

(22)

where \( \sigma \) is the standard deviation of the Gaussian window and \( K \) is a normalization constant that will be chosen later. The normalization is not important for our psychophysical task because the observer’s ratings are based on the relative magnitude of the centroid rather than on the absolute magnitude. For example, in a two-response task where the subject says right or left, only the sign of the centroid is needed. In a five-response task, such as in this experiment, the subject must first examine the stimuli and decide on four criteria to produce five approximately equally populated response categories. The criteria will change in different runs as the stimulus offset varies.

For the stimuli used here, the pattern \( P(x) \) can be expanded in a Fourier series with integer frequencies going from 1 c/degree to 11 c/degree.

\[
P(x) = \sum_f \left\{ a_f \sin(2\pi fx) + b_f \cos(2\pi fx) \right\}
\]  

(23)

The cosine terms of \( P(x) \) do not contribute to the centroid because antisymmetry of the integrand causes those terms to vanish. Using this expansion, Equation 22 can be rewritten as

\[
\text{centroid} = \sum_f a_f F(f)
\]  

(24)

where \( F(f) \) is given by

\[
F(f) = K \int_{-\infty}^{\infty} dx \, x \exp \left( -\frac{x^2}{2\sigma^2} \right) \sin(2\pi fx)
\]  

(25)

\[
= f \exp \left(-\frac{(2\pi f)^2}{2\sigma^2}\right)
\]  

(26)

where the normalization is chosen to be \( K = (2\pi\sigma^2)^{3/2} \) to simplify Equation 26.

Equation 24 shows that the location of the centroid is proportional to a weighted sum of the stimulus coefficients. If the human observer were using a centroid mechanism, then the classification image would match the weighting function \( F(f) \). Figure 11 shows the shape of \( F(f) \) for five values of \( \sigma \): 0, 0.5, 1.0, 1.5, and 2.0 min. For \( \sigma = 0 \), \( F(f) = f \), and the template only depend on the slope of the luminance distribution at \( x = 0 \), which is equivalent to using a dipole (a pair of adjacent opposite polarity lines) template. We call this \( \sigma = 0 \) condition the peak detector because of its locality, although slope detector might be a better name for this mechanism. The normalization \( K \) has been chosen so that at low values of \( f \), \( F(f) \) is independent of \( \sigma \), as seen in Figure 11.

The function \( F(f) \) has a peak at \( f = (2\pi\sigma)^3 \). For \( \sigma = 1/60 \) degree, corresponding to the third curve from the top in Figure 11, the peak is at 60/2π≈9.5 c/degree. The classification images for small Vernier offsets have peaks above 9.5 c/degree corresponding to a centroid mechanism with a Gaussian window with \( \sigma < 1 \) min. This is such a narrow window that it is reasonable to call the position mechanism a peak or slope or dipole mechanism.

In addition to the centroid mechanism, we also show the optimal template given by \( G(f) = f^a \text{env}(f) \), where \( \text{env}(f) \) is the envelope of the test pattern, \( a_n \). Equation 3. The function \( G(f) \) has an arbitrary scale factor for convenience in plotting.

One might think that the classification image with \( \sigma < 1 \) min would look like a dipole. That would be the case if our stimuli hadn’t been band limited. Because the frequency spectrum of a dipole is linearly proportional to spatial frequency, an analytic expression for the band-limited dipole is approximately given by

\[
D(y) = \int_0^g df \, f \sin(2\pi fy) = \frac{\sin(2\pi gy) - gy \cos(2\pi gy)}{y^2}
\]  

(27)

Because the stimuli go from the 1st through the 11th harmonic, we take \( g = 11.5 \) to be the upper limit of integration. This upper limit is obtained by converting the summation over discrete frequencies to an integral by replacing each discrete component at \( f \) with a rectangular distribution going from \( f-0.5 \) to \( f+0.5 \).
Acknowledgments

Supported by National Eye Institute Research Grants RO1EY01728 and RO1EY04776 and Core Grant P30EY07551. We are especially grateful to Hope Queener for developing the postprocessing tools using Matlab. Commercial Relationships: None.

References


