Stimulus information contaminates summation tests of independent neural representations of features

Steven S. Shimoza

Miguel P. Eckstein

Craig K. Abbey

Many models of visual processing assume that visual information is analyzed into separable and independent neural codes, or features. A common psychophysical test of independent features is known as a summation study, which measures performance in a detection, discrimination, or visual search task as the number of proposed features increases. Improvement in human performance with increasing number of available features is typically attributed to the summation, or combination, of information across independent neural coding of the features. In many instances, however, increasing the number of available features also increases the stimulus information in the task, as assessed by an optimal observer that does not include the independent neural codes. In a visual search task with spatial frequency and orientation as the component features, a particular set of stimuli were chosen so that all searches had equivalent stimulus information, regardless of the number of features. In this case, human performance did not improve with increasing number of features, implying that the improvement observed with additional features may be due to stimulus information and not the combination across independent features.

Keywords: visual search, ideal observer, spatial frequency, orientation

Introduction

A common concept in cognitive neuroscience is the modularity of information processing within the brain. Simply put, it is assumed that different parts of the brain process different types and aspects of information (e.g., the different sensory modalities). This concept has been assumed within the visual sensory modality as well. A common aspect in the modeling of visual processing, particularly in the field of attention, proposes that different types of visual information are analyzed by separate retinotopic “feature” maps with, for example, color and motion as representative features (Neisser, 1967; Treisman & Gelade, 1980; Livingstone & Hubel, 1988; Wolfe, Cave, & Franzen, 1989). Within the domain of spatial vision, in particular, it has been speculated that orientation and spatial frequency are separable and independent features of visual analysis (Burbeck & Regan, 1983; Morgan, 1992; Heeley, Buchanan-Smith, & Heywood, 1993; Heeley & Buchanan-Smith, 1994; Vincent & Regan, 1995; Olzak & Thomas, 1991, 1992; Thomas & Olzak, 1990, 1996; Chua, 1990). Here we examined the evidence for the independent neural processing of spatial frequency and orientation in a visual search task. On each trial, one target and three distractor grating patterns appeared simultaneously at different locations on a computer display, and the observer had to choose the location of the target (see Figure 1). The target could differ from the distractors either just in orientation or just in spatial frequency (single-feature searches), or both in orientation and spatial frequency (the 2-feature search). Common models of separable and independent features derived from simple detection or discrimination tasks (Thomas, Gille, & Barker, 1982; Klein, 1985; Ashby & Townsend, 1986; Graham, 1989; Kadlec & Townsend, 1992; Wickens & Olzak, 1992) and visual search (Eckstein, 1998; Eckstein, Thomas, Palmer, & Shimozaki, 2000) make specific predictions about summation across the independent features, so that performance improves as the number of features available to perform the task increases.

In particular, in this study we consider the influence of the information content of the stimuli on the interpretation of summation across independent features. The stimulus information can be assessed by an ideal observer (see Appendix B) that predicts the best possible performance for a particular visual task, or equivalently, that uses all the information possible for a given task (Green & Swets, 1974; Barlow, 1978; Burgess, Wagner, Jennings, & Barlow, 1981). For example, the ideal observer has been used previously to explain human observers’ performance in recognizing objects across different views (Liu, Knill, & Kersten, 1995; Braje, Tjan, & Legge, 1995; Tjan, Braje, Legge, & Kersten, 1995), in
visual search as a function of number of elements (Shaw, 1980, 1982; Palmer, Ames, & Lindsey, 1993; Palmer, 1995), and in the perception of symmetry (Liu & Tjan, 1998).

![Figure 1. Stimulus for a spatial frequency feature trial in Experiment 2. One target (5 cpd vertical Gabor patch, left location) and three distractors (2 cpd vertical Gabor patches) appear against a background of white noise. The observers had to choose the location containing the target. See the text for details.](image)

One aspect important for this work is that the ideal observer only considers the visual stimulus in its analysis, and does not propose any featural analysis. As shown later, in many instances, increasing the number of features also increases the amount of stimulus information. Therefore, any improvement in human performance with increasing number of features in visual search could be due to an increase in stimulus information, rather than a result of summation across independent features. In other words, stimulus information may explain observed human visual performance that would otherwise be attributed to perceptual processing inherent to the human visual system.

In the first experiment, the differences in spatial frequency and orientation between the target and the distractors were relatively small. For these stimuli, two common models of summation across independent features predict similar increases in performance as the stimulus information model in the 2-feature search, compared to the single-feature searches. This demonstrates how both predicted and observed human performance can be consistent with both independent feature summation models and an increase in stimulus information. In the second critical experiment, all the targets were chosen so that all tasks had the same stimulus information content (equivalent ideal observer performance), regardless of the number of features available to perform the task. For these stimuli, predictions from the stimulus information and the summation models can be distinguished. The summation models predict an increase of performance in the 2-feature search, as in the first experiment, whereas the stimulus information predicts no difference in performance in the 2-feature search.

### The Ideal Observer and Stimulus Information

The ideal observer is defined as the observer that uses all available information, both in the image and the prior information, to optimize performance in the task. Thus, the ideal observer performance is only limited by the stimulus information, and not by any intrinsic sources of inefficiency in the processing of the image that might be present in human observers. As a consequence, the ideal observer can be cast as describing the amount of information available to the human observer in the stimulus, or the stimulus information. Because of its objectivity and optimality, it has been used as a tool in many visual contexts to assess the upper limit of performance, including simple detection and discrimination (Green & Swets, 1974; Barlow, 1978; Burgess et al., 1981; Kersten, 1984; Eckstein, Ahumada, & Watson, 1997), object recognition (Braje et al., 1995; Liu et al., 1995; Tjan et al., 1995), perceptual learning (Gold, Bennett, & Sekuler, 1999; Abbey, Eckstein, & Shimozaki, 2001), and reading (Legge, Klitz, & Tjan, 1997). Humans may or may not be able to use parts of this information in their own performance of the task, and, as they are not ideal, can never use all the information. The amount of information used by the human observer can be assessed by comparing the ideal observer performance with human performance (often measured as efficiency, \((d'_{\text{human}}/d'_{\text{ideal}})^2\)). Several authors have modeled how humans might be suboptimal for a given task, such as the inability to optimally use the signal information (sampling inefficiency), internal or equivalent noise, or intrinsic uncertainty (e.g., Burgess et al., 1981; Pelli, 1985; Solomon & Pelli, 1994; Eckstein et al., 1997; Gold et al., 1999).

In this study, and in other studies with ideal observer analyses, image (external) noise is needed to generate less than perfect performance for the ideal observer. For the visual search localization of a target among distractors studied in this study, the ideal observer calculates the likelihood of the data (the responses of the model) for each location, given signal presence at that location and distractor presence at the other locations. The ideal observer then chooses the location with the highest likelihood. In the case of white (uncorrelated) noise in this task, the ideal observer reduces to a cross-correlation (dot product) of an ideal template with the stimulus (Green & Swets, 1974) (see Figure 2 and Appendix B). Figure 2 depicts the ideal observer in a 2-feature search trial for Experiment 2. The target is a high spatial frequency (5 cpd) horizontal Gabor patch in the left.
location amongst 3 distractor low spatial frequency (2 cpd) vertical Gabor patches. Therefore, the target differs from the distractors along both features of orientation and spatial frequency. The ideal template in white uncorrelated noise is simply the difference between a template matching the signal and a template matching the distractor. The ideal observer computes the cross-correlation (dot product) of the ideal template with the stimulus at each location, and chooses the location with the maximum dot product (in other words, the best match with the template) across the four locations. An important aspect of the ideal observer for this study is that there is no featural analysis or description in the model, such as spatial frequency or orientation. Any improvement of the ideal observer performance across conditions simply reflects the stimulus information in the task.

The location having the most evidence for target presence along a single feature, across all available features on a particular trial. In both these models, each feature is coded separately with independent internal noise.

Linear summation has been used as a test for the independence of two features by several authors (Thomas et al., 1982; Klein, 1985; Ashby & Townsend, 1986; Graham, 1989; Kadlec & Townsend, 1992; Wickens & Olzak, 1992; Eckstein, 1998; Eckstein, et al., 2000). In the 2-feature search task (in which the target differs from the distractors along 2 features), this model assumes a linear combination of information across features, weighted by the sensitivity for each feature. This linear combination across features predicts better performance in the 2-feature search task, compared to the single-feature search tasks. Figure 3 illustrates the linear summation model in a single 2-feature trial in Experiment 2. Starting on the left, the model assumes two independent responses at each location, one corresponding to orientation ($x_{d-o}$ for the target location, $x_{d-o}$ for the distractor locations), and the other corresponding to spatial frequency ($x_{d-sf}$ for the target location, $x_{d-sf}$ for the distractor locations). These responses are weighted separately by the sensitivity of the observer to that particular feature ($d'_o$ and $d'_s$), and then summed to give a single combined response for each location ($x_{d-linear}$ for the target location, $x_{d-linear}$ for the distractor locations). The model then chooses the location with the maximum value for the combined response as the target location.

**Overview of Independent Feature Models**

Generally, two independent features, as has been proposed with respect to orientation and spatial frequency, can be characterized as two independent neural sources of information that differentiate the target from the distractors. There are two common models of summation across independent features that have been developed principally in the field of spatial vision, summarized in Appendix C. The first independent feature model is linear summation, which posits a linear combination of information across the features. The second independent feature model is probability summation, which uses a maximum value rule to choose the location having the most evidence for target presence.

![Ideal Observer Diagram](image)

Figure 2. Description of the ideal observer in a 2-feature trial in Experiment 2. First the ideal observer computes the ideal template, which is the difference between the target and the distractor. Then the ideal observer cross-correlates (takes the dot product) this difference with the stimuli at each location. The location with the maximum dot product value is the stimulus that best matches the template, and is chosen as the target location.

![Linear Summation Diagram](image)

Figure 3. Schematic of the linear summation model. A 2-feature search from Experiment 2 is depicted with the target in the left locations. An independent response is generated for both spatial frequency and orientation at each location. These responses are weighted by the $d'$ for the particular feature and summed. The location with the maximal weighted linear response ($x_{d-linear}$ or $x_{d-linear}$) is chosen as the target location.

For this model, the predicted improvement in performance in the 2-feature search task can be described geometrically (see Figure 4). In Figure 4, each dimension represents the sensitivity to a single feature in units of $d'$ from signal detection theory (Green & Swets, 1974), and linear summation is represented as the vector sum of the sensitivity for each feature. Also, independent features are...
represented as orthogonal axes for each feature, and the length of the 2-feature vector, which represents performance in the two-feature task, becomes the hypotenuse of the two single-feature vectors.

\[
d'_{\text{2-Feature}} = \sqrt{(d'_{\text{Feature 1}})^2 + (d'_{\text{Feature 2}})^2}
\]

Figure 4. Geometric description of linear summation. Each axis represents the sensitivity to each feature in \(d'\) units. The sensitivity when both feature cues are available is equal to the Euclidean vector sum of the individual sensitivities. The orthogonal axes represent independence of the two features.

The second independent features model is probability summation (Graham, 1989; Eckstein, Whiting, & Thomas, 1996; Tyler & Chen, 2000), shown in Figure 5 (see Appendix C). This model assumes an independent internal response for each feature at each location, one for orientation \(x_{t-o}\) for the target location, and \(x_{d-o}\) for the distractor locations, and one for spatial frequency \(x_{t-sf}\) for the target location, and \(x_{d-sf}\) for the distractor locations. The model then chooses the location with the maximal independent featural response (uncombined, unlike the linear summation model) as the target location. In other words, the model chooses the location with the most evidence for target presence along a single feature, amongst the evidence across all features. As the number of features available to perform the task increases, the probability of any of the internal responses to the target (along any one of the available features) taking the maximum value also increases. Thus, this decision rule predicts better performance in the 2-feature visual search task, compared to the single-feature task, similar to the linear summation model. Probability summation, however, is a weaker form of summation than linear summation, and generally predicts a smaller increase in performance in the 2-feature search, relative to the single-feature searches.

Figure 5. Schematic of the probability summation model. A 2-feature search from Experiment 2 is depicted with the target in the left locations. An independent response is generated for both spatial frequency and orientation at each location, and the location with the maximal featural response is chosen as the target location.

In the two experiments of this study, observers participated in a 4-alternative forced choice (AFC) visual search localization task. On each trial, one target Gabor patch and 3 distractor Gabor patches appeared for 200 ms against a background of Gaussian white luminance noise (\(\sigma = 3.88\) cd/m\(^2\), mean luminance = 24.75 cd/m\(^2\)). These stimuli appeared in the center of four static square boxes included to reduce the intrinsic uncertainty in the task (uncertainty in the exact location of the signal, e.g., Burgess & Ghandeharian, 1984; Pelli, 1985; Eckstein et al., 1997). The boxes were 2 deg in length, and were centered 3.44 deg to the right, left, upward, and downward from a central fixation point. A uniform luminance mask of 38.8 cd/m\(^2\) appeared for 300 ms immediately following the search display. A high-contrast copy of the target was continuously shown at the bottom of the display. The target and distractor locations were randomized on each trial, and the observer indicated his or her choice of the target’s location for that trial by using a computer mouse.

In each experiment, there were three different types of searches based on the two feature dimensions of orientation and spatial frequency. In the single-feature searches, the target differed from the distractors along a single feature dimension, with one condition for each of the two feature dimensions. In the 2-feature search, the target differed from the distractors along both feature dimensions. The distractors were always vertically oriented 2 cycles/deg (cpd) Gabor patches with a 1-octave bandwidth, full-width half-height. The targets in Experiment 1 were relatively close to the distractors in spatial frequency and orientation, while the targets in Experiment 2 were relatively distant from the distractors in spatial frequency and orientation (summarized in Appendix C).
Table 1. Stimulus Parameters for the Target Gabor Patches

**Experiment 1 (close values)**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Spatial frequency of target</th>
<th>Orientation of target</th>
<th>Octave bandwidth</th>
<th>Contrast</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial frequency</td>
<td>2.5 cpd</td>
<td>Vertical</td>
<td>0.789</td>
<td>0.117</td>
</tr>
<tr>
<td>Orientation</td>
<td>2.0 cpd</td>
<td>13/15 deg from vertical(^a)</td>
<td>1.00</td>
<td>0.117</td>
</tr>
<tr>
<td>2-Feature</td>
<td>2.5 cpd</td>
<td>13/15 deg from vertical(^a)</td>
<td>0.789</td>
<td>0.117</td>
</tr>
</tbody>
</table>

**Experiment 2 (orthogonal values)**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Spatial frequency of target</th>
<th>Orientation of target</th>
<th>Octave bandwidth</th>
<th>Contrast</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial frequency</td>
<td>5.0 cpd</td>
<td>Vertical</td>
<td>0.387</td>
<td>0.0664</td>
</tr>
<tr>
<td>Orientation</td>
<td>2.0 cpd</td>
<td>Horizontal</td>
<td>1.00</td>
<td>0.0664</td>
</tr>
<tr>
<td>2-Feature</td>
<td>5.0 cpd</td>
<td>Horizontal</td>
<td>0.387</td>
<td>0.0664</td>
</tr>
</tbody>
</table>

Distractors were always 2.0 cpd vertical Gabor patches, 1 octave bandwidth, full-width half-height. Contrast = Michelson Contrast = (maximum luminance - minimum luminance)/(maximum luminance + minimum luminance). \(^a\)Due to differences between observers, two observers performed the orientation and 2-feature tasks with 13-deg targets, and one observer performed the orientation and 2-feature tasks with 15-deg targets. See the text for details.

In Experiment 2, all the targets were chosen to be statistically orthogonal to the distractors. As discussed later, this property of orthogonality had the consequence that all the conditions in Experiment 2 had the same stimulus information content (same ideal observer performance), regardless of the number of available features. The octave bandwidths of the targets were adjusted to have the same energy and spatial extent as the distractors. An example of one search display from a spatial frequency trial from Experiment 2 is shown in Figure 1. The range of spatial frequencies (2 to 5 cpd) was chosen to be near the peak of the contrast sensitivity function (CSF) to reduce differences in detectability between the stimuli. Also, the CSF for stimuli in image noise has been shown to be flatter than the CSF without image noise (Rovamo, Frassila, & Nasanen, 1992).

For each experiment, an observer performed in 800 trials of each type of search (spatial frequency, orientation, and 2-feature), broken into 8 sessions of 100 trials with the same type of search. Sessions were grouped into blocks of three, with one session of each type of search, and the order of the sessions were randomized within these blocks. Stimuli were presented on a monochrome monitor with a viewing size of 32.51 x 24.38 cm and a resolution of 1,024 x 768 pixels (Image Systems Corp., Minnetonka, MN), sitting 50 cm from the observer. At this distance, each pixel subtended 0.034 deg of visual angle. Luminance calibrations were performed with software and equipment from Dome Imaging Systems, Inc. (Luminance Calibration System, Waltham, MA).

The percentage of correct trials (percent correct, or PC) was obtained for each session, and was transformed to an index of detectability (\(d'\)) using the standard M-AFC transformation from signal detection theory (Green & Swets, 1974, see Appendix A). This index (\(d'\)) is the normalized distance between two Gaussian distributions describing the observer’s response to the target and to the distractor over a large number of trials, and typically varies in a 4-AFC task from 0 (chance performance) to about 4 (nearly perfect performance). Predictions of performance in the 2-feature task for independent feature and ideal observer models were derived from the human observers’ performance in the single-feature tasks (see Appendix B and Appendix C for details). Analyses of variance were performed at an alpha level of .05 using the statistical package GANOVA (Woodward, Bonett, & Brecht, 1990).

Four observers with normal uncorrected or corrected visual acuity participated in the experiments. Three observers participated in both experiments, two initially naïve female observers (K.F., aged 23 years, and A.P., aged 17 years), and the first author (S.S., male, aged 37 years). A fourth naïve male observer was added to the second experiment, (D.V., aged 29 years).

### Experiment 1

As summarized in Table 1, the differences in spatial frequency and orientation between the targets and the distractors in Experiment 1 were relatively small. The spatial frequency of the targets in Experiment 1 for the spatial frequency and 2-feature searches were 2.5 cpd, giving a difference of 0.5 cpd between the targets and the distractors (2.0 cpd). The orientation of the targets in the orientation and 2-feature searches were 15 deg from vertical for K.F., and 13 deg from vertical for S.S. and A.P. The orientation differences were selected separately for each observer to give relatively equivalent levels of performance in the spatial frequency and orientation tasks, and variation in performance across observers led to the use of the different orientations for the targets.

As shown later, for the targets and distractors in Experiment 1, the two independent feature models and the ideal observer assessment of stimulus information predict a comparable increase in performance in the 2-feature task. Thus, for these stimuli, a result of increasing with increasing number of available features cannot be interpreted strictly as the summation across independent features.
Results for Experiment 1

Figure 6. Empirical $d'$s for Experiment 1, by observer. Error bars indicate standard errors of the mean.

Figure 6 gives performance expressed as $d'$ for each observer in Experiment 1. A clear effect can be seen, with $d'$ for the 2-feature search significantly larger than those for the single-feature searches across the observers ($d'_2/$ vs. $d'_{sf}$, F(1,21)=43.15, MSE=0.073, p < .0001; $d'_2/$ vs. $d'_{so}$, F(1,21)=44.21, MSE=0.058, p < .0001). The $d'$ for the spatial frequency and orientation searches were nearly equal, as expected from the separate adjustments of the target orientation for each observer.

Figure 7 gives the ratios of $d'_2/d'_{sf}$ on the left, and $d'_2/d'_{so}$ on the right. Also included are the predictions for the ideal observer and the two independent feature models. First, it should be noted that all three models predict similar ratios that are greater than one, with the probability summation model predicting a slightly smaller ratio than the other two models. Second, the ratios for the three observers also were significantly greater than one ($d'_2/d'_{sf}$, t(2) = 8.39, standard error = 0.038, p = .0139; $d'_2/d'_{so}$, t(2) = 12.11, standard error = 0.022, p = .0067), reflecting the improvement in performance for the $d'$s in the 2-feature search. Third, the empirical ratios tended to fall between the predictions of the probability summation model on the low end, and both the ideal observer and the linear summation models on the high end. Across observers, the empirical ratios were significantly smaller than the linear summation predictions ($d'_2/d'_{sf}$, F(1,21)=10.09, MSE=0.021, p = .0457; $d'_2/d'_{so}$, F(1,21)=12.22, MSE=0.022, p = .022). For both probability summation and the ideal observer, the differences from the empirical ratios across observers approached but did not quite achieve significant levels (probability summation: $d'_2/d'_{sf}$, F(1,21)=4.062, MSE=0.022, p = .0568; $d'_2/d'_{so}$, F(1,21)=2.267, MSE=0.021, p = .1470), and ideal observer: $d'_2/d'_{sf}$, t(2) = 3.848, p = .0614; $d'_2/d'_{so}$, t(2) = 3.664, p = .0681).
Table 2. Absolute Ideal Observer Predictions, $d'$, Experiment 1 (close values)

<table>
<thead>
<tr>
<th></th>
<th>Single-feature</th>
<th>Orientation</th>
<th>2-Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.P. and S.S.</td>
<td>4.711</td>
<td>4.232</td>
<td>6.271</td>
</tr>
<tr>
<td>K.F.</td>
<td>4.711</td>
<td>4.793</td>
<td>6.666</td>
</tr>
</tbody>
</table>

Absolute performance of the ideal observer ($d'_{\text{ideal}}$) for the Experiment 1 in the three conditions can be found in Table 2, calculated as described in Appendix C. Observer K.F. had slightly higher $d'_{\text{ideal}}$ values for the orientation and 2-feature task, corresponding to the slightly larger orientation differences for her (15 deg), compared to A.P. and S.S. (13 deg). Also note that $d'_{2\text{f,ideal}}$ is about 1.4 times greater than the single-feature searches (spatial frequency and orientation), corresponding to the predictions for the ratios of $d'_{2\text{f}}/d'_o$ and $d'_{2\text{f}}/d'_o$ by the ideal observer.

The absolute performance of the human observers may be compared to the ideal observer; typically efficiency $(d'_{\text{human}}/d'_{\text{ideal}})^2$ is used to express this comparison (Barlow, 1978; Burgess et al., 1981). Figure 8 depicts the absolute efficiencies of the human observers in Experiment 1. These efficiencies ranged from about 0.12 to 0.22, which coincides to the efficiencies found in other studies of simple detection and discrimination (e.g., Barlow, 1978; Burgess et al., 1981; Burgess & Ghandeharian, 1984; Eckstein et al., 1997). Observers varied slightly in their absolute performance, with A.P. being more efficient than the other observers (A.P. vs. other observers, F(1,21)=64.71, p <.0001). Finally, observers were more efficient in the orientation tasks, relative to the spatial frequency (Experiment 1: F(1,21)=7.434, p =.0126) and 2-feature tasks (Experiment 1: F(1,21)=18.83, p =.0003).

Experiment 2

In Experiment 1, all the models predicted ratios of $d'_{2\text{f}}/d'_o$ and $d'_{2\text{f}}/d'_o$ greater than one, and the observers’ results were similar to these predictions. As seen from these results, it can be difficult to distinguish the effects of additional stimulus information from the effects of summation across independent features. In the second experiment, a stimulus set was constructed so that the stimulus information was constant across conditions, and therefore predicted equal performance across all three types of searches, regardless of the number of available features. As in the first experiment, the independent feature models predicted an increase of performance in the 2-feature search. Therefore, for the stimuli in Experiment 2, the summation models predict ratios of $d'_{2\text{f}}/d'_o$ and $d'_{2\text{f}}/d'_o$ greater than one, whereas the stimulus information predicts ratios equal to one.

As shown in Table 1, the differences in spatial frequency and orientation between the targets and distractors in Experiment 2 were larger than those in Experiment 1. The spatial frequency of the targets in the spatial frequency and 2-feature tasks was 5 cpd, giving a difference of 3 cpd between the targets and distractors, and the orientation difference between the targets and distractors in the orientation and 2-feature search was 90 deg. Because the differences between the targets and the distractors were larger than the previous experiment, a lower Michelson contrast of 6.64% was used for all the stimuli. The same three observers in Experiment 1 participated in this Experiment, along with an additional observer, a 28-year-old initially naïve male (D.V.).

Results for Experiment 2

Figure 9 gives the results for $d'$ for all the observers in Experiment 2. Most notably for these stimuli, there was no improvement found in the 2-feature search, as was found in the first experiment. Figure 10 gives the ratios of $d'_{2\text{f}}/d'_o$ on the left, and $d'_{2\text{f}}/d'_o$ on the right, with the predicted ratios of the linear summation, probability summation, and ideal observer models. As discussed earlier, the ideal observer predicts ratios about equal to one, and the two summation models predict ratios larger than one. The empirical ratios were clearly closer to the predictions of the ideal observer. The empirical ratios across the four observers were not significantly greater than the predicted ratios of the ideal observer (the ratios of $d'_{2\text{f}}/d'_o$ for S.S. were significantly smaller than the ideal observer, t(7) = -9.844, p = <.0001). Conversely, the empirical ratios across the four observers were significantly smaller than the linear summation predictions ($d'_{2\text{f}}/d'_o$, F(1,28)=283.7, MSE=0.015, p <.0001; $d'_{2\text{f}}/d'_o$, F(1,28)=249.6, MSE=0.013, p <.0001), and the probability summation predictions ($d'_{2\text{f}}/d'_o$, F(1,28)=92.56, MSE=0.014, p <.0001; $d'_{2\text{f}}/d'_o$, F(1,28)=73.26, MSE=0.013, p <.0001). As expected, a
significant experiment-by-type-of-search interaction was found for the \( d' \)’s for the three observers common to both experiments (\( F(2,41)=25.31, \text{MSE}=0.056, p < .0001 \)), indicating the different pattern of results across the two experiments. Also, the empirical ratios were significantly different from each other for the three observers across the two experiments, for both \( d'_{2f}/d'_{sf} \) (\( F(1,42)=22.34, \text{MSE}=0.035, p < .0001 \)) and \( d'_{2f}/d'_{o} \) (\( F(1,42)=41.11, \text{MSE}=0.045, p < .0001 \)).

Absolute performance of the ideal observer (\( d'_{\text{ideal}} \)) for Experiment 2 in the three conditions is listed in Table 3, calculated as described in Appendix C. Note that \( d'_{2f,\text{ideal}} \) for Experiment 2 is nearly equal to the \( d' \)’s for the single-feature searches, corresponding to predictions for the ratios of \( d'_{2f}/d'_{sf} \) and \( d'_{2f}/d'_{o} \) by the ideal observer equal to one for Experiment 2. The absolute efficiencies of the human observers for Experiment 2 are shown in Figure 11. The efficiencies ranged from about 0.05 to 0.125, a lower range than in Experiment 1. In fact, the three observers common to both experiments (A.P., K.F., S.S.) were more efficient in Experiment 1 than in Experiment 2 (\( F(1,42)=173.4, p < .0001 \)), most likely due to the lower contrast necessary for Experiment 2 (with the larger differences in spatial frequency and orientation) leading to increased intrinsic uncertainty (uncertainty about the exact location of the stimulus locations, see Pelli, 1985; Burgess & Ghandeharian, 1984; Eckstein et al., 1997). As in Experiment 1, A.P. was more efficient than the other observers (A.P. vs. other observers, \( F(1,28)=62.23, p < .0001 \)), and observers were more efficient in the orientation tasks, relative to the spatial frequency (\( F(1,28)=17.69, p = .0002 \)) and 2-feature tasks (\( F(1,28)=9.967, p = .0038 \)).

Figure 9. Empirical \( d' \)’s for Experiment 2, by observer. Error bars indicate standard errors of the mean.

Figure 10. Empirical and predicted ratios of \( d' \) for Experiment 2, by observer. The predicted ratios were derived from the linear summation, probability summation, and ideal observer models. The left graph summarizes \( d'_{2f}/d'_{sf} \), and the right graph summarizes \( d'_{2f}/d'_{o} \). Error bars indicate standard errors of the mean.
The results of this study might be interpreted to suggest that there is no need to posit the existence of independent features for spatial frequency and orientation. This is consistent with the coding of spatial frequency and orientation in the primary visual cortex (V1), where it is accepted that neurons in V1 are coded conjointly in spatial frequency and orientation information, with a relatively restricted range of sensitivity across both dimensions. Also, it is believed that such responses in V1 correspond well to human performance in many spatial tasks being described by channels with narrow bands of sensitivity in spatial frequency and orientation (typical estimates are 0.5 to 1.5 octaves and 15-30 deg for the respective bandwidths of spatial frequency and orientation [Graham, 1989]). Beyond the level of V1, however, it has been suggested by a number of authors that spatial frequency and orientation are treated as two independent channels.

For example, Regan has made this suggestion (Regan, 2000) based on his studies on suprathreshold discriminations of simple gratings (Burbeck & Regan, 1983; Regan, 1985; Vincent & Regan, 1995). These studies found that, in general, discriminations along one dimension for two simple gratings are not affected by variations along the other dimension. Chua (1990) came to a similar conclusion using an identification/discrimination task requiring judgments of both spatial frequency and orientation (dual judgments). Chua performed analyses of judgments of 16 stimuli varying across four levels of both spatial frequency and orientation using the confusion matrices along with an information transmission approach (roughly, analyzing the correlation of the errors with the stimulus values and the responses for each dimension). He found that the responses along one dimension were not contingent upon the stimulus values along the other dimension, suggesting that the two features are independent at a decisional level. Finally, Olzak and Thomas (1991, 1992) and Thomas and Olzak (1990, 1996) have proposed two independent classes of mechanisms, one that summates orientation information over a broad range of spatial frequencies (known as cigars for their shapes in polar plots of Fourier space), and conversely, another that summates spatial frequency information over a broad range of orientations (known as donuts, again, for their shapes in polar plots of Fourier space). Using suprathreshold two-dimensional grating patterns of two superimposed Gabor patches of varying spatial frequencies and orientations, they found that judgments for one Gabor patch typically were contingent upon the characteristics of the other Gabor patch in manner consistent with these summation mechanisms. For example, orientation judgments of a Gabor were affected by superimposed orthogonal Gabors, but only those Gabors of a similar spatial frequency. Thomas and Olzak (1996) suggest that the orientation and spatial frequency summation mechanisms under their conditions are obligatory, and that there is limited
direct access to the responses representative of the primary visual cortex.

The present studies did differ in several respects from the previous studies described above. A relatively fast precip and stimulus duration of 200 ms was chosen to negate the effects of eye movements, and the addition of image noise was necessary to generate predictions of less-than-perfect performance for the ideal observer. Also, Olzak and Thomas (1991, 1992) and Thomas and Olzak (1990, 1996) typically have used compound stimuli of two superimposed grating patterns, unlike the simple grating patterns used here. Finally, it is possible that performing a visual search task, as opposed to a simple detection or discrimination, might change the performance of an observer.

Only the predicted ratios of \( d'_{2f}/d'_o \) and \( d'_{sf}/d'_o \) for the ideal observer were compared directly to human performance, and absolute performance of the human observers, as measured by efficiency (see "Results" sections), was several times less than that predicted by the ideal observer. The relevant point is that the ideal observer and the human observers appear to use information from the stimulus similarly. In cases where there is additional featural information, but not additional stimulus information (Experiment 2), the human observers do not show an improvement in the 2-feature task. Therefore, for this task, the human observers’ performance appears to be determined by the stimulus information, and not the featural information. Also, one might consider the consequences of adding internal noise to the ideal observer model, which is equivalent across conditions. By this method, we may more closely match the absolute performance of the human observer with a degraded ideal observer. Regardless of the level of the internal noise added, however, the predicted ratios of \( d'_{2f}/d'_o \) and \( d'_{sf}/d'_o \) for the degraded ideal observer would be the same.

One issue that might be problematic for the modeling of the independent feature models is that the improvement with summation across increasing number features is based on the stochastic independence in the noise of each feature analyzer. In fact, if the noise is perfectly correlated across both feature analyzers, then the summation models predict no improvement in performance from the 1-feature to the 2-feature search task. One possibility is that the stochastic independence of the responses of the orientation and spatial frequency mechanisms is violated due to the fact that the feature mechanisms view the same stimuli on each trial, and therefore the same sample of image noise.

To test this hypothesis, we implemented the mechanisms of Olzak and Thomas (1991, 1992) and Thomas and Olzak (1990, 1996) for independent analyses of orientation (cigar) and spatial frequency (donut) in a linear summation model (see Appendix D). The energy of the image noise, being white in the Fourier domain, is thus distributed across Fourier space. As a result, the correlation of the common image noise in the feature mechanisms is equivalent to the amount of correlation, or overlap, of the two mechanisms themselves in Fourier space. As the cigar and donut mechanisms are specifically designed to sample different parts of this space, the overlap is relatively small (.107 for Experiment 1 and .0347 for Experiment 2). Thus, the performance ratios \( (d'_{2f}/d'_o \) and \( d'_{sf}/d'_o \)) for an observer linearly combining information across the cigar and donuts were estimated to be only slightly less than the predictions for the independent linear combination model. This decrease in ratios across observers was about 0.07 for Experiment 1, and about 0.025 for Experiment 2, smaller compared to the decrease in the predicted ratios when comparing the probability summation predictions against the independent linear summation predictions (about 0.21).

Another aspect to consider is that humans were relatively inefficient compared to the ideal observer in performing the visual search tasks (see "Results" sections). While this inefficiency can be modeled in several ways (as mentioned in the Introduction), it is assumed that at least part is due to internal noise of the response. If independent internal noise is added to the responses of supposed orientation and spatial frequency feature mechanisms, such as cigars and donuts, the internal noise would tend to decorrelate the signals from the mechanisms, and thus further approach the stochastic independence assumption.

---

**Appendix A**

**M-AFC Transformation from Percent Correct (PC) to \( d' \)**

This section describes the standard transformation used to convert percent correct of the human observers to \( d' \) values for an M-AFC procedure (Green & Swets, 1974; MacMillan & Creelman, 1991; Palmer et al., 2000; Eckstein et al., 2000). It is assumed that one response is generated at each of the M locations, which is determined by a univariate Gaussian distribution, with one Gaussian describing the response to the target, and another Gaussian describing the response to the distractor. The distributions have unit variance, with the mean of the distractor distribution equal to zero, and the mean of the target distribution equal to \( d' \). This is a standard set of assumptions for signal detection theory:

- response to the target \( x_t \) = Gaussian \((\mu_t = d', \sigma_t = 1)\),
- response to the distractor \( x_d \) = Gaussian \((\mu_d = 0, \sigma_d = 1)\).

Each location leads to an independent response determined by the appropriate distribution to the stimulus placed at that location; thus, there is one target response \( x_t \) and \( M-1 \) distractor responses \( x_d \). The maximum response across all locations for each trial is chosen as the target location. Therefore, a correct
The response (decision variable) of the ideal observer is the dot product (cross-correlation) of the ideal template with the stimulus at that location (where superscript \( T \) indicates the transpose of the vector).

\[
\lambda = \text{template}^T \cdot \text{stimulus} \tag{B.1}
\]

The ideal template in white noise is the difference between the target and the distractor (Green & Swets, 1974).

\[
\text{template} = t - d \tag{B.2}
\]

The stimuli may be described as the sum of the target and the added image noise, or the distractor and the added image noise:

\[
\text{stimulus}_t = t + n \tag{B.3}
\]

\[
\text{stimulus}_d = d + n . \tag{B.4}
\]

Therefore,

\[
\lambda_t = \text{template}^T \cdot \text{stimulus}_t = (t - d)^T (t + n) \tag{B.5}
\]

\[
\lambda_d = \text{template}^T \cdot \text{stimulus}_d = (t - d)^T (d + n) . \tag{B.6}
\]

The standard deviation of \( \lambda \) is as follows (Green & Swets, 1974).

\[
\sigma_\lambda = \sqrt{\text{template}^T K \text{template}} = \sqrt{(t - d)^T K (t - d)} \tag{B.7}
\]

The image noise was chosen to be white, and thus uncorrelated. Therefore, the covariance matrix is a diagonal matrix: \( K = \sigma^2_{\text{image}} I \), where \( I \) is the identity matrix.

Substituting into the previous equation,

\[
\sigma_\lambda = \sqrt{(t - d)^T K (t - d)} = \sqrt{(t - d)^T \sigma^2_{\text{image}} t (t - d) \tag{B.8}
\]

The following equation describes \( d' \) for the ideal observer:

\[
d'_{\text{ideal}} = \frac{\lambda_t - \lambda_d}{\sigma_\lambda} = \frac{\lambda_t}{\sigma_\lambda} - \frac{\lambda_d}{\sigma_\lambda} \tag{B.9}
\]

The expected value (mean) of \( n \) is zero,

\[
\lambda = \text{template}^T \cdot \text{stimulus} \tag{B.1}
\]

The ideal template in white noise is the difference between the target and the distractor (Green & Swets, 1974).

\[
\text{template} = t - d \tag{B.2}
\]

The stimuli may be described as the sum of the target and the added image noise, or the distractor and the added image noise:

\[
\text{stimulus}_t = t + n \tag{B.3}
\]

\[
\text{stimulus}_d = d + n . \tag{B.4}
\]

Therefore,

\[
\lambda_t = \text{template}^T \cdot \text{stimulus}_t = (t - d)^T (t + n) \tag{B.5}
\]

\[
\lambda_d = \text{template}^T \cdot \text{stimulus}_d = (t - d)^T (d + n) . \tag{B.6}
\]

The standard deviation of \( \lambda \) is as follows (Green & Swets, 1974).

\[
\sigma_\lambda = \sqrt{\text{template}^T K \text{template}} = \sqrt{(t - d)^T K (t - d)} \tag{B.7}
\]

The image noise was chosen to be white, and thus uncorrelated. Therefore, the covariance matrix is a diagonal matrix: \( K = \sigma^2_{\text{image}} I \), where \( I \) is the identity matrix.

Substituting into the previous equation,

\[
\sigma_\lambda = \sqrt{(t - d)^T K (t - d)} = \sqrt{(t - d)^T \sigma^2_{\text{image}} t (t - d) \tag{B.8}
\]

The following equation describes \( d' \) for the ideal observer:

\[
d'_{\text{ideal}} = \frac{\lambda_t - \lambda_d}{\sigma_\lambda} = \frac{\lambda_t}{\sigma_\lambda} - \frac{\lambda_d}{\sigma_\lambda} \tag{B.9}
\]

The expected value (mean) of \( n \) is zero,
\[
\begin{align*}
\dd'_{\text{ideal}} &= \frac{(t - d)^T (t + n) - (d + n)}{\sigma_{\text{image}} \sqrt{(t - d)^T (t - d)}} \\
&= \frac{(t - d)^T (t - d)}{\sigma_{\text{image}} \sqrt{(t - d)^T (t - d)}} \\
&= \frac{\sqrt{(t - d)^T (t - d)}}{\sigma_{\text{image}}} \\
&= \frac{\sqrt{t^T t - 2d^T t + d^T d}}{\sigma_{\text{image}}}.
\end{align*}
\]

This equation was used to predict ideal observer performance (\(d'\)) for all conditions (spatial frequency, orientation, and 2-feature), and the predicted ratios of \(d'_2/d'_d\) and \(d'_{2f}/d'_{1f}\) were calculated from these individual \(d'\)s. Notably, the \(d'\) for the ideal observer is a function of the image noise, the energies of the target and the distractors (\(t, d, d\)), and the correlation between the target and the distractor (the dot product, \(d\)). In Experiment 2, the targets for all conditions were chosen to be effectively orthogonal to the distractors (correlations equal to zero). Thus, \(d'_{\text{ideal}}\) for all conditions in Experiment 2 depend only on the energies of the targets and distractors. The energies were also equalized across conditions, leading to predictions of equal performance across all conditions in Experiment 2.

**Appendix C Independent Feature models**

The following section gives descriptions of the two common models of summation across independent features, and their expressions for \(d'\) in the 2-feature task.

**Linear Summation**

A schematic for the linear summation model can be found in Figure 3. The linear summation model assumes two responses at each location, one for each feature. The responses are described by two Gaussian distributions of unit variance, one for the target, and one for the distractor, using a standard SDT assumption.

Response to the spatial frequency of the target
\[
x_{t,sf} = \text{Gaussian}\ (\mu_{t,sf} = d'_{sf}, \sigma_{t,sf} = 1)
\]

Response to the spatial frequency of the distractor
\[
x_{d,sf} = \text{Gaussian}\ (\mu_{d,sf} = 0, \sigma_{d,sf} = 1)
\]

Response to the orientation of the target
\[
x_{t,o} = \text{Gaussian}\ (\mu_{t,o} = d'_{o}, \sigma_{t,o} = 1)
\]

Response to the orientation of the distractor
\[
x_{d,o} = \text{Gaussian}\ (\mu_{d,o} = 0, \sigma_{d,o} = 1)
\]

The linear summation model uses a weighted linear combination at each location of the responses to each feature as its decision variable (\(x_{\text{linear}}\)). The weights for the linear combination are the \(d'\)s for each task, which are the optimal weightings in this case. On each trial, the maximum value amongst \(x_{\text{linear}}\) and the three \(x_{d-o}\)'s is chosen as the target location, where the weighted linear combination of the target responses is

\[
x_{t-linear} = d'_{sf}x_{t-sf} + d'_{o}x_{t-o}
\]

and the weighted linear combination of distractor responses is

\[
x_{d-linear} = d'_{sf}x_{d-sf} + d'_{o}x_{d-o}
\]

The means for \(x_{t-linear}\) and \(x_{d-linear}\) are defined in terms of the means for the responses to spatial frequency and orientation.

\[
\mu_{t-linear} = d'_{sf}\mu_{t-sf} + d'_{o}\mu_{t-o} = (d'_{sf})^2 + (d'_{o})^2
\]

\[
\mu_{d-linear} = d'_{sf}\mu_{d-sf} + d'_{o}\mu_{d-o} = 0
\]

Assuming that the responses to the two features are independent, and with the standard deviations of the responses to each feature = 1,

\[
\sigma_{\text{linear}} = \sqrt{(d'_{sf}\sigma_{sf})^2 + (d'_{o}\sigma_{o})^2} = \sqrt{d'_{sf}^2 + d'_{o}^2}
\]

The \(d'_{2f}\) for the linear summation model is expressed in terms of the weighted linear combinations, \(x_{t-linear}\) and \(x_{d-linear}\).

\[
d'_{2f, linear} = \frac{\mu_{t-linear} - \mu_{d-linear}}{\sigma_{\text{linear}}}
\]

\[
\therefore d'_{2f, linear} = \frac{d'_{sf}^2 + d'_{o}^2}{\sqrt{d'_{sf}^2 + d'_{o}^2}}
\]

\[
d'_{2f, linear} = \sqrt{d'_{sf}^2 + d'_{o}^2}
\]

Predictions for \(d'_{2f, linear}\) for each observer were found from Equation 6. The MAFC conversion in Appendix A may be used to convert \(d'_{2f, linear}\) to the predicted percent correct in the 2-feature task.

**Probability Summation**

A schematic of the probability summation model can be found in Figure 5, in which the target for a two-feature search is located in the left position. The probability summation model assumes that each location gives two independent responses, one for each feature, and chooses the location with the maximum response across both
features as the target location. The responses are described by unidimensional Gaussian distributions of unit variance, one for the target, and one for the distractor, a standard assumption from signal detection theory.

Response to the spatial frequency of the target

\[ x_{t-sf} = \text{Gaussian} (\mu_{t-sf} = d'_s, \sigma_{t-sf} = 1) \]

Response to the spatial frequency of the distractor

\[ x_{d-sf} = \text{Gaussian} (\mu_{d-sf} = 0, \sigma_{d-sf} = 1) \]

Response to the orientation of the target

\[ x_{t-o} = \text{Gaussian} (\mu_{t-o} = d'_o, \sigma_{t-o} = 1) \]

Response to the orientation of the distractor

\[ x_{d-o} = \text{Gaussian} (\mu_{d-o} = 0, \sigma_{d-o} = 1) \]

Percent correct for the probability summation model in the 2-feature task is \( PC_{2f, \text{prob sum}} \).

Let \( M \) be the total number of locations. Then

\[ PC_{2f, \text{prob sum}} = p(x_{t-sf} \text{ is max}) + p(x_{t-o} \text{ is max}) \] (C.9)

For \( x_{t-o} \) to have the maximum value, all other responses must be less than \( x_{t-sf} \), including the response to orientation for the target (\( x_{t-o} \)), and the response to spatial frequency and orientation for the distractors (\( x_{d-sf}, x_{d-o} \)). Therefore,

\[
p(x_{t-sf} \text{ is max}) = \int_{-\infty}^{\infty} \left( \prod_{i=1}^{M-1} p(x_{d-sf_i} < x) \prod_{i=1}^{M-1} p(x_{d-o_i} < x) \right) dx 
\]

\[
= \prod_{i=1}^{M-1} \int_{-\infty}^{\infty} p(x_{d-sf_i} < x) p(x_{d-o_i} < x) dx . \] (C.10)

Substituting the Gaussian assumptions,

\[ p(x_{t-sf} = x) = g(x - d'_s) \]

\[ p(x_{t-o} < x) = \int_{-\infty}^{x-d'_o} g(y) dy = G(x - d'_o) \]

\[ p(x_{d-sf} < x) = \int_{-\infty}^{x} g(y) dy = G(x) \]

\[ p(x_{d-o} < x) = \int_{-\infty}^{x} g(y) dy = G(x) . \]

Therefore,

\[
p(x_{t-sf} \text{ is max}) = \int_{-\infty}^{\infty} g(x-d'_s) G(x) dx \]

\[
= \int_{-\infty}^{\infty} g(x-d'_s) G(x) dx \] (C.11)

Similarly, for \( x_{t-o} \) to have the maximum value, all other responses must be less than \( x_{t-o} \).

\[
p(x_{t-o} \text{ is max}) = \int_{-\infty}^{\infty} \left( \prod_{i=1}^{M-1} p(x_{d-sf_i} < x) \prod_{i=1}^{M-1} p(x_{d-o_i} < x) \right) dx 
\]

\[
= \prod_{i=1}^{M-1} \int_{-\infty}^{\infty} p(x_{d-sf_i} < x) p(x_{d-o_i} < x) dx . \] (C.12)

Therefore,

\[
PC_{2f, \text{prob sum}} = \int_{-\infty}^{\infty} g(x-d'_o) G(x) dx + \int_{-\infty}^{\infty} g(x-d'_o) G(x) dx \] (C.13)

Equation C.14 was used to find a predicted \( PC_{2f, \text{feature}} \) for the probability summation model from the \( d'_s \) and the \( d'_o \) for each observer.\(^6\) The \( PC_{2f, \text{feature}} \) then was converted to \( d'_{2f} \) for the probability summation model by using the same conversion to \( d' \) described above in Appendix A.

Note that this equation is an exact prediction of probability summation (Eckstein et al., 1996, 2000; Tyler & Chen, 2000), and not an approximation, such as the Quick pooling model, which has been a common approximation used by others (Quick, 1974; Graham & Robson, 1987; Graham, 1989).\(^7\)
Appendix D
Linear Summation Across Cigars and Donuts

A possible violation of independence for the tests of the independent feature models is that, for this task, any proposed set of independent feature analyzers for orientation and spatial frequency views the same sample of image (external) noise at each location. A simulation of a linear combination across specific proposed feature mechanisms was performed to assess the effect of the correlated external noise in the predictions of improvement in performance in the 2-feature task. The specific model of orientation and spatial frequency feature mechanisms used in the simulation were proposed originally by Olzak and Thomas (1991, 1992) and Thomas and Olzak (1990, 1996), and are known as cigars for orientation and donuts for spatial frequency. The cigar mechanism combines information across spatial frequencies for a narrow bandwidth of orientations, and is named for its cigarlike shape in polar Fourier space. The donut mechanism combines information across orientations for a narrow bandwidth of spatial frequencies, and thus has the shape of a donut in polar Fourier space.

The simulated cigar and donut mechanisms were constructed from a linear combination of individual Gabor mechanisms (and thus, analogous to the response of V1 neurons), with specific parameters based on the spatial model of Watson (1983). All the Gabor mechanisms had a spatial frequency bandwidth of an octave (half-height, full-width), and an orientation bandwidth of 38.2 deg. The cigar mechanisms were comprised of Gabors with peak spatial frequencies in octave steps from 0.5 cpd to 16 cpd (0.5, 1, 2, 4, 8, and 16 cpd), with two phases differing by 90 deg (quadrature), and with peak sensitivities at the target orientations (15 deg for Experiment 1, 90 deg for Experiment 2). The donut mechanisms were comprised of Gabors at orientations differing by 30 deg (i.e., 0, 30, 60, 90, 120, and 150 deg), with two phases differing by 90 deg, and with peak sensitivities at the target spatial frequencies (2.5 cpd for Experiment 1, 5 cpd for Experiment 2). We assumed that performance in the single-feature tasks was mediated by the output of a single cigar or donut. Thus, for Experiment 1, orientation performance was determined by a cigar tuned to 15 deg, and spatial frequency performance was determined by a donut tuned to 2.5 cpd. For Experiment 2, orientation performance was determined by a cigar tuned to 90 deg, and spatial frequency performance was determined by a donut tuned to 5 cpd. Performance in the 2-feature task was determined by the linear combination of the cigar and donut mechanisms, weighted by the sensitivity of each mechanism.

In Gaussian white image noise, the correlation of the external noise entering the feature mechanisms is equivalent to the correlation of the mechanisms themselves. Intuitively, the energy of the white image noise is distributed across Fourier space, and thus the correlation of the external noise entering the feature mechanisms depends upon the overlap of these mechanisms in Fourier space. These correlations (ρ) between the cigar and donut mechanisms were relatively low, .107 for Experiment 1, and .0347 for Experiment 2. This is expected, as the cigar and donut mechanisms are specifically designed to sample different parts of Fourier space. As a result of the low correlations, the predictions for improvement in the 2-feature task for the cigar and donut mechanisms were similar (slightly less than) to the predictions for the independent linear summation model (see below).

For predictions of the linear summation of the outputs of the cigar and donut mechanisms, a slight modification of the independent linear summation model is necessary. As shown in Appendix C, the equation
describing $d'$ for the linear combination of the feature mechanisms is as follows (C.6, C.7).

$$d'_{2f, \text{linear}} = \frac{\mu_{\text{linear}} - \mu_{\text{distractor}}}{\sigma_{\text{linear}}} = \frac{d^2_{sf} + d^2_o}{\sigma_{\text{linear}}} \quad \text{(D.1)}$$

In this case of the cigars and donuts, the standard deviation of the combined response includes the correlation (covariance = $2\rho(d'_{sf} \sigma_{sf})(d'_o \sigma_o)$) between the single feature responses:

$$\sigma_{\text{linear}} = \sqrt{(d'_{sf} \sigma_{sf})^2 + (d'_o \sigma_o)^2 + 2\rho(d'_{sf} \sigma_{sf})(d'_o \sigma_o)}$$

$$= \sqrt{d^2_{sf} + d^2_o + 2\rho(d'_{sf})(d'_o)} \quad \text{(D.2)}$$

So that

$$d'_{2f, \text{linear}} = \frac{d^2_{sf} + d^2_o}{\sigma_{\text{linear}}} = \frac{d^2_{sf} + d^2_o}{\sqrt{d^2_{sf} + d^2_o + 2\rho(d'_{sf})(d'_o)}} \quad \text{(D.3)}$$

Table 4 gives the predicted ratios of $d'_{o}/d'_{d}$ and $d'_{2f}/d'_{o}$ for the linear summation across cigars and donuts, for the independent linear summation model, and the difference between the predicted ratios. As with the other summation models, the predicted ratios for summation across cigars and donuts were larger than one. Also, the differences in predicted ratios between the independent cigar/donut linear summation models were small, about 0.07 for Experiment 1 and about 0.025 for Experiment 2. These differences were much less than the difference in predicted ratios for independent linear summation and the probability summation (about 0.21).

**Acknowledgments**

The authors wish to thank James Thomas for his observations and insight, and Kristine Fazio, Ana Prokhorova, and Darko Vodopich for their participation as observers. Funded by National Institutes of Health Grant 53455 and National Aeronautics and Space Administration NAG-1157. Commercial relationships: none.

**Footnotes**

1. Note that others (e.g., Geisler & Davila, 1985) have talked about an ideal observer analysis in terms of an ideal decision after suboptimal processing of the image reflecting the human visual system. In this latter treatment of the ideal observer, the model performance reflects both stimulus information and constraints in the human visual system.

2. A common variant of this model is linear summation with equal weighting across the two features (unweighted). The stimulus parameters were chosen to give approximately equal performance in the orientation and spatial frequency searches, and therefore the weighted and unweighted linear summation models gave nearly equivalent predictions for the 2-feature search.

3. As described in Appendix B, the stimulus information in the present task as assessed by the ideal observer depends upon (1) the energies of the target and the distractors, (2) the image noise, and (3) the correlation between the target and the distractors. All the targets in Experiment 2 were chosen to equate these three stimulus parameters across all conditions, including, most importantly, the correlations of the targets with the distractors. This ensures the same stimulus information for each task. Specifically, all the targets were orthogonal to the distractors (correlations = 0).

4. However, the analyses of Shaw (1980, 1982) and Palmer (Palmer et al., 1993; Palmer, 1995; Palmer, Verghese, & Pavel, 2000) of set-size effects in visual search suggest that performance in visual search for many tasks can be directly predicted from simple detection and discrimination results.

5. The internal noise is added as a scalar to the decision variable of the ideal observer (the dot product of the stimulus with the ideal template, with the ideal template defined as the difference between a template matching the target and a template matching distractor) (see Figure 2 and Appendix B).

6. In the highly unlikely case of a tie, the model chooses the distractor location over the target location. It might be more appropriate for the model to choose randomly in this case; however, this has little effect in the predictions.

7. Best fits of the predictions of the exact probability summation model to the Quick pooling formula (Graham, 1989; Quick, 1974) ($d'_{2f} = (d'_{o}^{\frac{1}{k}} + d'_{o}^{\frac{1}{k}})^{\frac{1}{2}}$) for the observed data in this study found summation exponents ($k$) from 2.96 to 3.72. Probability summation has been inferred from exponents ranging from 3 to 5 (Graham, 1989), congruent with the predicted exponents for the exact probability summation model.

**References**


