Efficient integration across spatial frequencies for letter identification in foveal and peripheral vision

Anirvan S. Nandy
Department of Psychology, University of Southern California, Los Angeles, CA, USA

Bosco S. Tjan
Department of Psychology and Neuroscience Graduate Program, University of Southern California, Los Angeles, CA, USA

Objects in natural scenes are spatially broadband; in contrast, feature detectors in the early stages of visual processing are narrowly tuned in spatial frequency. Earlier studies of feature integration using gratings suggested that integration across spatial frequencies is suboptimal. Here we re-examined this conclusion using a letter identification task at the fovea and at 10 deg in the lower visual field. We found that integration across narrow-band (1-octave) spatial frequency components of letter stimuli is optimal in the fovea. Surprisingly, this optimality is preserved in the periphery, even though feature integration is known to be deficient in the periphery from studies of other form-vision tasks such as crowding. A model that is otherwise a white-noise ideal observer except for a limited spatial resolution defined by the human contrast sensitivity function and using internal templates slightly wider in bandwidth than the stimuli is able to account for the human data. Our findings suggest that deficiency in feature integration found in peripheral vision is not across spatial frequencies.

Keywords: spatial frequency channels, summation, letter identification, fovea, periphery


Introduction

Reading and object recognition are usually associated with the clear and sharp central vision that is afforded by the fovea. Compared to the fovea, the periphery is far less capable of these types of form vision, even after its poor spatial resolution has been compensated for by magnification and contrast enhancement. For example, reading in the periphery is laboriously slow and objects often cannot be identified in a cluttered scene. At present, we do not sufficiently understand why peripheral form vision is qualitatively inferior to that of central vision. This is not only a crucial issue for basic research but is also vital for developing effective rehabilitation regimens and adaptive technologies for patients with central vision loss.

Past studies, particularly those on “crowding” (the inability to identify objects against a cluttered background), suggest that peripheral deficits in form vision result from inadequate selection and inappropriate integration of simple features into complex features at a relatively early stage of visual processing (Blake, Tadin, Sobel, Raissian, & Chong, 2006; He, Cavanagh, & Intriligator, 1996; Levi, Hariharan, & Klein, 2002; Nandy & Tjan, 2007; Pelli, Palomares, & Majaj, 2004). However, the specific nature of these deficits in feature selection and integration is unknown.

A conventional starting point for investigating feature processing is to measure an observer’s spatial tuning properties when performing a form-vision task. Chung, Legge, and Tjan (2002) found that for both foveal and peripheral vision, the peak spatial tuning frequency and tuning bandwidth for identifying an isolated letter can be adequately modeled by an ideal observer with a limited spatial resolution as described by the subject’s contrast sensitivity function (CSF) at the test eccentricity. Chung and Tjan (2007) extended this result to letter identification under crowding. Because the ideal-observer models in these studies made use of all stimulus-level information within the imposed spatial resolution limits, their results imply that feature selection along the dimension of spatial frequency (for the purpose of letter identification) is optimal in both the fovea and the periphery.

In the current study, we ask the complementary questions: Is feature integration across spatial frequencies efficient in the fovea and in the periphery for the identification of an isolated letter? “Feature” refers to any aspect of a stimulus that carries information relevant to a given task. A feature is therefore task dependent. A feature that is useful for detecting a target may not be useful for discriminating it from a distracter. Letters, similar to most visual forms, comprise of a broad range of spatial-frequency components that bear shape information. Efficient integration across spatial frequencies is thus a prerequisite for efficient form vision.

Classical findings on grating detection and discrimination suggest that integration across features is inefficient when there is a large difference in the spatial frequency of
the features (e.g., Graham, Robson, & Nachmias, 1978; Quick, Mullins, & Reichert, 1978). For example, the probability of detecting a compound grating that is comprised of two simple gratings with spatial frequencies more than a factor of two (one octave) apart is found to be approximately equal to the sum of the probabilities of detecting each of component gratings presented alone. This “probability summation” is consistent with the idea that each component grating is detected by a different narrowly tuned spatial frequency channel (Blakemore & Campbell, 1969; Campbell & Robson, 1968), which is “blind” to the other grating of a very different spatial frequency.

Probability summation is essentially the absence of summation (integration). In comparison, the probability of detection for an ideal observer (a statistically optimal decision rule) is determined by the sum of the contrast energy of the component gratings. Energy summation, which is optimal in this case, predicts a much larger effect of summation than probability summation when the detectabilities of the components are comparable. Moreover, when the contrast energy of the components are sub-threshold, while their sum is supra-threshold, probability summation predicts the compound grating to be sub-threshold, while energy summation predicts it to be supra-threshold.

Let us now return to letter identification, a more complex form-vision task. Our goal was to measure and compare how efficiently the foveal and peripheral vision systems integrate features across spatial frequencies. We wanted to determine if the form-vision deficits in the periphery are attributable to suboptimal feature integration along the spatial-frequency dimension. We adapted a typical summation paradigm. In the critical experiment, we measured contrast thresholds for identifying band-pass filtered letters at two frequencies that are two octaves apart ($f/2$ and $2f$). We also measured the threshold for identifying letters composed by adding these two frequency components together.

To preview, we found the periphery to be equally efficient at integrating features across spatial frequencies as the fovea. More surprisingly, both fovea and periphery exhibit optimal summation in the letter identification task, in sharp contrast to the sub-optimal probability summation for a simple grating detection task. The significant implications of these results for identifying form-vision deficits in the periphery and for the understanding of form-vision processes in general will be addressed in the General Discussion section.

**Analysis procedure**

In this section, we will describe the necessary steps to quantify the extent of integration across spatial frequencies. A broad-band stimulus like the letter “a” shown in Figure 1A can be decomposed into narrow-band components by a set of band-pass filters whose center frequencies form a progression in frequency space (Figure 1B).

We can then choose two narrow-band components that are sufficiently far apart in frequency space so as to have negligible overlap in their frequency contents. Moreover, if the separation is further than the typical bandwidth of a spatial-frequency channel (between 1 to 2 octaves), we can ensure that the components will be processed by different spatial frequency channels in the visual front-end. Let the center frequencies of these components be $f_1$ and $f_2$, respectively. We can form a composite stimulus by simply adding the two narrow-band components (Figure 1C).

Let $c_f$ be the contrast threshold for identifying a narrow-band letter at center frequency $f$ and $CS_f = 1/c_f$ be the corresponding contrast sensitivity. If there is no overlap in the frequency domain between the components with center frequencies $f_1$ and $f_2$ (i.e., the components are orthogonal, with a zero dot product), then for an ideal observer limited by an additive Gaussian equivalent input noise, the contrast sensitivity for the composite can be predicted from those for the components (Appendix C):

$$CS_{f_1+f_2}^2 = CS_{f_1}^2 + CS_{f_2}^2$$  \hspace{1cm} (1)

To quantify a visual system’s ability to integrate information across spatial frequencies, we define an index of integration $\Phi$ as follows:

$$\Phi = \frac{CS_{f_1+f_2}^2}{CS_{f_1}^2 + CS_{f_2}^2}$$  \hspace{1cm} (2)

This index has a simple interpretation if we assume that the visual system is like a Gaussian-noise-limited maximum a posteriori ideal observer but differs from it by making use of only a constant fraction of the signal-to-noise ratio of the stimulus (Legge, Kersten, & Burgess, 1987; Peli, 1990; Tjan, Braje, Legge, & Kersten, 1995; see also Appendix C). If the component stimuli are handled independently by the visual system and if the information across frequency channels is optimally combined, then $\Phi$ will be equal to 1. If on the other hand the information across the frequency channels is not optimally combined, $\Phi$ will be less than 1.

There exists a third possibility where $\Phi$ can be greater than 1. In this case, even though the component stimuli are far apart in frequency space, the underlying spatial mechanisms utilized for the task are not independent. Example scenarios include inappropriately broad channels for the orthogonal components and performance-dominating “late” noise situated after the outputs of the spatial mechanisms have been combined. In the first scenario, thresholds for identifying the orthogonal components in...
the stimulus are elevated by the use of inappropriately broad channels, which admit more noise from positions or spatial frequencies where there is no signal. The integration index is greater than 1.0 not because the composite is more efficiently identified, but because the components are poorly identified. In the second scenario, the threshold for the composite is lower than expected because there is only one dominant noise source as opposed to two when the components are combined.

The three categories of the integration index are summarized as follows:

\[ \Phi \begin{cases} 
< 1 & \text{sub-optimal integration} \\
= 1 & \text{optimal integration} \\
> 1 & \text{non-orthogonal integration}
\end{cases} \tag{3} \]

We note that this interpretation of the integration index is contingent on the validity of Equation 1, which in turns relies on the observer being a maximum a posteriori observer with additive Gaussian equivalent input noise. The derivation of the base case for Equation 1 relies on the linear proportionality between signal contrast and the Euclidean distance between pairs of alternatives in the internal decision space. For a 2-way discrimination task, this means that we are assuming that the psychometric function of \( d' \) vs. contrast is linear, which has been shown to be the case for a contrast discrimination task when the target contrast is above detection threshold (“supra-threshold”) (Legge et al., 1987). The details of this derivation are provided in Appendix C. Appendix C also shows that this base case can be generalized to other types of observer models for which the psychometric function is not linear. In particular, we are able to show that Equation 1 holds either analytically or is a good approximation for observers with contrast-dependent noise (“multiplicative” noise) and in supra-threshold conditions for observers with a nonlinear transducer.

We also note that in so far as Equation 1 is valid, the interpretation for the cases of \( \Phi < 1 \) and \( \Phi > 1 \) are unequivocal, as described in Equation 3. However, there exist conditions such that \( \Phi = 1 \) does not necessarily imply optimal feature integration. For example, a visual system can have an integration index of 1 or close to 1 by having nonorthogonal spatial mechanisms for the components and a suboptimal integration mechanism to combine the results. Of course, in this example, the nonorthogonality would have to precisely balance out the suboptimal integration. Furthermore, if the primary goal of measuring \( \Phi \) is to compare feature integration performance across experimental conditions (in our case, foveal vs. peripheral vision), the ambiguity in the interpretation of \( \Phi = 1 \) is less of a concern.

To empirically distinguish between the three cases of feature integration described in Equation 3, the choice of the center frequencies, \( f_1 \) and \( f_2 \), of the component stimuli is important. If there is a large difference in contrast sensitivity between the components, then it follows from Equation 1 that the sensitivity for the composite under optimal integration will be very similar to the sensitivity for the more sensitive component, making it difficult to distinguish between optimal and sub-optimal integration. We chose to address this issue by first measuring an observer’s spatial tuning function (contrast sensitivity vs. stimuli center frequency) for letter identification. Previous work (Chung et al., 2002) has shown that such a tuning function is roughly symmetric about the peak tuning frequency when frequency is expressed in log units. Given the tuning function, we chose two components of roughly equal sensitivity by selecting their center frequencies as plus and minus one octave from the peak tuning frequency:

\[ f_{\text{low}} = f_{\text{peak}} / 2 \]
\[ f_{\text{high}} = f_{\text{peak}} \times 2 \] \( \tag{4} \)

The two center frequencies are therefore two octaves apart. To ensure orthogonality at the stimulus level, we used components with a bandwidth of one octave. With this arrangement, we expect the components to be encoded by different spatial frequency channels since the typical channel bandwidth is believed to be between 1 and 2 octaves (Blakemore & Campbell, 1969; Stromeyer & Julesz, 1972). The composite stimulus was constructed by summing the two components.

**Methods**

Experiments were conducted at the fovea and at 10 deg in the lower right quadrant of the visual field. Two experiments were conducted. In both the experiments, 26 lowercase filtered target letters were presented in a letter identification task. The first experiment was to measure the spatial tuning functions; the second experiment was to measure the index of integration. In addition to the two main experiments, for the purpose of ideal-observer analysis, we also measured the contrast sensitivity function (CSF) with static sinusoidal gratings at both the fovea and 10 deg.

**Subjects**

Four subjects (ASN, BW, PLB, and JS), including one of the authors, participated in the experiments. All subjects had normal or corrected to normal vision and...
three of the subjects (BW, PLB, and JS) were naïve to the purpose of the study. All had (corrected) acuity of 20/20 or better in both eyes. Subjects viewed the stimuli binocularly in a dark room with a dim night-light. Written informed consent was obtained from the naïve subjects before the commencement of data collection.

**General procedure**

All the experiments followed a block design and each experimental condition consisted of 4 blocks with 60 trials per block. The blocks were randomly ordered with the constraint that the \(n\)th repeat of a particular center-frequency block occurred only after all frequencies had been tested for at least \(n - 1\) blocks. This was done in order to distribute the blocks of each condition evenly throughout the experiment to prevent practice effects from confounding experimental manipulations.

For Experiments 1 and 2, a filtered letter stimulus (see Stimuli section for details) was presented in each trial and the task of the subject was to identify the presented letter by pressing the appropriate key. Details of the filtering conditions are provided in the appropriate experiment sections below. The stimuli were presented at a viewing distance of 105 cm. Depending on the eccentricity condition, the subjects had to fixate either at a fixation mark at the center of the screen (foveal viewing) or at a green LED 10 deg above and to the left of the center of the screen (peripheral viewing). The letter contrast was adjusted using the QUEST procedure (Watson & Pelli, 1983) as implemented in the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) to estimate threshold contrast for reaching an accuracy level of 50% (chance is 1/26). Following Chung et al. (2002), we defined the contrast of a filtered letter in units of nominal contrast with respect to an unfiltered letter. Specifically, we assign a (nominal) contrast of \(x\)% to a filtered letter that was derived from an unfiltered letter of \(x\)% Weber contrast and a filter with a modular transfer function of unit peak amplitude.

Prior to the main experiment, letter acuity was measured at 10 deg in the inferior right peripheral field. The subjects were asked to identify any of the 26 lowercase letters, while the size of the presented letter was varied using QUEST to achieve an identification accuracy of 79%.

![Figure 1](https://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/933526/)
Stimuli

The 26 lowercase letters in Arial font (Mac OS 9) were generated separately on a background of 256 × 256 pixels. Letter size was set to twice the subject’s peripheral letter acuity in both foveal and peripheral viewing conditions. The letter size (the height of a lowercase “x”) used for each of our subjects is shown in Table 1. The letter images were each spatially filtered by a set of raised cosine filters (Alexander, Xie, & Derlacki, 1994; Chung et al., 2002; Chung, Levi, & Legge, 2001; Peli, 1990). Each filter has a bandwidth (full-width at half-height) of 1 octave and is radially symmetric in the log-frequency domain. The transfer function of the filter at radial frequency \( f_r \) is given by

\[
G(f_r) = \frac{1}{2} \left[ 1.0 + \cos \left( \pi \cdot \frac{\log(f_r) - \log(f_{\text{ctr}})}{\log(f_{\text{cut}}) - \log(f_{\text{ctr}})} \right) \right],
\]

where \( f_{\text{ctr}} \) is the spatial frequency corresponding to the peak amplitude of the filter (center frequency) and \( f_{\text{cut}} \) is the frequency at which the amplitude of the filter drops to zero (cutoff frequency). Figure 2 depicts examples of filtered images for the letter “a”.

The stimuli were displayed at the center of a 19-in. CRT monitor (Sony Trinitron CPD-G400) and the monitor was placed at a distance of 105 cm from the subject. After calibration and gamma-linearization, the monitor has 11 bits (2048 levels) of linearly spaced contrast levels. The experiments and the monitor were controlled from a Mac G4 running OS 9.2.2. All 11 bits of the contrast levels were addressable to render the stimulus for each trial. This was achieved by using a passive video attenuator (Pelli & Zhang, 1991) and a custom-built contrast calibration and control software implemented in MATLAB. Only the green channel of the monitor was used to present the stimuli.

The stimuli were presented according to the following temporal design: (a) a fixation beep immediately followed by a fixation screen for 500 ms, (b) stimulus presentation for 250 ms, (c) subject response period (variable) with positive feedback beep for correct trials, and (d) 500-ms delay before onset of the next trial.

On each trial, we collected the identity and contrast of the target letter, and the response of the subject for subsequent data analysis.

Contrast sensitivity functions

For the purpose of an ideal-observer analysis, the contrast sensitivity function (CSF) was measured using a 2IAFC task in which vertically oriented cosine phase Gabors were presented in one of two intervals. For each trial, the task of the subject was to identify the interval in which the grating was presented. Gabors were presented at 0.1, 1.0, 2.0, 4.0, 8.0, and 16.0 cycles/deg for the foveal viewing condition and at 0.5, 1.0, 2.0, 4.0, and 8.0 cycles/deg for peripheral viewing. The subjects were tested with one spatial frequency in each block, and there were 4 blocks of 60 trials each for each spatial frequency. The space constant (size at 1/e) of the Gaussian envelope of the gratings was set to 4.7 deg. QUEST procedure was used to adjust the contrast of the gratings to achieve an accuracy level of 79%.

### Table 1

<table>
<thead>
<tr>
<th>Subject</th>
<th>Letter size</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASN</td>
<td>0.96(^\circ)</td>
</tr>
<tr>
<td>BW</td>
<td>1.15(^\circ)</td>
</tr>
<tr>
<td>PLB</td>
<td>1.11(^\circ)</td>
</tr>
<tr>
<td>JS</td>
<td>1.18(^\circ)</td>
</tr>
</tbody>
</table>

The unfiltered letters are passed through a set of unit-gain raised cosine filters (Equation 5) at different center frequencies in half-octave steps.

![Figure 2](https://jov.arvojournals.org/pdfaccess.ashx?url=data/journals/jov/933526/)
To incorporate the CSFs into a model, the data points at which the measurements were made were fitted by a bi-parabolic function for the fovea and a biphasic function (flat at low frequencies followed by parabolic roll-off) for the periphery. Appendix A specifies the equations used for the fit.

**Experiment 1: Measuring the letter tuning function (LTF)**

The letter tuning function (Chung et al., 2002) or LTF was measured by presenting letter stimuli filtered with a set of six unit-gain 1-octave wide raised cosine filters (Equation 5) whose center frequencies \( f_{\text{cen}} \) were logarithmically spaced at 1.25, 1.77, 2.5, 3.54, 5.0, and 7.07 cycles/letter, respectively. Figure 2 shows an example of the filtered stimuli. For each trial, the nominal contrast of the filtered letters was adjusted using a QUEST procedure to achieve a performance threshold of 50% correct (chance was 1/26 or 3.8%). There were 240 trials per center frequency, broken into 4 blocks of 60 trials each.

**Results**

Figure 4A shows the LTF (contrast sensitivity for letter identification vs. filter center frequency) obtained for our subjects in the fovea (blue) and in the periphery (red). Each data point was estimated from 240 trials with QUEST. The 95% confidence intervals were estimated using a bootstrap procedure (Efron & Tibshirani, 1993). The six data points for each viewing condition were well described by a parabolic function in log–log \( r^2 > 0.97 \). The equation of the parabolic function is given by

\[
\log(CS_f) = \log(A) - \frac{4}{\sigma^2 \log(2)} \left( \log(f) - \log(f_{\text{peak}}) \right)^2 \tag{6}
\]

where \( CS_f \) is the contrast sensitivity at frequency \( f \), \( A \) is the peak sensitivity, \( f_{\text{peak}} \) is the peak tuning frequency, and \( \sigma \) is the bandwidth of the function in octaves.

The peak tuning frequencies as determined by the fitted parabolic functions are shown in Table 2 in units of cycles/letter (1 letter = 1 x-height). For the same letter size, the average peak tuning frequency in the periphery was lower than that in the fovea \( t(6) = 3.7258, p < 0.01 \). The average tuning bandwidth was about 2.05 and 1.69 (octaves) in the fovea and periphery, respectively (Table 2). The tuning bandwidths in the periphery were slightly but significantly lower than that in the fovea \( t(6) = 7.3294, p < 0.01 \). As will be shown with the ideal-observer analysis, an LTF with a narrower bandwidth is indicative of the observer using perceptual templates of broader bandwidths. These results are generally consistent with those obtained in Chung et al. (2002). By using the same letter size in both the fovea and the periphery conditions, the current study is more sensitive to differences in spatial tuning properties between foveal and peripheral vision.

**Experiment 2: Estimating the integration index (\( \Phi \))**

For each subject in Experiment 1, we generated the stimuli for estimating the integration index (Equation 2) with respect to the subject’s peak tuning frequencies using Equation 4. Specifically, we generated (a) the low frequency component by setting the center frequency of the raised cosine filters (Equation 5) to one half of \( f_{\text{peak}} \); (b) the high frequency component by setting the center frequency to twice \( f_{\text{peak}} \); and (c) the composite stimulus by summing the two components at the same contrast ratio as the components would in an unfiltered letter (i.e., a contrast ratio of 1 in units of nominal contrast). See Figure 3 for an example of the stimuli for this experiment. Akin to Experiment 1, we measured the nominal contrast thresholds of these three stimulus categories by using an adaptive QUEST procedure to attain a performance level of 50% accuracy. The integration index was then calculated according to Equation 2.

**Results**

Figure 4B shows the contrast sensitivities for the component and the composite stimuli overlaid on the LTF curves obtained in Experiment 1. As expected, the contrast sensitivities of the components lie on or very close to the respective LTF curves indicating the robust test-retest reliability. The sensitivities to the composite stimuli are significantly higher than that of the components for all subjects and all viewing conditions.

Figure 4C shows the integration indices obtained from the fovea (blue) and periphery (red). The confidence

<table>
<thead>
<tr>
<th>Subject</th>
<th>( f_{\text{peak}} )</th>
<th>Bandwidth</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fovea ( 10^\circ )</td>
<td>Fovea ( 10^\circ )</td>
</tr>
<tr>
<td>ASN</td>
<td>2.54</td>
<td>1.96</td>
</tr>
<tr>
<td>BW</td>
<td>2.99</td>
<td>2.09</td>
</tr>
<tr>
<td>PLB</td>
<td>2.46</td>
<td>2.14</td>
</tr>
<tr>
<td>JS</td>
<td>2.82</td>
<td>2.00</td>
</tr>
</tbody>
</table>

Table 2. Peak tuning frequencies (cycles/letter) and bandwidths (full-width at half-height in octaves) of the letter tuning functions obtained in Experiment 1.
spatial frequency components and the composite stimuli for the spectra in frequency space, they have nonoverlapping spectra.

The typical finding in grating detection or discrimination, integration across spatial frequency components two octave apart was optimal for all four subjects in the fovea. Even more surprisingly, peripheral integration was optimal for subjects ASN, BW, and PLB. For subject JS, we find that Φ is significantly greater than 1 in the periphery. This could be due to the nonorthogonality of the underlying mechanism or perceptual templates used to encode the composite. We will return to a more detailed discussion of this in the section on the Ideal observer analysis.

Also shown in Figure 4C (green symbol) are the predicted integration indices for probability summation (see Equation D2). This would be the case if the visual system were to use information from the component it was most sensitive to, as opposed to integrating information from both components. Subjects’ performance in both foveal and peripheral viewing conditions was significantly better than that predicted by probability summation.

To further confirm that these results are qualitatively different from those obtained with grating detection and discrimination, we measured contrast thresholds for subjects ASN, BW, and PLB in a compound-grating orientation discrimination task (Appendix D). The spatial frequencies of the components gratings were the same as the center frequencies of the component letters in Experiment 2. Indeed, integration was suboptimal in this case and in the range predicted by probability summation (Figure D1).

These results suggest that for a letter-identification task, the visual system can indeed integrate information across spatial frequency channels and do so optimally in both the fovea and periphery.

What is the most concise model that can quantitatively account for our letter tuning function (LTF) and index of integration data? Chung et al. (2002) found that for both foveal and peripheral vision and over a range of letter sizes, the LTFs can be accurately modeled by an ideal observer with a limited spatial resolution defined as the subject’s contrast sensitivity function (CSF) at the test eccentricity. When it comes to accounting for both the peak tuning frequency and tuning bandwidth, this CSF-limited ideal-observer model has no free parameter. Chung et al. essentially argued that to a good approximation, the visual system is optimal in selecting the spatial-frequency features for a letter identification task, after discounting a linear limitation in spatial resolution. Their model provided a mechanistic explanation of the observed letter channel (Solomon & Pelli, 1994) without postulating any specialized channels.1

We measured the CSF for subjects ASN, BW, and PLB at the test eccentricities used in Experiments 1 and 2. The CSFs and the fits to an analytic form (Appendix A) are shown in Figure 5. CSF for subject JS could not be collected due to lack of availability of the subject.

When applied to the composite letters used in Experiment 2, this CSF-limited ideal-observer model predicts an integration index of 1.0, which matched the results of all of our subjects in the foveal condition, and three of the four subjects in the peripheral condition. The model, however, cannot account for the integration index of 1.68 of JS in the periphery. Quantitatively, this model also slightly overestimated the tuning bandwidth for all subjects and underestimated the peak tuning frequency for the fovea condition for subject BW. Similar quantitative deviations were observed in Chung et al. (2002) as well as in Chung and Tjan (2007).

The CSF-limited ideal-observer model of Chung et al. does not include any front-end spatial frequency channels. The presence of low-level spatial frequency channels followed by transducer nonlinearity will impose a lower bound on the bandwidth of the perceptual templates utilized by an observer. This is because if, for example, everything is seen through two-octave channels, then the basic features that comprise a perceptual template cannot be less than two octaves wide. We can capture this effect of front-end spatial-frequency channels by specifying the bandwidth of the templates used by the ideal-observer model.

We simulated a family of CSF-limited ideal-observer models, each using internal templates of a different bandwidth, ranging from 1 (same as the stimuli) to 8 octaves. The center frequency of the templates matches the center frequency of the signal. We also tested a model that uses unfiltered letters as templates (i.e., templates of infinite bandwidth). These letter templates were generated by filtering the unfiltered letter images by raised cosine
Figure 4. (A) The letter tuning functions—LTFs: contrast sensitivity (1/threshold contrast) for identifying the 26 letters at 50% correct vs. spatial frequency—obtained from four human observers for the narrow-band letter stimuli (Experiment 1) are shown for both the fovea (blue) and the periphery (red). The dotted gray lines mark the peak tuning frequency. (B) The observers’ contrast sensitivities for identifying the stimuli used in Experiment 2 are shown overlaid on the corresponding letter tuning functions. The square data points represent the sensitivity to the low (left) and high (right) spatial frequency components, respectively; the horizontal lines represent the sensitivity to the composite stimulus. (C) Integration indices for the fovea and the periphery are shown along with the predicted indices for probability summation (green dots). Error bars are bootstrapped 95% confidence intervals.
filters (Equation 5) at the same center frequencies as the stimuli, but with different cutoff frequencies.

Our model therefore deviated from the true ideal observer in two respects: (a) its spatial resolution was limited by the subject’s CSF, and (b) the perceptual templates it used were with bandwidth that might not match that of the stimuli. Barring these limitations, the models were optimal and made their decision by maximizing the posterior probability among the list of candidate templates that are known to the system. Figure 6 shows a schematic of a model. Appendix B provides the detailed formulation of the ideal-observer model.

The CSF and template limited ideal-observer models were executed to perform the tasks that were identical to those in the human experiments (Experiments 1 and 2), using the same stimulus sets as the corresponding human.

---

Figure 5. Observers’ contrast sensitivity function (CSF) for the fovea and the periphery are shown relative to the corresponding letter tuning functions. The ranges of measurement of the CSFs encompass the letter tuning functions. The analytical functions used to fit the CSF (Appendix A) are used as the front-end filters in the ideal-observer models. Error bars are bootstrapped 95% confidence intervals.

---

Figure 6. Schematic of the ideal-observer models. The stimulus is first linearly filtered by the human CSF. The result is then perturbed by white Gaussian pixel noise. This noisy and filtered signal is then fed into an ideal observer that chooses among a set of responses (defined by the internal templates) the one that maximizes the posterior probability.
subjects. Figure 7 shows the results of the model simulations overlaid on the corresponding human data. The ideal-observer model’s tuning curves were fitted by parabolic functions (Equation 6, $r^2 = 0.97$). The internal noise of the model was adjusted such that at the 50% correct threshold criterion, the peak amplitude of the model’s most sensitive tuning function (obtained with the 1-octave templates that matched the bandwidth of the stimuli) is the same as the peak amplitude of the subject’s tuning function. Figure 8 compares the bandwidths and peak tuning frequencies of the tuning functions for our family of ideal-observer models to that of the human observers.

Results

As the bandwidth of the templates increases, the model’s peak tuning frequency remains unchanged, while its tuning bandwidth decreases (Figures 7 and 8). In general, the bandwidth of the letter tuning function (or the letter channel) is at its widest when the bandwidth of the templates matches the bandwidth of the stimuli. This theoretical result means that compared to a CSF-limited ideal observer, the efficiency of a CSF and template limited observer with templates broader than the signal decreases as the center frequency of the signal moves away from the peak tuning frequency of the CSF-limited ideal observer.
This result makes sense because the alternative would mean that at some signal frequency away from the peak tuning frequency, the efficiency would approach 1.0, which is impossible because the templates used by the observer is broader than the signal by a constant octave.

We found that a CSF and template limited ideal-observer model using templates with bandwidth around 2 octave accounts quite well for the human data in the fovea and the periphery, both in terms of matching the human letter tuning functions (Figure 7A) and in reproducing the observed integration indices (Figure 7B).² Ideal-observer models that use internal templates beyond 3 octaves wide produce tuning functions that are significantly narrower than the human curves (Figure 8A) and also lead to integration indices that are significantly greater than 1.0. The latter is due to the fact that the templates used for the

---

Figure 8. (A) Bandwidths of the models’ letter tuning functions are plotted as a function of template bandwidth. Letter tuning function bandwidth decreases monotonically with increasing template bandwidth. For comparison, the bandwidths of the human letter tuning functions are shown by the horizontal solid lines (dotted lines representing the bootstrapped 95% confidence intervals). (B) A similar comparison is made for the peak frequencies of the letter tuning functions. The model peak tuning frequencies are invariant to template bandwidth.
component stimuli are no longer orthogonal. The integration indices for these nonorthogonal channels increase monotonically with increasing degree of overlap.\(^3\)

These results corroborate a two-stage processing hypothesis: a front-end system consisting of independent feature detectors or spatial frequency channels that are narrowly tuned (about 2 octaves) followed by a back-end system that optimally combines the detected features across spatial frequencies.

Due to events beyond our control, subject JS was not available for the CSF measurement, and as a result, no model simulation could be performed on his data set. Nevertheless, we can account for JS’s peripheral integration index, which is significantly greater than 1, from the simulation results for the other subjects. Ideal-observer models that use 3–4 octave wide internal templates yield indices similar to that of JS in the periphery. This suggests that the peripheral visual field for subject JS may be mediated by spatial frequency channels that have wider bandwidths as compared to the other subjects.

### General discussion

Our results show that while contrast sensitivity to filtered letters were significantly lower in the periphery than in the fovea, the periphery is as efficient at integrating features across spatial frequencies as the fovea. That both fovea and periphery exhibit optimal integration across spatial frequencies in a letter identification task is surprising and is in sharp contrast to the sub-optimal probability summation for a compound-grating detection or discrimination task. What might account for this apparent discrepancy? There are at least two possibilities.

#### Letters as “natural” stimuli

First, the answer may lie in the overlearning of broadband edges and bars. Compound gratings are not familiar objects, and it is possible that the visual system never develops the mechanisms to optimally combine information across the spatial frequency components in arbitrary phase alignments. In contrast, natural (or overlearned) stimuli like letters are necessarily broadband, and feature-detection must be distributed across many narrowly tuned mechanisms (or channels) found in the early stages of visual processing. More importantly, unlike that of a compound grating, the narrowband components extracted from a natural stimulus are often confined to a specific set of phase configurations that lead to the sharp edges and contours and other common broadband features. It is reasonable to expect the visual system to develop mechanisms that are efficient detectors of these naturally occurring combinations of spatial frequency and phase

arrangements. Using an orientation discrimination task, Thomas and Olzak (2001) showed the existence of an efficient summing circuit for gratings that is phase-aligned. The role of familiarity is also hinted at by a finding in Meinhardt (2001), showing that learning a spatial-frequency discrimination task with gratings over a period of weeks resulted in broadening of the spatial tuning function for grating detection (i.e., better cross-frequency summation). However, to our knowledge, there is no study to date showing the emergence of optimal summation following learning a fixed but arbitrary phase alignment between the components of a compound grating.

One way to test if the efficient integration with letter stimuli depends on naturally occurring phase alignments common to broadband shapes with sharp edges is to randomly jitter the components in a composite letter stimulus. We would expect suboptimal integration in the jittered condition if specific phase alignment is required. Conversely, one can systematically measure integration indices using gratings with relative amplitude and phase that are either consistent with those found in natural images or are parametrically deviated from the naturally occurring ones.

To test whether familiarity to a fixed but unnatural phase alignment plays an important role, we can train participants to identify composite letters and gratings with components that are relatively displaced by a fixed amount and compare the integration indices between the naturally occurring phase alignment and the trained and untrained displacements.

#### Letter identification and cue integration

A second explanation for the optimal summation observed with letters is that the sensitivity to the components, against which the sensitivity of the composite was compared, is not limited by the front-end spatial frequency channels but by a second-stage feature integration process. It is worth noting that whereas optimal summation has rarely been observed in simple grating detection and discrimination tasks, optimal cue integration is relatively common in tasks such as 3-D depth and slant perception (Blake, Bülthoff, & Sheinberg, 1993; Ernst & Banks, 2002; Hillis, Watt, Landy, & Banks, 2004), texture-defined pattern detection and identification (Landy & Kojima, 2001; Meinhardt, Persike, Messenholl, & Hagemann, 2006), and visual search (Shimozaki, Eckstein, & Abbey, 2002).

A common theme among the tasks showing optimal cue integration is that the low-level features that make up the individual cues are supra-threshold such that the visibility of the individual cues are not limited by the sensitivity of the front-end feature detectors. Another common theme is that perceiving the quantity of interest (e.g., slant) using any individual cue (e.g., stereo-disparity) requires integration across low-level features (e.g., local disparity in a random dot stereogram). While different cues may be optimally integrated (e.g., disparity and texture), the low-level integration process subserving each individual cue is
often suboptimal (e.g., statistical efficiency for detecting disparity-defined edges is below 20%; Harris & Parker, 1992; Wallace & Mamassian, 2004). This last point is akin to our findings.

Let us assume that oriented narrowband gratings (Gabor) are the features extracted by the early stages of visual processing. To identify a narrowband letter, these oriented Gabor features must be integrated spatially. Different spatial configurations of these Gabor features form different letters. The highest reported detection efficiency for Gabor is about 70% (Burgess, Wagner, Jennings, & Barlow, 1981), while the highest reported efficiency for identifying 2-octave band-pass filtered letters is 42% (Parish & Sperling, 1991), with the more typical values being in the single digits (Gold, Bennett, & Sekular, 1999; Pelli, Burns, Farell, & Moore-Page, 2006; Tjan et al., 1995). The large difference in efficiency is indicative of the presence of additional processing between the stages of Gabor detection and narrowband letter identification. For our tasks, to identify a letter at a single spatial frequency band, the Gabor features must be integrated across spatial locations and orientations; to identify letters rendered at two spatial frequency bands, feature integration must also occur across spatial frequencies. Our results suggest that in spite of the inefficiency in cross-location or cross-orientation integration of visual features, the process of integration across spatial frequencies is optimal within the second-stage feature space, irrespective of eccentricity.

What constitutes a letter channel?

Our results also elaborate the notion of a letter channel (Majaj, Pelli, Kurshan, & Palomares, 2002; Solomon & Pelli, 1994). The finding of optimal summation across letter components widely separated in spatial frequencies clearly argues against treating a letter channel as analogous to a spatial frequency channel. Chung et al. (2002) have shown that to a good approximation the feature selection process underlying the identification of isolated band-pass filtered letters can be explained by the two factors that limit the performance of their CSF-limited ideal-observer models: the letter-identity information in the spatial frequency spectra of the stimulus, called the letter sensitivity function, or LSF, and the observer’s contrast sensitivity function, or CSF. Chung et al. showed that an observer’s letter tuning function (LTF) is proportional to the product of CSF and LSF, that is: \( \text{LTF}(f) = k \times \text{CSF}(f) \times \text{LSF}(f) \). This simple model argues against the need to posit an active channel selection mechanism that scales with the size of the stimulus. This model corresponds to our ideal-observer model with 1-octave templates. Our simulation results show that while the Chung et al. model is able to predict human performance for cross-spatial-frequency integration, a better fit to the human letter tuning functions can be achieved with 2-octave templates at the same center frequency as the test stimulus. With or without modification to the template bandwidth, the view expressed in Chung et al. (2002) against an active scale-dependent channel selection process remains unchanged. Our results also support the general claim of Chung et al. that the periphery visual system, like the foveal system, uses the appropriate set of spatial-frequency features from the input when performing a letter identification task.

Feature integration in the periphery

If cross-spatial-frequency selection and integration are optimal in the periphery, why is form vision so much poorer in the periphery than in the fovea? It is likely that the cause of form-vision deficits in the periphery is not about feature integration across spatial frequency, but across spatial locations. It is known that spatial uncertainty is high in the periphery (Hess & Field, 1993; Hess & McCarthy, 1993; Levi & Klein, 1996; Levi, Klein, & Yap, 1987), and there is emerging evidence that crowding in the periphery is due to uncertainty about target or feature locations (Nandy & Tjan, 2007; Strasburger, 2005). This view is consistent with the second explanation of our finding of optimal integration, i.e., from the perspective of cue combination.

Summary

Our findings suggest that both the fovea and the periphery are equally efficient at integrating information across spatial frequency channels. In fact, this integration is optimal for letter stimuli. In light of the sub-optimal summation that is the norm for compound gratings (also replicated in our study), our results suggest the existence of a stage of feature integration that combines information across spatial frequency channels efficiently, regardless of eccentricity. This stage plays an integral role in the identification of complex broadband stimuli. Taking into account the CSF and the bandwidth of the front-end spatial frequency channels, a simple ideal-observer model explains our data.

Appendix A

Analytical form for fitting the CSFs

The fovea CSF data points were fitted with the following bi-parabolic function:

\[
\log(CSF) = \begin{cases} 
\log(A) - \frac{4}{\sigma_1 \log(2)} \left( \log(f) - \log(f_{peak}) \right)^2 & \text{for } f \leq f_{peak} \\
\log(A) - \frac{4}{\sigma_2 \log(2)} \left( \log(f) - \log(f_{peak}) \right)^2 & \text{for } f > f_{peak}
\end{cases}
\]

(A1)
The periphery CSF data points were fitted with a biphasic function, which is flat to the left of the peak frequency and has a parabolic roll-off to the right:

$$\log(CS_f) = \begin{cases} \log(A) & \text{for } f \leq f_{\text{peak}} \\ \log(A) - \frac{4}{\sigma^2 \log(2)} \left( \log(f) - \log(f_{\text{peak}}) \right)^2 & \text{for } f > f_{\text{peak}} \end{cases}$$

(A2)

In both Equations A1 and A2, $A$ is the peak sensitivity of the CSF, $f$ is the spatial frequency, $CS_f$ is the contrast sensitivity at frequency $f$, $f_{\text{peak}}$ is the spatial frequency at peak sensitivity, and $\sigma$ is the bandwidth of one limb of the function in octaves. Equations A1 and A2 provide accurate interpolation of the CSF within the relevant range of spatial frequencies needed for simulating our ideal-observer model.

### Appendix B

#### Ideal-observer model

Here we briefly describe the mathematical formulation of the CSF and template limited ideal-observer models used in this study. A schematic of the model is shown in Figure 6, which is an extension of the model used in Chung et al. (2002). As in Chung et al. we assumed that a linear filter derived from the human CSF and an additive white Gaussian noise source were situated between the input signal, $S$, and the ideal observer (a maximum a posteriori (MAP) classifier). A more detailed description of the decision rule for such a classifier can be found elsewhere (Tjan et al., 1995). Here we restate the derivations that are specific to our current application.

Let $G(f)$ denote the transfer function of the CSF (Appendix A) and $N_{\text{int}}$ be the noise source with each noise pixel being normally distributed with a mean of zero and standard deviation $\sigma$. Then the resulting input that is fed into the MAP classifier is given by

$$I = \mathcal{F}^{-1}\{\mathcal{F}\{S\} \cdot G\} + N_{\text{int}},$$

(B1)

where $\mathcal{F}\{\cdot\}$ and $\mathcal{F}^{-1}\{\cdot\}$ represent forward and inverse Fourier transform operations and $S$ is the filtered letter stimulus.

Templates used by the model were generated by filtering broadband letters with raised cosine filters of various bandwidths at the same center frequency as that of the stimulus. Because the templates can be of a different bandwidth than the stimulus, the contrast of the internal templates is estimated as part of the decision procedure. In other words, the ideal-observer model has an uncertainty in stimulus contrast.

To be maximally correct on average, the MAP classifier selects the response (corresponding to the internal template $T$) that is the most probable given the input $I$:

$$P(T_j|I) = \int P(cT_j|I) \, dc = \frac{P(I|cT_j)P(cT_j)}{P(I)} \, dc \text{ (Bayes rule),}$$

(B2)

where $c$ is the internal template contrast, which is not necessarily the same as the stimulus contrast. Since the prior probability $P(cT_j)$ is constant within the range of feasible contrasts (all letters are equally likely), and $P(I)$ does not depend on $T_j$, we can further simplify the posterior probability as follows:

$$P(T_j|I) = k \int P(I|cT_j) \, dc = K \int \exp \left( -\frac{1}{2\sigma^2} \| I - cT_j \|^2 \right) \, dc$$

(B3)

The chosen response of the MAP classifier is then the signal corresponding to the template $T_{\text{MAP}}$ where

$$j_{\text{MAP}} = \arg \max_j P(T_j|I)$$

(B4)

Since taking the integral in Equation B3 is nontrivial, we adopted the following approximation by replacing integration with maximization. That is, we computed:

$$j_{\text{MAP}} = \arg \max_j P(T_j|I)$$

$$= \arg \max_j \int P(I|cT_j) \, dc$$

$$= \arg \max_j \left\{ \max_c P(I|cT_j) \right\}$$

$$= \arg \max_j \left\{ \max_c \left[ \exp \left( -\frac{1}{2\sigma^2} \| I - cT_j \|^2 \right) \right] \right\}$$

$$= \arg \max_j \left\{ \min_c \| I - cT_j \|^2 \right\}$$

(B5)

### Appendix C

#### Optimal integration for ideal observer and ideal-observer-like visual system

Claim 1: Let $c_f$ be the contrast threshold in nominal contrast units for identifying a narrow-band letter at center frequency $f$. If there is no overlap in the frequency...
domain between the components with center frequencies $f_1$ and $f_2$ (i.e., the components are orthogonal, with a zero dot product), then for an ideal observer limited by an invariant (stimulus-independent) additive white input noise (at the front-end), the contrast sensitivity (reciprocal of the threshold contrast) for the composite can be predicted from those for the components:

$$1/c_{f_1+f_2}^2 = 1/c_{f_1}^2 + 1/c_{f_2}^2$$  \hspace{1cm} (C1)

Proof: An ideal observer (a maximum a posteriori classifier) with white additive input noise makes decisions by maximizing the posterior probability given by Equation B3. The posterior is computed in terms of the squared Euclidean distance between the input image ($I$) and the letter templates at the test contrast ($cT$), normalized by the noise variance, which is a constant for invariant additive noise. As a result, the pairwise Euclidean distances between the letters at the test contrast jointly determine the average accuracy (Tjan et al., 1995, Appendix B). Hence, without loss of generality, we can confine our derivation to the Euclidean distance between two randomly chosen letters. At a given criterion accuracy, the Euclidean distance between the generic letter pair is a constant.

Let $A$ and $B$ be the templates of two letters such that at contrast $c$, the signals corresponding to these letters are $cA$ and $cB$, and their Euclidean distance is $c\|A-B\|$. For a composite at contrast $c$, the nominal contrast of its components is also by definition. This definition of contrast conveniently reflects the equality: $cA_{f_1+f_2} = c(A_{f_1} + A_{f_2}) = cA_{f_1} + cA_{f_2}$, where $c$ is the nominal contrast of $A_{f_1}, A_{f_2}$, and $A_{f_1+f_2}$. Now consider the Euclidean distance between two composite letters at nominal contrast $c$:

$$c \| A_{f_1+f_2} - B_{f_1+f_2} \|$$

$$= c \sqrt{(A_{f_1+f_2} - B_{f_1+f_2})^2}$$

$$= c \sqrt{A_{f_1}^2 + A_{f_2}^2 - 2A_{f_1}B_{f_1} - 2A_{f_2}B_{f_2} + B_{f_1}^2 + B_{f_2}^2}$$

$$= c \sqrt{(A_{f_1} - B_{f_1})^2 + (A_{f_2} - B_{f_2})^2}$$  \hspace{1cm} (C2)

This is because the dot products of the cross-frequency terms (e.g., $A_{f_1}A_{f_2}$ or $A_{f_1}B_{f_2}$) are zero since the components are orthogonal. Equivalently:

$$c^2 \| A_{f_1+f_2} - B_{f_1+f_2} \|^2$$

$$= c^2 \| A_{f_1} - B_{f_1} \|^2 + c^2 \| A_{f_2} - B_{f_2} \|^2$$  \hspace{1cm} (C3)

As noted earlier, at a given criterion accuracy, the Euclidean distance between a generic pair of letter templates is a constant. Hence,

$$c_{f_1}^2 \| A_{f_1} - B_{f_1} \|^2 = c_{f_2}^2 \| A_{f_2} - B_{f_2} \|^2$$

$$= c_{f_1+f_2}^2 \| A_{f_1+f_2} - B_{f_1+f_2} \|^2$$  \hspace{1cm} (C4)

where $c_{f_1}$, $c_{f_2}$, and $c_{f_1+f_2}$ are the threshold nominal contrasts for the components $f_1$ and $f_2$ and the composite $f_1+f_2$, respectively. Expressing the composite of Equation C4 in terms of its components using Equation C3, we have:

$$c_{f_1}^2 \| A_{f_1} - B_{f_1} \|^2 = c_{f_2}^2 \| A_{f_2} - B_{f_2} \|^2$$

$$= c_{f_1+f_2}^2 \| A_{f_1+f_2} - B_{f_1+f_2} \|^2 + c_{f_1+f_2}^2 \| A_{f_2} - B_{f_2} \|^2$$  \hspace{1cm} (C5)

Solving the set of equations in Equation C5 results in Equation C1, hence proving Claim 1:

$$1 = \frac{c_{f_1}^2 \| A_{f_1} - B_{f_1} \|^2}{c_{f_1}^2 \| A_{f_2} - B_{f_2} \|^2}$$

$$1 = \frac{c_{f_1+f_2}^2 \| A_{f_1+f_2} - B_{f_1+f_2} \|^2 + c_{f_1+f_2}^2 \| A_{f_2} - B_{f_2} \|^2}{c_{f_1+f_2}^2 \| A_{f_1+f_2} - B_{f_1+f_2} \|^2}$$

$$\Rightarrow 1 = \frac{c_{f_1+f_2}^2 + c_{f_1}^2 + c_{f_2}^2}{c_{f_1}^2 + c_{f_2}^2}$$

$$\Rightarrow \frac{1}{c_{f_1+f_2}^2} = \frac{1}{c_{f_1}^2} + \frac{1}{c_{f_2}^2}$$  \hspace{1cm} (C6)

**Corollary 1.1:** Equation C1 holds if the additive input noise of the ideal observer is a multivariate Gaussian and not necessarily uncorrelated or white.

Proof: A linear transformation (a matrix multiplication) that decorrelates the noise can be applied simultaneously to the input, noise, and letter templates, such that a white-noise ideal observer formulation is applicable to the transformed input. Let $P$ be the “pre-whitening” matrix. We can follow the derivation of Equation C1 in Claim 1 by replacing $A$ and $B$ by $PA$ and $PB$, respectively.

**Corollary 1.2:** Equation C1 holds if a linear filter is placed at the front-end of an ideal observer (as in the CSF-limited ideal observer) before the input noise.

Proof: We can filter both the input and the templates and apply a Gaussian-noise ideal observer to the transformed input. If $F$ is linear filter, we can derive Equation C1 by replacing $A$ and $B$ by $FA$ and $FB$, respectively, in Claim 1.

**Corollary 1.3:** Equation C1 holds for an observer that is otherwise a Gaussian-noise ideal observer but with sampling efficiencies $\eta_1$ for $f_1$ and $\eta_2$ for $f_2$.

Proof: If the sampling efficiency is $\eta$, then the effective Euclidean distance between a generic pair of letters is
scaled by η. In the derivation of Claim 1, this scaling can be implemented by replacing \(\|A_{f_i} - B_{f_i}\|\) by \(\eta_1\|A_{f_i} - B_{f_i}\|\) and \(\|A_{f_j} - B_{f_j}\|\) by \(\eta_2\|A_{f_j} - B_{f_j}\|\), starting at Equation C5.

**Claim 2**: Equation C1 holds if a stimulus-dependent Gaussian noise source is added after the invariant input noise, such that the variance of this second noise source is proportional to the sum of the contrast energy of the signal and the variance of the invariant input noise. Such a noise source is often called a “multiplicative” noise. The invariant noise is commonly referred to as an “additive” noise. The additive and multiplicative noises are stochastically independent.

**Proof**: As stated earlier, the relevant quantity in computing the posterior is the ratio between the squared Euclidean distance between the input and a template, and the variance of the noise (Equation B3). In the derivation of Claim 1, the variance of the noise is ignored because it is a constant. In the case of having a contrast-dependent noise, the proof of Claim 1 can be replayed to arrive at a version of Equation C6 by considering the equality in the “effective” Euclidean distance between a pair of generic terms, we have:

\[
\frac{\sigma^2 + \left( \sigma^2 + \|X_{f_1+f_2}\|^2 \right) c_{f_1+f_2}^2 / a}{c_{f_1+f_2}^2} m
\]

Divide both side of the equality by \((1 + m)\sigma^2\) and rearrange terms, we have:

\[
\frac{1}{c_{f_1+f_2}^2} = \frac{1}{c_{f_1}^2} + \frac{1}{c_{f_2}^2} + \frac{m}{(1 + m)\sigma^2 a} \times \left( \|X_{f_1}\|^2 + \|X_{f_2}\|^2 - \|X_{f_1+f_2}\|^2 \right)
\]  

**Corollary 2.1**: Equation C1 holds under supra-threshold conditions for an observer whose front-end consists of an additive “peripheral” noise, followed by a pixel-wise logarithmic compressive nonlinearity, followed by a second additive “central” noise, and is otherwise ideal.

**Proof**: Legge et al. (1987) showed in their Appendix A that for a contrast discrimination task when the pedestal is above detection threshold (supra-threshold), an observer with a nonlinear transducer situated between two invariant noise sources is equivalent to an observer with contrast-invariant periphery noise followed by a contrast-dependent central noise, without any nonlinearity in between. They did so by expressing the nonlinearity with a Taylor series and consider the net signal-to-noise ratio. They showed that the variance of the contrast-dependent noise is inversely proportional to the squared derivative of the transducer function under supra-threshold conditions. If we are to require the
variance of the central noise to be proportional to input contrast energy, as is required by Claim 2, then the derivative of the nonlinearity has to be inversely proportional to input contrast, thus implying that the nonlinearity must be logarithmic (and thus compressive) in contrast. Claim 2 applies under these conditions, implying Equation C1.

Although the derivation of Legge et al. concerns a contrast discrimination task, it is equally applicable to a n-way identification task since as we have mentioned in the derivation of Claim 1, all that is relevant for determining the performance of a maximum-a posteriori observer for an n-way discrimination task is the generic pairwise distance (\(d'\) in the case of a 2-way discrimination) between a pair of alternatives.

**Corollary 2.2:** For an observer with an expansive nonlinearity in between the peripheral and central noise, the threshold contrast sensitivities to the components and the composite stimuli approach Equation C1 in suprathreshold conditions as stimulus contrast increases.

Proof: Recall that Legge et al. (1987) showed that the variance of equivalent central contrast-dependent noise is inversely proportional to the squared derivative of the transducer function. For an expansive nonlinearity, the derivative increases as a function of contrast. High contrast thus reduces the variance of the central noise and diminishes its influence. In the limit, the central noise becomes irrelevant and the performance is only limited by the invariant periphery noise, meeting the conditions of Claim 1. Hence, Equation C1 holds in the limit of high stimulus contrast for an expansive nonlinearity.

### Appendix D

#### Grating orientation discrimination task

Our goal in this supplementary experiment was to measure the index of integration for gratings in a way that is comparable to the spatial frequencies and size of the letter stimuli in Experiment 2. The component gratings were tested at the same frequencies as that of the center frequencies of the letter components for subjects ASN, BW, and PLB in Experiment 2. For ASN, this corresponded to \(f_{\text{low}} = 1.63\) and \(f_{\text{high}} = 6.51\) cpd in the fovea and \(f_{\text{low}} = 1.3\) and \(f_{\text{high}} = 5.2\) cpd in the periphery. The compound gratings were formed by combining (in sine phase) the corresponding component gratings according to the ratio of their detection thresholds, estimated from the subject’s CSF. For example, the detection threshold in the periphery of ASN was 0.86% in Weber contrast for \(f_{\text{low}}\) and 1.92% for \(f_{\text{high}}\); as a result, the contrast ratio of the components in the composite was 0.86 (\(f_{\text{low}}\)) to 1.92 (\(f_{\text{high}}\)). Similar in principle to the definition used for letters, we define the nominal contrast of a component or the composite to be the corresponding Weber contrast of the \(f_{\text{low}}\) component in the composite.

The task was a 2-way discrimination task between horizontally and vertically oriented Gabor patches. The space constant of the Gaussian envelope of the gratings was set to the same size as that of the letter stimuli (0.96° for ASN, 1.15° for BW, 1.11° for PLB). QUEST procedure was used to adjust the contrast of the gratings to achieve an accuracy level of 74%. As with Experiment 2, the component and the composite gratings were tested in interleaving blocks. Stimulus presentation timing was identical to that of Experiment 2. The number of trials per condition was 180 divided over three blocks.

The integration index was estimated as in Experiment 2:

\[
\Phi = \frac{CS_{\text{low}}^2 + CS_{\text{high}}^2}{CS_{\text{low}}^2 + CS_{\text{high}}^2} \tag{D1}
\]

Figure D1 shows the obtained integration indices in the fovea and in the periphery. Also shown are the predicted indices for probability summation which were estimated.

![Figure D1](https://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/933526/ on 11/05/2018)
by considering only the response to the maximally sensitive component:

\[ \Phi_{\text{prob. sum.}} = \frac{(\max[CS_{\text{low}}^2, CS_{\text{high}}^2])^2}{CS_{\text{low}}^2 + CS_{\text{high}}^2} \] (D2)

As has been showed in the past (Graham et al., 1978), the process of integration of compound gratings is suboptimal and in the range predicted by probability summation. We found no practice effect on integration index across blocks of trials (highest positive change in integration index between first and final blocks was 0.15 for BW in the fovea; highest negative change was 0.22 for PLB in the periphery).

Footnote

1 We chose to analyze our results using the ideal-observer model of Chung et al. (2002) because excluding the front-end CSF filter, the model is the white-noise ideal observer formulation for our octave bandwidth letter stimuli. We note that the ideal observer tuning function of Chung et al. (2002, Figure 10) is band-pass while that of Solomon and Pelli (1994, Figures 4c and 4d) is low-pass, even though both groups used a white-noise ideal observer. This discrepancy in the ideal-observer tuning functions is superficial since the two groups measured the tuning function in different units. Following the standard engineering conventions, Solomon and Pelli measured amplitude gain per unit linear bandwidth per unit radius. In contrast, staying native to the stimulus units, Chung et al. measured “letter sensitivity” (equivalent to amplitude gain) in units of per octave bandwidth per 2π radii. Chung et al. used octave units because they tested their human observers with octave-wide filtered letters. For Chung et al., the linear bandwidth decreases in proportion to decreasing spatial frequency. As a result, letter sensitivity also decreases with frequency, reaching zero at DC (since the linear bandwidth of a one-octave width stimulus at DC is zero).

We can derive the exact formula equating the letter sensitivity \( H(f) \) of Chung et al. to the amplitude gain \( G(f) \) of Solomon and Pelli. Let \( \hat{E}(f) \) be the fraction of signal energy utilized by the ideal observer for the letter-identification task over all orientations and for all radial spatial frequencies less than \( f \). The gain of Solomon & Pelli is equivalent to

\[ G(f) = \sqrt{\frac{1}{2\pi^2 f} \frac{d\hat{E}}{df}} \] (N1)

The letter sensitivity of Chung et al. is

\[ H(f) = \sqrt{\frac{d\hat{E}}{d(\log_2 f)}} \] (N2)

Applying chain rule and rearranging terms, we have

\[ G(f) = \sqrt{\frac{1}{2\pi^2 f} \frac{d\hat{E}}{d(\log_2 f)}} \frac{d(\log_2 f)}{df} = H(f) \sqrt{\frac{1}{2\pi \ln(2)f^2}} \]

\[ H(f) \propto fG(f) \sqrt{2\pi \ln(2)} \]

Solomon and Pelli found that \( G(f) \) is low-pass. By Equation N3, the sensitivity of Chung et al., \( H(f) \), should be bandpass, which is the case. Furthermore, Equation N3 suggests that the low-frequency falloff of \( H(f) \) should approach a log–log slope of 1.0 at low radial frequencies, which is also the case (Chung et al., 2002, Figure 10).

2 BW’s fovea LTF is peculiar since it straddles the 4-octave model LTF at the low frequencies and the 1-octave model LTF at the high frequencies. Our model used a single template bandwidth and underestimated the peak tuning frequency by 0.2 octave, an error that is less than but comparable to those observed in the foveal condition in Chung et al. (2002).

3 We observe that the integration index of the model with 2-octave wide templates is very slightly but consistently lower than the observer model with the templates matched to the signal (1-octave wide). Recall that a template for the composite condition is equal to the sum of the templates for the corresponding components. A component template that is 2-octaves wide (full width at half height) can “see” very little of the other signal component in the composite condition because the 1-octave signal components are two octaves apart. Thus, as far as the narrow-band signals are concerned, these 2-octave templates appear independent. However, the component templates see identical noise in the frequency range where they overlap and are not independent with respective to noise. As a result, the signal-to-noise ratio seen by a composite template is lower than the sum of the signal-to-noise ratios seen by the component templates, hence a slight reduction in integration index. The 2-octave templates are not special. If the 1-octave signal components were three octaves apart, then we would observe a slightly decreased integration index for models with template bandwidths between one to three octaves.
Acknowledgments

This research was supported by the National Institutes of Health Grants EY016391 and EY016093 to BST.

Commercial relationships: none.
Corresponding author: Bosco S. Tjan.
Email: btjan@usc.edu.
Address: Department of Psychology, 3620 South McClintock, SGM 501, University of Southern California, Los Angeles, CA 90089-1061, USA.

References


