

Scale dependence and channel switching in letter identification

Ipek Oruç

Department of Ophthalmology and Visual Sciences,
University of British Columbia, Vancouver, BC, Canada



Michael S. Landy

Department of Psychology and Center for Neural Science,
New York University, New York, NY, USA



Letters are broadband visual stimuli with information useful for discrimination over a wide range of spatial frequencies. Yet, recent evidence suggests that observers use only a single, fixed spatial-frequency channel to identify letters and that the scale of that channel, in units of letter size, is determined by the size of the letter (scale dependence). We report two letter-identification experiments using critical-band masking. With sufficiently high-amplitude, low- or high-pass masking noise, observers switched to a different range of spatial frequencies for the task. Thus, letter channels are not fixed for a given letter size. When an additional white-noise masker was added to the stimulus to flatten the contrast-sensitivity function, the letter channel used by the observer still depended on letter size, further supporting the hypothesis that letter identification is scale dependent.

Keywords: letter identification, letter channels, critical-band masking, scale invariance, channel switching, equivalent input noise, ideal observer

Citation: Oruç, I., & Landy, M. S. (2009). Scale dependence and channel switching in letter identification. *Journal of Vision*, 9(9):4, 1–19, <http://journalofvision.org/9/9/4/>, doi:10.1167/9.9.4.

Introduction

How are objects recognized? Whereas much is known about the initial coding of visual images, there is comparatively little understood about how the results of early visual coding are interpreted to recognize a face or identify a letter. The identification of letters is a useful test case for studying object recognition. People have a huge amount of practice at this task. The number of different stimulus categories (letters) is small and yet the physical forms vary widely (in shape, font, size, etc.) while continuing to be recognizable.

Recent studies suggest that letters are typically identified using a single spatial-frequency channel (e.g., Solomon & Pelli, 1994). The preferred spatial frequency of that channel depends only on the size and the complexity of the type font (Majaj, Pelli, Kurshan, & Palomares, 2002). In this paper, we describe two experiments intended to determine the degree of flexibility observers have in the choice of the spatial frequency channel used for letter identification. We begin by reviewing previous studies of the properties of spatial channels used in letter identification, determined primarily by noise masking. Then, we discuss implications of the shape of the contrast-sensitivity function letter-identification mechanisms, which will serve to motivate the experiments that follow.

Letter channels

Letters are broadband visual stimuli and there is information available in the letter spectra at a broad range of scales that is useful for letter identification (Figure 1). When observers identify band-pass-filtered letters masked by noise, human efficiency at letter identification is scale invariant and depends in a complex way on the parameters of the noise masker (static or dynamic, broad band or spatially filtered) with a peak efficiency at 1–3 cycles/letter (e.g., Gold, Bennett, & Sekuler, 1999; Parish & Sperling, 1991). Thus, humans can identify letters using information from a wide range of spatial frequency.

Yet, when subjects identify unfiltered (i.e., broadband) letters masked by noise, only a relatively narrow range of spatial frequencies is used for the task. The primary evidence for this comes from the use of the critical-band-masking paradigm in a letter-identification task (Solomon & Pelli, 1994), a technique we employ in the experiments described below. In Solomon and Pelli's experiments, observers identified unfiltered letters masked by either low- or high-pass noise. For a fixed noise power spectral density (hereinafter referred to as "noise level") one finds that threshold contrast for letter identification rises as noise bandwidth increases, i.e., as the cutoff frequency increases (for low-pass noise) or decreases (for high-pass noise) (Figure 2). The derivative of this curve with respect to noise cutoff indicates the incremental effect on threshold

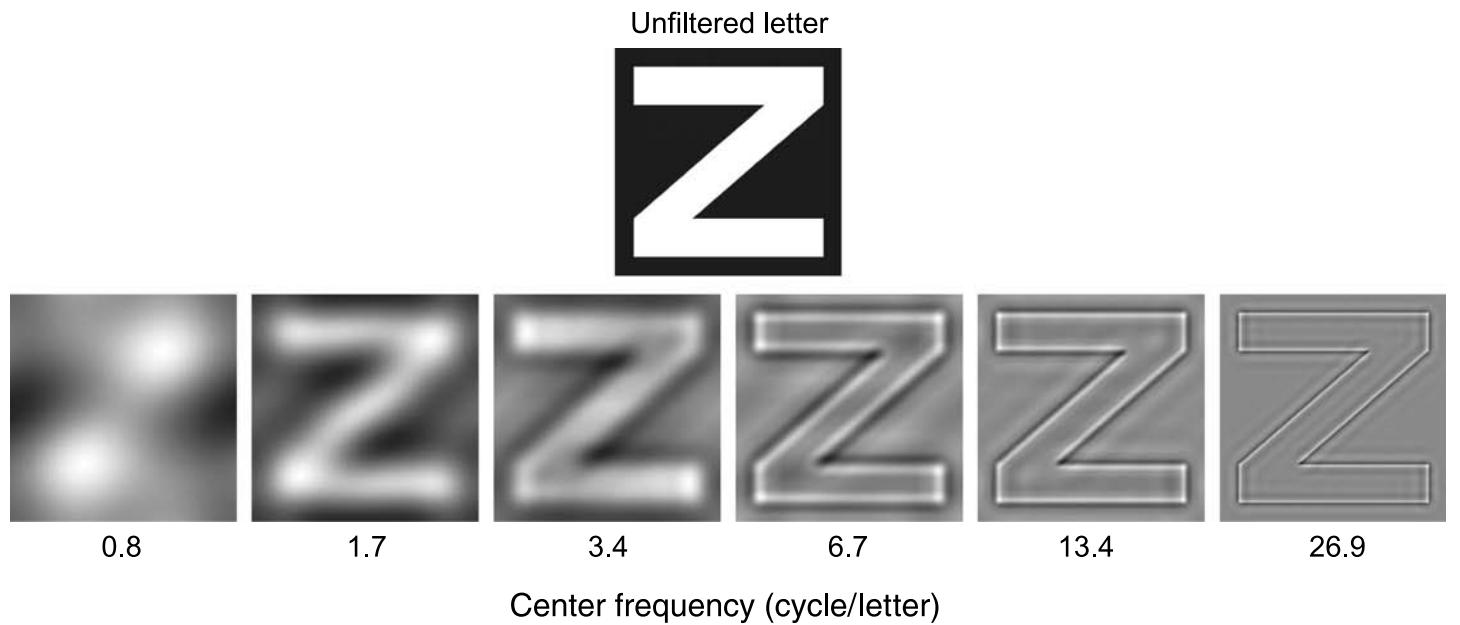


Figure 1. Band-pass-filtered letters. (Top) Unfiltered letter Z. (Bottom) Letter Z band-pass filtered using 2-octave-wide filters at six center frequencies. The identity of the letter is evident in band-pass-filtered versions of the broadband image for a wide range of center frequencies.

of noise at that spatial frequency and thus yields an estimate of the weighting of each spatial frequency used by the observer (Patterson, 1974) for 1-dimensional (e.g., auditory) stimuli. For 2-dimensional stimuli such as static visual patterns, an additional scaling by the frequency is required to yield the average power gain across orientations (Solomon & Pelli, 1994). The results of this experiment indicate that subjects use a relatively narrow band of spatial frequencies roughly 1.5 octaves wide to identify letters, despite the fact that useful information is available over a far broader range. Majaj et al. (2002) have replicated this result and extended it to a wide variety of fonts and sizes.

When the critical-band-masking technique is used with narrow-band target stimuli (i.e., sine-wave gratings), the channel used by the subject appears to depend on the characteristics of the masking noise. For example, when observers are asked to detect a sine-wave grating of frequency f , they will typically use a channel centered on f to perform the task because that is the channel that is most sensitive to the target stimulus. However, when confronted with, e.g., a low-pass noise masker, they will shift to a channel with a higher peak spatial frequency (Figure 3A illustrates this for the task of letter discrimination rather than grating detection). This *off-frequency looking* implies that the subject uses a channel less sensitive to the test grating but also less responsive to the masking noise, resulting in a higher signal-to-noise ratio (Pelli, 1981; Perkins & Landy, 1991). With a low-pass masker, observers use a higher frequency channel, and with high-pass noise they use a lower frequency channel.

An interesting aspect of the results of critical-band-masking experiments using broadband-letter stimuli (Majaj et al., 2002; Solomon & Pelli, 1994) is that the channel used for letter identification appears to be fixed. The channels inferred using low- and high-pass maskers were identical or nearly so (Majaj et al., 2002). That is, nearly the same channel was used to identify letters independent of the characteristics of the noise masker. Observers did not switch channels to avoid masking noise even though, unlike the case of detecting a sine-wave grating, there was useful information in the letter stimuli at these neighboring spatial frequencies. (Note that for letter stimuli we refer to this strategy as channel switching, instead of off-frequency looking, because letters are typically broad band.) Majaj et al. (2002) took this as evidence for a mechanism by which the channel to be used was selected based on the signals to be discriminated, independent of the noise condition. Oruç, Landy, and Pelli (2006) found a similar inability to switch channels for identification of texture-defined letters. In the experiments described below, we demonstrate that letter-identification channels are not fixed. We find that the degree of channel switching increases with masking-noise level.

If an observer uses a fixed channel, then thresholds should display additivity of masking. To explain this, suppose one first measured threshold energy E_0 of a target with no masker, along with threshold energies E_{low} and E_{high} in the presence of low- and high-pass noise maskers sharing the same cutoff frequency f , and E_{all} in the presence of all-pass (white) noise. If the observer used a fixed channel, and threshold depended only on the signal-

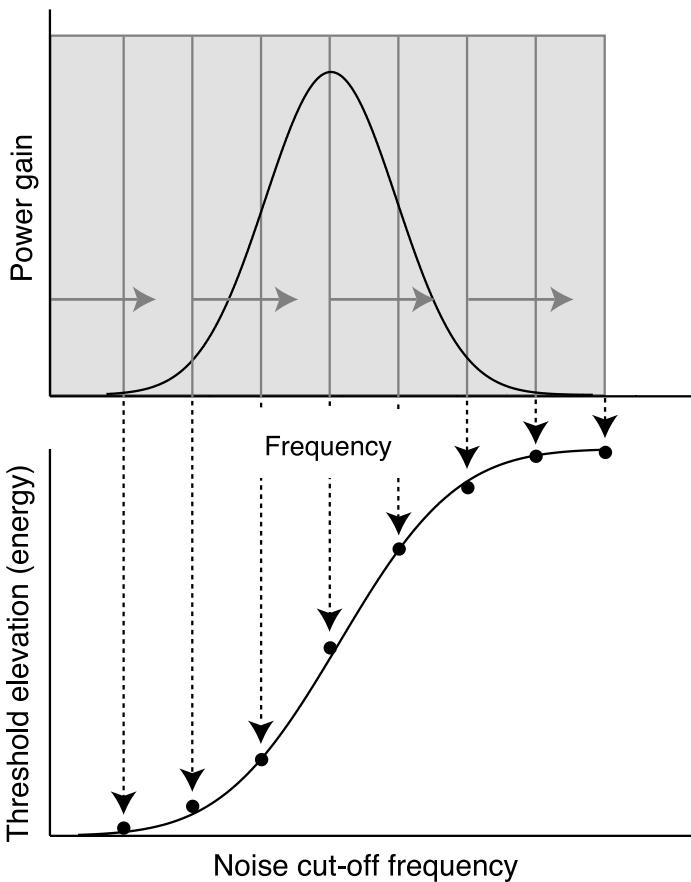


Figure 2. Critical-band masking. (Top panel) The solid curve indicates the sensitivity profile of a channel used to identify a stimulus, e.g., a letter. Low-pass noise with various cutoff frequencies is added to the stimulus. (Bottom panel) Threshold elevation as a function of noise cutoff frequency (relative to threshold without added noise). Noise is effective in elevating threshold to the extent that the channel is sensitive to the noise. For 1-dimensional stimuli, the derivative of the threshold elevation curve provides an estimate of the channel's power gain.

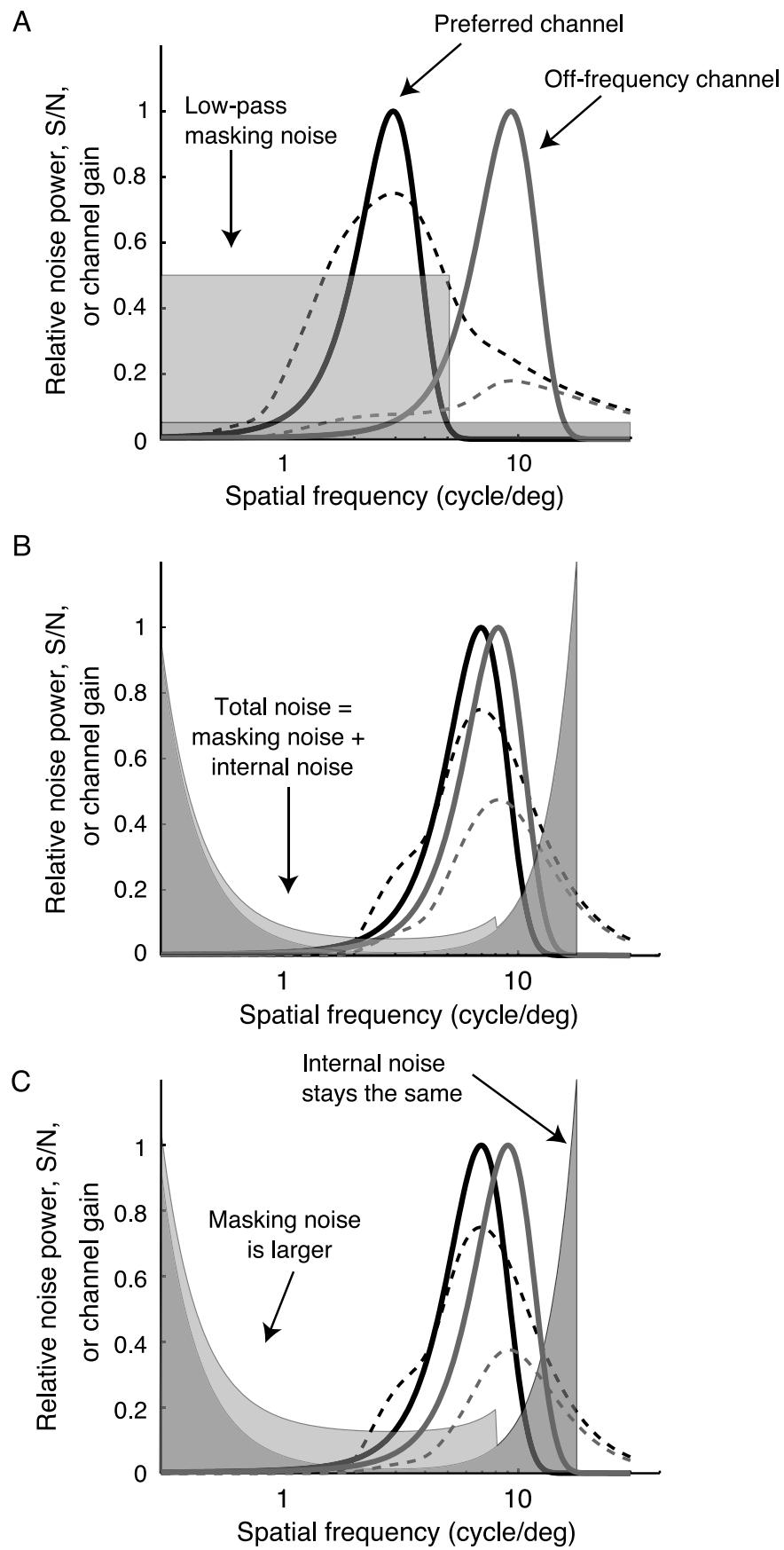
to-noise ratio passed by that channel, then threshold elevations (e.g., $E_{\text{low}}^+ = E_{\text{low}} - E_0$) should sum, i.e., $E_{\text{all}}^+ = E_{\text{low}}^+ + E_{\text{high}}^+$. On the other hand, with off-frequency looking, observers should be able to reduce threshold for the low- and high-pass maskers, so that $E_{\text{all}}^+ > E_{\text{low}}^+ + E_{\text{high}}^+$. Majaj et al. (2002) reviewed the data of their own study as well as that of Pelli (1981) and concluded that the excess masking (at most 0.2 log units on average, or about 33% higher contrast energy required in white noise than predicted) was negligible for both letters and sine-wave gratings to have any practical importance. Observers were unable to switch channels to any useful degree. Note that this argument did not take into account the data of Perkins and Landy (1991), who found excess masking an order of magnitude larger than this for the detection of a sine-wave grating in the presence of two band-pass noise maskers with frequency ranges above and below that of the target grating. In the experiments below, we will attempt to

resolve the apparent contradiction between these results. In our experiments described below, we find that excess masking increases with masking-noise level.

Another important aspect of letter identification is scale invariance. That is, is the same mechanism used for object recognition independent of object size? Many studies indicate that human observers are scale invariant in recognizing objects (Biederman & Cooper, 1992; Furmanski & Engel, 2000; Landy & Bergen, 1991; also see Wiskott, 2006). Parish and Sperling (1991) found that letter-identification performance for band-pass-filtered letters in band-pass-filtered noise was unaffected by changes in viewing distance across a 30:1 range. Pelli, Burns, Farell, and Moore-Page (2006) found that efficiency fell slowly with increasing letter size. Similarly, Legge, Pelli, Rubin, and Schleske (1985) found that reading rates were largely unaffected over a nearly 50-fold range of text sizes, with reduced reading speeds only for extremely large or small text. However, Chung, Legge, and Tjan (2002) and Majaj et al. (2002) found scale dependence: letters were identified differently at different scales. The peak frequency of the channel used to identify letters, when described in units relative to the letter size (cycle/letter), was lower for smaller letters. In other words, small letters were identified by their coarse strokes while large letters were identified by their edges and fine details. Below, we test whether scale-invariance is due to the shape of the contrast-sensitivity function. To understand this idea, we must first describe the concept of equivalent input noise and its implications for letter-identification mechanisms.

Equivalent input noise and its implications

The ability of human observers to detect sine-wave gratings depends on spatial frequency. Ahumada and Watson (1985) suggested that the contrast sensitivity function (CSF) can be modeled as an *equivalent input noise* (hereinafter referred to as “equivalent noise” and symbolically as N_{eq}). That is, insensitivity to a given spatial frequency can be treated as if the observer were, in fact, an ideal discriminator hampered both by noise in the visual display plus equivalent noise. This equivalent noise is high for those spatial frequencies at which the observer's sensitivity is low, and low where the sensitivity is high. The intensity of the equivalent noise at each spatial frequency is set so that an ideal observer, when presented with sine-wave gratings plus the equivalent noise, would reproduce the observer's CSF. As it is usually measured, the CSF is band pass. The high-frequency cutoff is due to low-pass filtering by the eye's optics, while the low-frequency attenuation is the result of neural processing (e.g., Banks, Geisler, & Bennett, 1987). Thus, equivalent noise is highest at low and high spatial frequencies and has a minimum where sensitivity is highest (approximately 4 cycles/deg).



In considering the results of studies of letter discrimination using masking noise, it is important to consider the CSF. Effectively, in these studies observers discriminate letters in the face of masking noise in the stimulus itself (the noise masker in the display) plus the equivalent noise. Thus, it may not be in the observer's best interest to switch channels. For example, in switching channels to one centered on a higher spatial frequency to avoid low-pass noise, the observer will reduce the amount of masker noise passed by that channel but may also increase the equivalent noise passed by the channel (Figure 3B). This may explain why Majaj et al. (2002) did not observe more substantial channel switching. However, suppose the low-pass noise masker level was increased substantially, so that it dominated the equivalent noise in neighboring parts of the spectrum (Figure 3C). In this case, it would be beneficial to switch channels to higher spatial frequencies to improve signal-to-noise ratio.

Indeed, Solomon (2000) observed this phenomenon for the task of detecting a Gabor masked by low- or high-pass-filtered noise. As low-pass noise level was increased, observers switched to channels with higher center frequencies. If the observer had used a fixed channel, then one would predict that threshold would be proportional to noise level when noise level was high (so that stimulus noise dominated equivalent noise). Instead, the log-log slope of threshold elevation was less than 1 consistent with channel switching, especially for low-pass noise maskers with cutoff frequencies close to the target frequency. **Experiment 1** tests whether observers will also channel switch when the task is letter identification if the masker level is sufficiently high.

The shape of the CSF (and hence of the equivalent noise) may also be the reason for the finding of scale dependence (Chung et al., 2002; Oruç et al., 2006). If the size of a letter is reduced, its spatial frequency content is shifted to higher spatial frequencies. Scale invariance implies using a “preferred” channel (stated in units of cycle/letter) independent of letter size. Reducing the size of a letter shifts this “preferred” channel to higher spatial frequencies (now stated in retinal units of cycle/deg). In this part of the spectrum, the equivalent noise is increasing with spatial frequency. For small letters, observers would improve signal-to-noise ratio if they switched to a lower

Figure 3. Increased noise power spectral density results in larger channel switching. (A) The black dashed curve indicates the signal-to-noise ratio for letter discrimination using octave-wide channels of varying center frequency (the abscissa) in the face of flat internal noise (dark shaded area). Solid line: Tuning curve of the channel with the highest S/N. With added low-pass masking noise (light shaded area and gray curves), S/N is reduced and the optimal channel moves to a higher spatial frequency (“channel switching”). (B) Channel switching is reduced with small letters and u-shaped equivalent input noise (based on a band-pass CSF; dark shaded area). (C) Channel switching is more prominent for a higher external noise level.

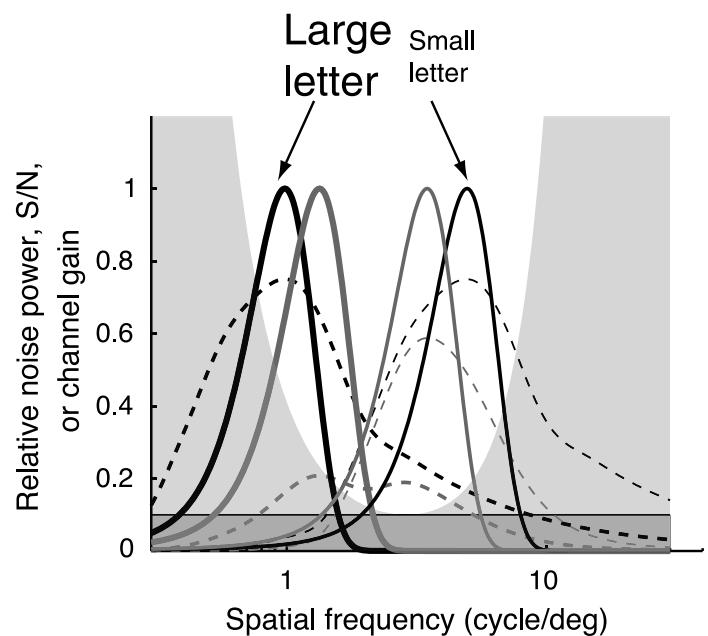


Figure 4. CSF-based account of scale dependence. Conventions same as Figure 3. For flat internal noise (dark shaded area, black curves), reducing letter size by a factor of 5 increases the peak frequency of the preferred channel in retinal units (cycle/deg), so that peak channel frequency remains fixed in units of cycle/letter, i.e., letter identification is scale invariant. However, in the face of equivalent input noise (light shaded area, gray curves) preferred channels are shifted toward moderate frequencies where equivalent noise is reduced, resulting in scale dependence.

frequency channel (Figure 4). Similarly, for large letters, the letter spectrum is shifted to a region in which equivalent noise decreases with increasing spatial frequency, implying that observers should shift to a higher frequency channel (Figure 4). This is qualitatively in agreement with the scale dependence found by Majaj et al. (2002); their observers used a lower frequency channel (relative to the letter, i.e., in units of cycle/letter) for small letters as compared to large ones.

To demonstrate this explanation, Chung et al. (2002) simulated a subideal observer that included a front-end layer to take into account the human CSF. This model, called the CSF-ideal observer, discriminated band-pass letters embedded in white noise by selecting the alternative with maximum likelihood. Similarly, they measured human contrast thresholds for the discrimination of band-pass-filtered letters. The CSF-ideal observer produced a similar pattern of results to the human observers in the sense that it produced a scale-dependent pattern of behavior: the CSF-ideal's peak sensitivity to band-pass-filtered letters shifted to higher frequencies for larger letters, and lower frequencies for smaller letters.

Chung et al. (2002) and Majaj et al. (2002) reach essentially opposite conclusions using different experimental designs. Majaj et al. used critical-band masking with unfiltered as well as band-pass-filtered letters and concluded,

given the negligible excess masking, that observers used a fixed channel for letter identification independent of the spatial frequency cutoff of the masking noise. When unfiltered letter size was varied, the estimated channel was not fixed (relative to letter size), leading them to conclude that the observer selected the channel used for letter identification based on letter size. Chung et al. used band-pass-filtered letters and no external noise. In comparing their experimental results with predictions of the CSF-ideal observer, they concluded that observers use all spatial frequencies as dictated by the characteristics of equivalent noise, i.e., weighted by the signal-to-noise ratio at each spatial frequency, and that there was no need to posit active selection of narrow-band channels internal to the observer to explain their data. In other words, Chung et al. appear to predict that observers will always switch channels (where, by “channel”, we mean the weighting function across spatial frequency, as they suggest there is no need to invoke a fixed-channel model) in response to stimulus information (filtered letters or masking noise). In contrast, Majaj et al. suggest the channel is fixed and does not change with changes in masking noise. As we pointed out, there is information available for letter discrimination across a wide range of spatial frequencies (Figure 1), and observers are capable of using much of that range when forced due to spatial filtering of the stimuli (Parish & Sperling, 1991). Almost certainly, human observers switch channels in these experiments.

If the observer had a flat CSF, and hence flat equivalent noise, the CSF-ideal model would predict scale-invariant behavior. In these circumstances, when a letter is reduced or enlarged in size, there is no need to switch channels to improve signal-to-noise ratio. Consistent with this idea, Oruç et al. (2006) found perfect scale invariance for identification of texture-defined (i.e., second-order) letters, and the second-order CSF is relatively flat (Jamar & Koenderink, 1985; Kingdom, Keeble, & Moulden, 1995; Landy & Oruç, 2002; Sutter, Sperling, & Chubb, 1995).

Experiment 2 tests this hypothesis by estimating letter channels for several letter sizes using critical-band masking with an additional pedestal of white noise. If the CSF is measured for sine-wave gratings masked with sufficiently high-amplitude white noise, the white noise dominates the equivalent noise, resulting in a relatively flat CSF (e.g., Rovamo, Franssila, & Näsänen, 1992). According to the CSF-based account, under these conditions an ideal observer will use spatial frequencies in proportion to the signal energy useful for discrimination (since noise power is now independent of spatial frequency), resulting in scale invariance. **Experiment 2** tests whether this holds for human observers.

Preview

In this paper, we present two experiments. In **Experiment 1**, we used critical-band masking to estimate letter-

channel frequency at four noise levels. We show that as noise level increases, observers switch channels. **Experiment 2** is similar to **Experiment 1**, with the addition of a pedestal of white noise (on top of the low- and high-pass noises) to flatten the CSF. Despite the addition of the pedestal, observers remained scale dependent. **Appendix A** describes an additional experiment in which observers’ CSFs are measured both with and without the white-noise pedestal and demonstrates the effectiveness of the white-noise pedestal in flattening the CSF. All results in **Experiments 1** and **2** were compared with the results of simulations of the CSF-ideal observer described in **Appendix B**.

General methods

Apparatus

The experiments were run on a computer equipped with a Cambridge Research Systems VSG 2/3 card. Stimuli were displayed on a SONY Trinitron 17 in. monitor (model GDM-200 PS) at 1024×768 resolution. The monitor was gamma-corrected using an OptiCAL photometer (Model OP200-E) by Cambridge Research Systems using a software that generates and saves a gamma-correction look-up table. Mean luminance was 40 cd/m^2 . The standard viewing distance d was 91.4 cm. In **Experiment 2**, some subjects also ran at distances of $d/3$, $d/2$, and $3d$ (30.5, 45.7, and 274.2 cm, respectively). All stimuli were 512×512 pixels in size ($16.0 \times 16.0 \text{ cm}$) and were generated using Matlab 7.0 and Adobe Illustrator 10.

Procedure

Subjects were seated in a dark room. They entered their responses using the computer keypad. Auditory feedback was provided after each response: a single beep for a correct response and two beeps for an incorrect response. Detection and identification thresholds at 82% correct were measured using the Quest procedure (Watson & Pelli, 1983) with the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) in Matlab 7.0. In a given session, 16 blocks of trials were run by each observer, which lasted about an hour. Each block consisted of two interleaved staircases of 40 trials each in which the staircases controlled letter contrast. Subjects completed 1–3 blocks for each condition yielding 2–6 independent threshold estimates, each obtained at the end of a 40-trial staircase run. Stimulus generation, experimental code, data analysis, and ideal observer simulations were programmed in Matlab 7.0, using tools from the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) and the CRS VSG Toolbox for Matlab.

Stimuli

Letter stimuli ([Experiments 1](#) and [2](#)) used the five characters **D**, **N**, **R**, **S**, and **Z** from the Sloan font (Pelli, Robson, & Wilkins, 1988, available at <http://www.psych.nyu.edu/pelli/software.html>). At the standard viewing distance, the full stimulus image subtended 10×10 deg; the letters were 4.7×4.7 deg (letters in Sloan font have equal width and height by design). These specific letters were chosen from the entire set of 10 available in the Sloan font, as in our previous work (Oruç et al., 2006) because these five have approximately equal ink areas that covered $12\% \pm 0.4\%$ (mean $\pm SD$) of the entire stimulus. Stimuli were bright letters on a gray background masked by low- or high-pass-filtered noise.

Clipping

Noise-masking studies require a delicate balance so as to have sufficient noise power to dominate internal noise and raise contrast thresholds while preserving enough dynamic range in the stimuli to be able to achieve threshold. The resulting stimulus pixel values are confined to the range of the frame buffer (i.e., they range from 0 to 255). Values outside of this range are clipped, and one has to be careful that clipping artifacts do not contaminate the results.

We were careful to keep the amount of stimulus clipping to a minimum. In [Experiment 1](#) Gaussian white noise was used; in [Experiment 2](#) and [Appendix A](#) binary noise was used to enable additional noise power without engendering additional clipping artifacts. In both cases, the nominal “white noise” was, in fact, low-pass filtered with a cutoff of 80 cycles/letter, reducing the RMS power of the noise by 36% from the original Gaussian or binary white noise. Across all conditions (noise cutoff frequencies, low- or high-pass noise, and various signal contrasts), the maximum amount of clipping affected only 3.5% of image pixels, and that was only for the combination of highest signal-contrast level, highest noise level, and high-pass noise condition. For the low-pass noise, or the next-lower noise level, at most 0.5% of pixels was clipped. Thus, we are convinced that clipping artifacts did not contribute to our results and do not affect our conclusions.

Experiment 1: Channel switching

In this experiment, we used the critical-band-masking paradigm and obtained two estimates of letter-channel frequencies; one from the high-pass and another from the low-pass noise sweeps. We repeated the same experiment at several noise levels. We determined whether observers

were induced to switch the channel they used for letter identification when the noise level was sufficiently high.

Methods

Subjects

Five subjects took part in this experiment. All had normal or corrected-to-normal vision. Subject IO was an author.

Noise

Noise masks were low- or high-pass-filtered zero-mean Gaussian white noise (composed of independent, normally distributed pixels). There was a total of 16 noise conditions: white noise, no noise, and 7 low- and 7 high-pass noises with cutoff frequencies $f_c = 0.5, 1, 2, 5, 10, 20$, and 40 cycles/letter. A low-pass noise with 80 cycles/letter cutoff frequency was used in the nominally white-noise condition. Cycle/letter is defined as cycle/deg \times letter width in deg. The low-pass noise masks were produced by filtering white noise using a Butterworth filter with frequency response:

$$B(f) = \frac{1}{1 + (f/f_c)^{2n}}, \quad (1)$$

where f denotes spatial frequency and f_c is the cutoff frequency. The value $n = 5$ was used to provide a relatively steep filter cutoff. The Butterworth filter is a standard engineering filter design that enables one to control the rate of attenuation (by assigning an appropriate value to n) and minimize ringing in the filtered image. We produced the high-pass noise masks by filtering nominally white-noise masks (i.e., low-pass noise with 80 cycles/letter cutoff) with inverted versions of these filters (i.e., $1 - B(f)$). The root-mean-squared (RMS) contrast of the white noise prior to filtering had one of four possible values: 0.2, 0.4, 0.6, and 0.8. As a result, the noise maskers had a two-sided power spectral density of 1.53×10^{-5} , 6.1×10^{-5} , 1.37×10^{-4} , and $2.44 \times 10^{-4} \text{ deg}^2$, respectively, as viewed from 91.4 cm.

Procedure

The task was letter identification. Each trial began with a fixation cross displayed for 150 ms. Then, one of the five possible letters was presented for 150 ms with a superimposed noise mask (the letter and noise contrast values were summed). This was followed by a 150-ms blank period. Finally, a screen with all five possible letters was displayed until the subject responded by key press as to which letter she or he thought had been presented. Auditory feedback was provided to indicate whether or not the selection was correct.

The trials were blocked by noise level and type. There were four different noise levels, completed in separate sessions in a random order. Each session consisted of 16 blocks of noise maskers (at a fixed noise level) including low- and high-pass noises with 7 cutoff frequencies as well as white- and no-noise conditions, ordered randomly. In each block two estimates of the contrast threshold were measured via two randomly interleaved staircases for a given noise masker. Subjects completed one or two sessions per noise level, resulting in 2–4 threshold estimates per condition. Contrast threshold for letter identification was measured by adjusting the letter contrast while holding noise level constant.

Data analysis

For each type of noise mask, threshold elevation was computed as the difference between masked and unmasked squared contrast thresholds (which is proportional to signal contrast energy). A cumulative Gaussian was fit to the threshold elevation values for each low-pass mask as a function of the logarithm of noise cutoff frequency, f_c , using a least-squares criterion. The derivative of the best fitting curve with respect to cutoff frequency divided by f provided an estimate of the power gain of the channel used to identify the letters, assuming a fixed channel was used across those conditions (Solomon & Pelli, 1994). A second, independent fit of a reversed cumulative Gaussian was made to the threshold elevation data for the high-pass noise conditions, and differentiated and divided by f to provide a second channel estimate.

Results

Figure 5A shows threshold elevation (relative to the no-noise condition) as a function of noise cutoff frequency for the highest noise level for subject OY. As expected, for low-pass noise masks, threshold rises gradually and then asymptotes with increasing noise cutoff frequency (solid curve). For high-pass noise masks threshold falls over the same range (dotted curve). Threshold elevation for the no-noise condition, which is by definition zero, is plotted as part of both data series, once at the 0 cycle/letter cutoff for the low-pass series, and again at the 80 cycles/letter cutoff for the high-pass series. Similarly, the threshold elevation for the white-noise condition is also plotted as part of both data series, at 80 cycles/letter cutoff for the low-pass, and at the 0 cycle/letter cutoff for the high-pass series. Assuming threshold elevation is linearly related to the noise power that falls within the pass band of the underlying channel, the derivative of the threshold elevation curve divided by f provides an estimate of the power gain of the underlying channel. **Figure 5B** shows estimates of the power gains of channels estimated using the data from the low- (solid

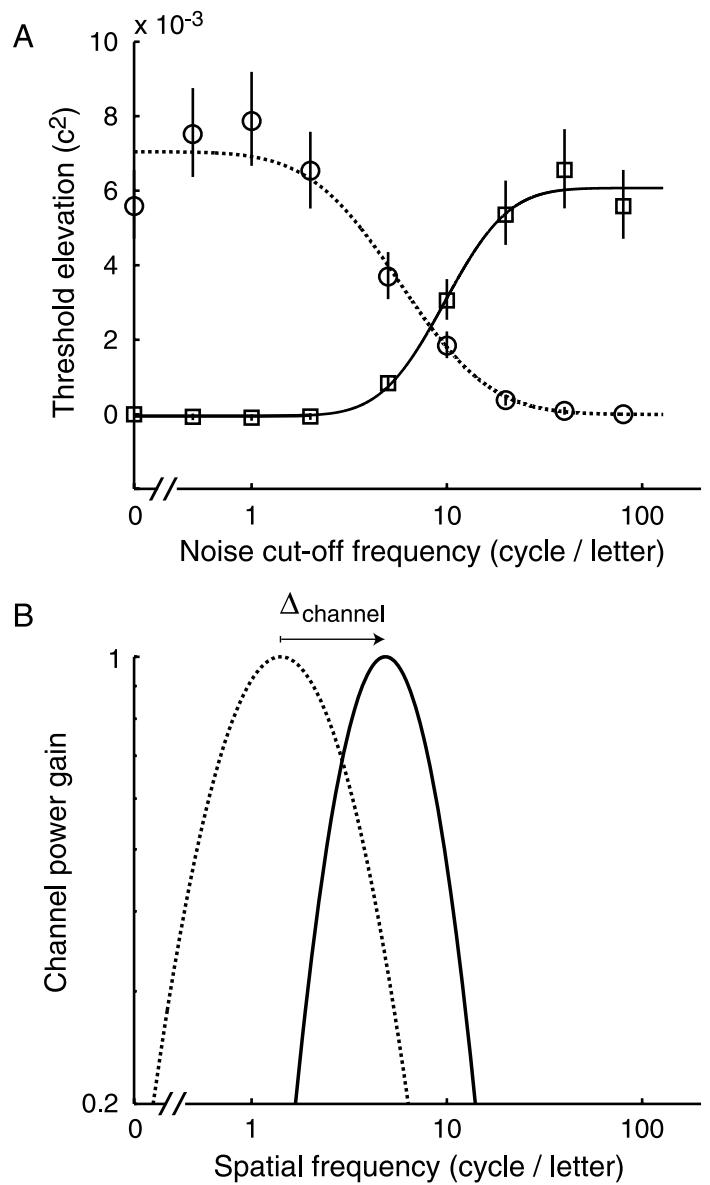


Figure 5. Estimation of letter channels. (A) Threshold elevation (squared contrast, which is proportional to signal energy) is plotted as a function of noise cutoff frequency for both low- (solid curve) and high-pass (dotted curve) noise conditions. Data are for subject OY in the highest noise-level condition. Cumulative Gaussians were fitted independently to the low- and high-pass data. Both data series include the no-noise condition (plotted at the 0 cycle/letter cutoff for the low-pass series, and again at the 80 cycles/letter cutoff for the high-pass series), which is by definition zero (i.e., baseline threshold subtracted from itself), and thus has no error bar. Also included in both data series is the white-noise condition, plotted at the 80 cycles/letter cutoff for the low-pass series, and again at 0 cycle/letter cutoff for the high-pass series. Error bars are 68% bootstrap confidence intervals. (B) The derivatives of the fitted curves divided by f provide estimates of the letter channel's power gain. The difference between the center frequencies of the two letter-channel estimates is Δ_{channel} .

curve) and high-pass (dotted curve) noise sweeps shown in [Figure 5A](#).

If the observer used a fixed channel to perform the task, independent of the characteristics of the noise masker, then the two estimates of channel sensitivity should be identical within measurement error. In particular, the center frequencies of the two channels (i.e., the frequencies resulting in peak sensitivity) should be the same. On the other hand, if the subject switches channels—to lower spatial frequencies in the face of high-pass noise, and to higher spatial frequencies in low-pass noise—then the degree of such channel switching can be estimated as the difference between the two channel center frequencies, $\Delta_{\text{channel}} = f_{LP} - f_{HP}$ ([Figure 5B](#)).

Δ_{channel} is plotted as a function of noise level in [Figure 6](#) for five subjects (solid line) as well as the group average. For each observer 2–4 estimates of Δ_{channel} were obtained

by separately fitting each set of high- and low-pass thresholds. The average and standard error of these estimates are shown in [Figure 6](#). We also simulated the CSF-ideal observer of Chung et al. (2002) as described in detail in [Appendix B](#) ([Figure 6](#), dashed lines). On each trial, the CSF-ideal observer was presented with a letter with added external masking and equivalent input noise (based on the observer's individual CSF, i.e., there were no free parameters for this prediction; [Appendix A](#)). The CSF-ideal observer chose the letter that was most likely given the noisy input stimulus. The CSF-ideal did not possess explicit “channels” but, nevertheless, produced simulated data similar to those in [Figure 5](#). The dashed lines in [Figure 6](#) show the values of Δ_{channel} estimated from the simulated CSF-ideal data. As expected ([Figure 3](#)), the CSF-ideal observer demonstrated increasing magnitude of channel switching (i.e., gave greater weight to a different

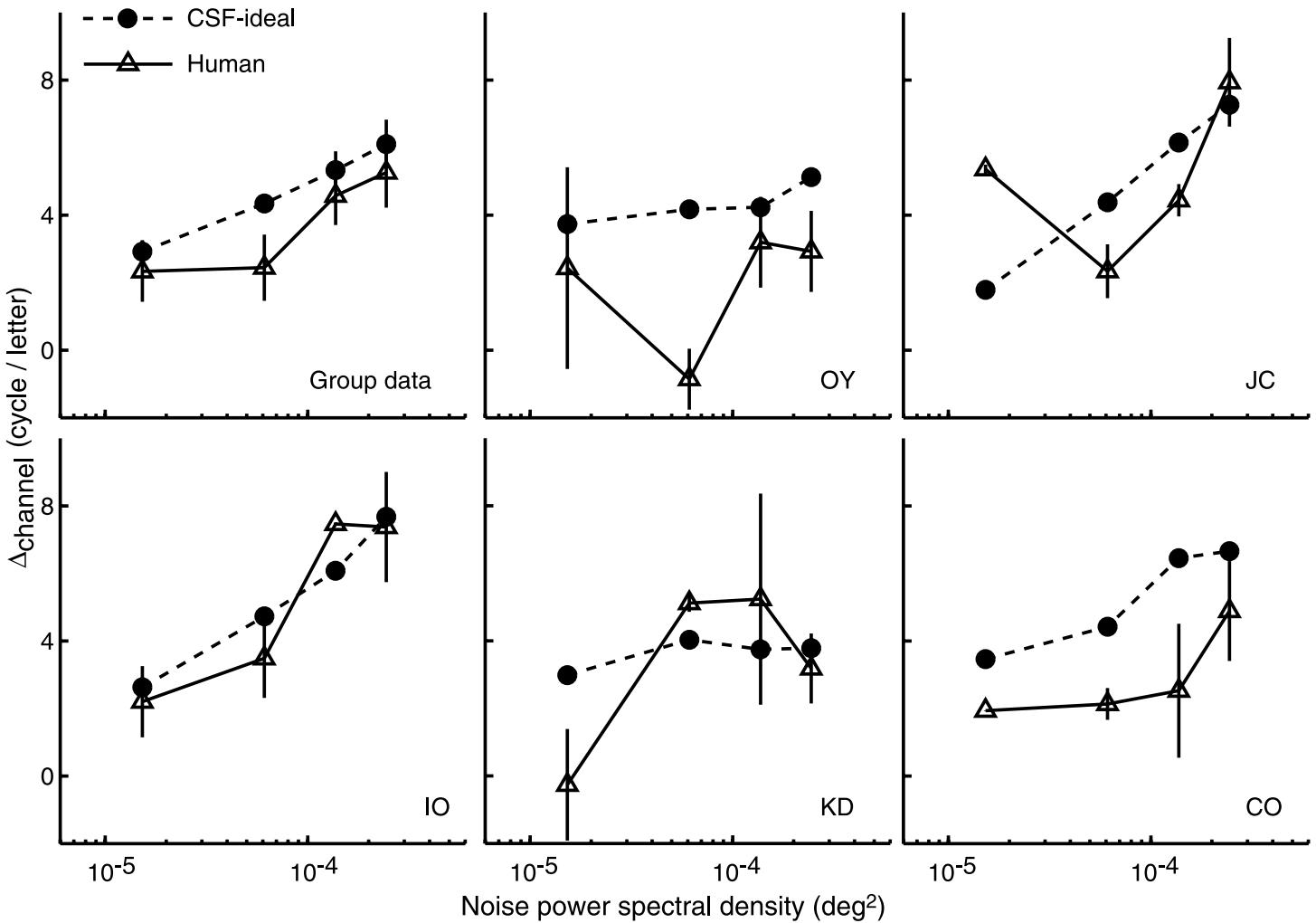


Figure 6. Channel switching. Δ_{channel} is plotted as a function of noise level for five observers and the group average (solid curves: human observers; dashed curves: corresponding CSF-ideal observers). For the group data (top left panel) error bars are SE across five observers for the human as well as the corresponding CSF-ideal results. For the individual observers' data (the remaining five panels) error bars are SE across separate fits to 2–4 sets of threshold elevation data for each observer. Human observers' data follow closely the individual predictions obtained from the CSF-ideal observer. The group data show both the CSF-ideal and the human observers increase channel switching, indicated by an increase in Δ_{channel} , with increases in noise level.

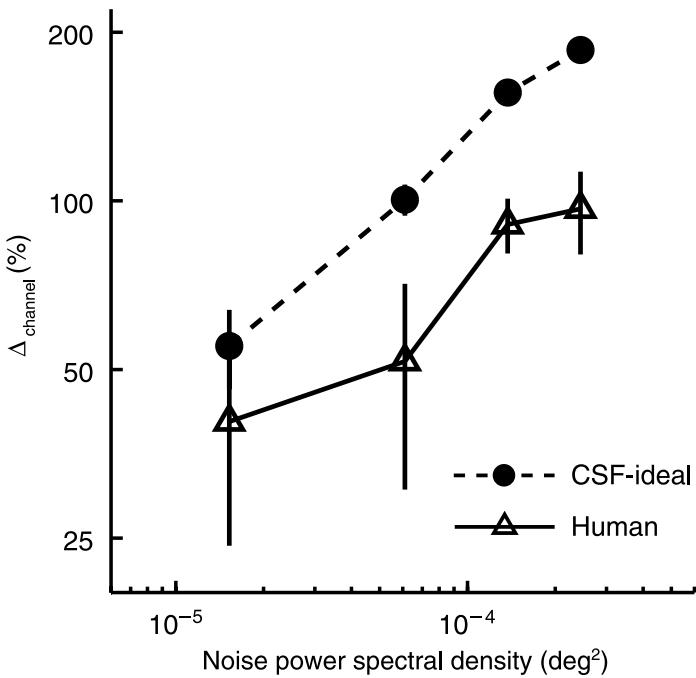


Figure 7. Average degree of channel switching. Δ_{channel} is plotted as a percentage (i.e., $((f_{LP} - f_{HP})/[(f_{LP} + f_{HP})/2]) \times 100\%$) as a function of noise power for the human observers and their corresponding CSF-ideal models. Error bars represent 68% bootstrap confidence intervals for percent Δ_{channel} estimates across the 5 subjects.

set of spatial frequencies) as the masking noise power was increased.

Although the data are somewhat variable, most observers' data are reasonably close to CSF-ideal model predictions.

The overall trend indicated by the group data (Figure 6, top left panel) shows an increase of Δ_{channel} with increasing noise level. Majaj et al. (2002) concluded that excess masking in their experiments (at most 0.2 dB on average) was negligible. However, when described in terms of the estimated shift of the peak frequency of the spatial frequency channel used to perform the task, we find combined shifts (leftward for high-pass noise plus rightward for low-pass noise) of over 90% of the average channel frequency in high-noise conditions (Figure 7).

We also performed an excess-masking analysis to bolster our claim that observers switch channels with sufficiently high masking-noise levels. As mentioned in the Introduction section, if observers use a fixed channel in all noise conditions, then the sum of the threshold elevations in response to low- and high-pass noise maskers with common cutoff frequency should equal the threshold elevation due to white noise. Here, we quantify the degree of channel switching by a *noise-additivity ratio* ($E_{low}^+ + E_{high}^+)/E_{all}^+$). If observers used a fixed channel, then this ratio should equal 1. If observers switch channels to escape the effects of masking noise, this ratio should fall below 1, especially near the preferred channel's peak frequency. In Figure 8, we plot the noise additivity as a function of masking-noise power. For each noise power, we plot the noise-additivity ratio for the noise cutoff frequency that results in the smallest noise-additivity ratio, i.e., near the intersection of the low- and high-pass threshold elevation curves.

Note that the prediction of noise additivity only holds for low- and high-pass noises whose power gains sum to one. This was true for the ideal filters used by Majaj et al. (2002) but is not true of the Butterworth filters used here (their amplitude gains sum to one, but their power gains

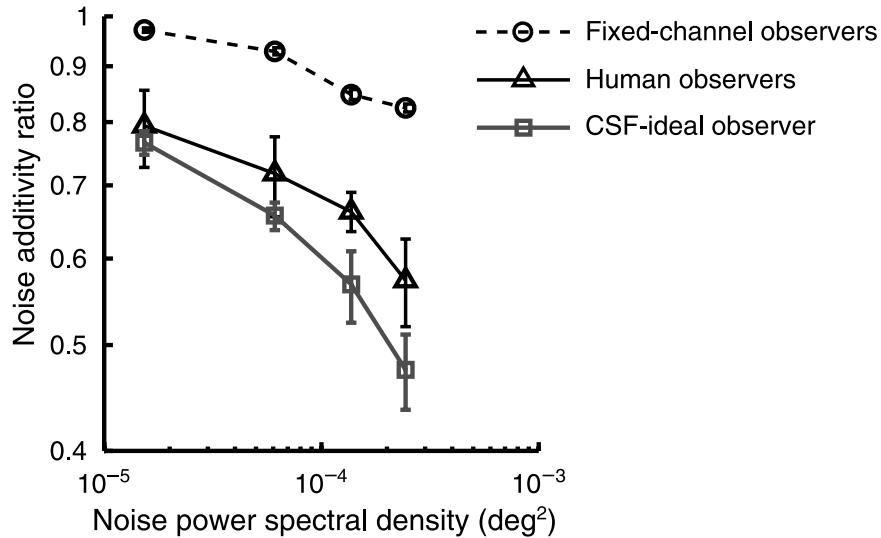
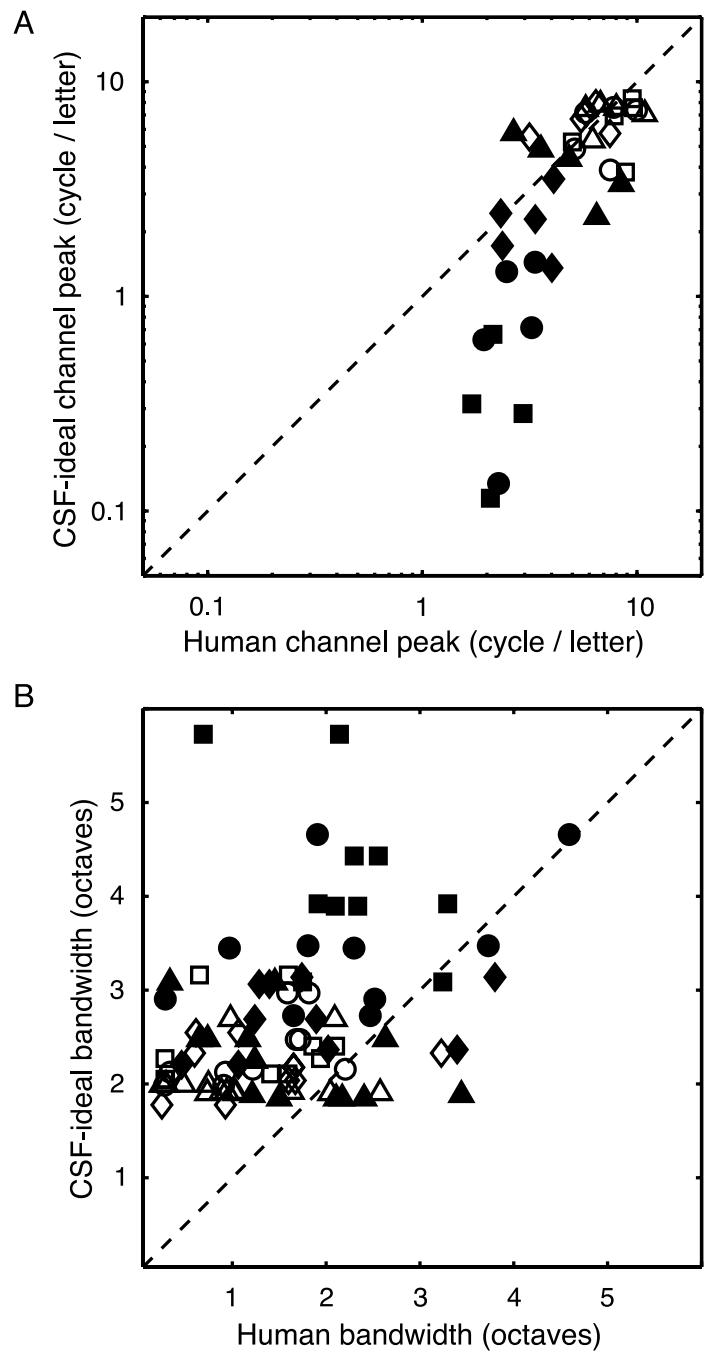


Figure 8. Noise-additivity analysis. The noise-additivity ratio ($(E_{low}^+ + E_{high}^+)/E_{all}^+$) is plotted as a function of noise level for the human observers as well as a simulated fixed-channel model and the CSF-ideal observer. The human observers display significantly more excess masking (leading to ratios below 1) than the fixed-channel model but slightly less than the CSF-ideal. Error bars represent 68% bootstrapped confidence intervals.

do not). For this reason, we do not compare human noise-additivity ratios to 1. Instead, we compare human noise-additivity ratios to those of a simulated fixed-channel observer faced with our Butterworth-filtered noise masks. For the simulations, we used a fixed channel with a Gaussian tuning function with a half-height bandwidth of 1.5 octaves (although the exact value of the bandwidth does not appear to be critical for our conclusions as similar results were obtained with bandwidths of 1 and 2 octaves). The peak frequency of this hypothetical channel for a given noise level and observer was chosen based on the results of [Experiment 1](#): we used the average of the peak-frequency estimates derived from the low- and high-pass noise sweeps. The input to this channel was the stimulus corrupted by both equivalent noise (based on each observer's CSF) and external noise (white, low- or high-pass). The noise-additivity ratio was calculated assuming that threshold elevation was proportional to the total noise power passed by the channel in each condition.

[Figure 8](#) shows the geometric average of human noise-additivity ratios (triangles) and the geometric average of fixed-channel predictions computed for each observer (circles) as a function of noise level. We also plot noise-additivity ratios for the CSF-ideal observer (squares) as a benchmark of the maximum possible advantage that can be obtained by switching channels. The fixed-channel and CSF-ideal results serve as worst-case and best-case scenarios in terms of the impact of channel switching (or lack of it) to compare to the human noise-additivity results. Error bars were computed based on a nonparametric bootstrap simulation in which we resampled noise-additivity ratios from the data set of all observers a large number of times with replacement. Geometric means obtained from the resampled data sets were then sorted and the 68% confidence intervals were given by the upper and lower 16th percentile values. At all noise levels human noise-additivity ratios fall well below those predicted by a fixed-channel model. In addition, the degree of excess masking increased faster for human observers than for the fixed-channel model with increases in noise level. The human noise-additivity ratios fall short of reaching the optimal values represented here by the CSF-ideal results. However, the difference between the human and CSF-ideal noise-additivity ratios is not substantial. These results confirm that human observers switch channels to avoid low- and high-pass noise maskers when the noise power is sufficiently high, consistent with our analysis of Δ_{channel} estimates.

The CSF-ideal-observer results predict increased channel switching with higher noise spectral density and are qualitatively in agreement with the human observers' results. However, we should point out that CSF-ideal predictions of channel parameters are poor. [Figure 9](#) shows a comparison of the estimates of letter-channel center frequencies and bandwidths compared with the corresponding CSF-ideal predictions. The peak frequency estimated from the human data almost never falls below



[Figure 9](#). Comparison of the CSF-ideal model results to the human data. (A) Peak channel spatial frequencies for the CSF-ideal model and the corresponding human channel peak estimates are shown in a scatterplot. Open symbols represent low-pass results, and solid symbols represent high-pass results. The markers \blacktriangle , \blacklozenge , \bullet , and \blacksquare indicate the four noise levels in increasing order. (B) Bandwidth estimates for the CSF-ideal observer and the corresponding human estimates are shown.

2 cycles/letter ([Figure 9A](#)). In contrast, under conditions encouraging the use of low spatial frequencies (high amplitude, high-pass noise), the “letter channel” estimated

using the CSF-ideal observer generally has a peak frequency lower than 1 cycle/letter. In addition, the estimated bandwidths of letter channels based on the human data are by and large narrower than those estimated using the CSF-ideal observer (Figure 9B). Note that our results indicate that human observers are able to switch channels in the face of different noise maskers. Thus, it is likely that they are using different channels at each cutoff frequency, rather than just one high- and one low-pass channel as we have plotted them. Seen this way, the channels we report should be interpreted as a composite of the channels used across all cutoff frequency conditions. However, our comparison to the CSF-ideal is valid as these data were analyzed in the same manner.

These results are in contrast with those of Chung et al. (2002), who argued that the CSF-ideal observer did an excellent job of accounting for contrast thresholds for the identification of band-pass-filtered letters. The results of Experiment 1 indicate that human discrimination performance is not governed solely by the information content of the stimuli as limited by equivalent noise.

Discussion

Majaj et al. (2002) used the critical-band-masking paradigm to estimate the channels used for letter identification. They did this for a variety of fonts including the Sloan font used here, and a range of letter sizes (0.18 to 55 deg) that includes the size used here (4.7 deg). They found hardly any evidence for a significant degree of channel switching. That is, they estimated nearly the same channel shape using low- and high-pass noises (i.e., Δ_{channel} was nearly zero).

On the other hand, we did find evidence for a substantial amount of channel switching when noise level was sufficiently high. More importantly, we found that the amount of channel switching increased with increasing noise level. We have already noted that Perkins and Landy (1991) also found substantial excess masking, most of which they attributed to channel switching.

Is it possible to reconcile these seemingly opposite findings? One possibility is that Majaj et al. (2002) used weaker noise maskers. In fact, their noise RMS contrast levels (before filtering) only ranged from 0.15 to 0.35, which is substantially lower than most of the contrast range we used (0.2–0.8) in the present study. However, the relevant variable for determining signal-to-noise ratio is not noise contrast but rather noise spectral density, which depends on both noise contrast as well as viewing distance (for a fixed pixel size on the display screen). Majaj et al. varied viewing distance over a hundred-fold range. Thus, their closest viewing distances resulted in noise maskers with high values of noise spectral density, which should have resulted in substantial excess masking. On the other hand, those viewing conditions were paired with large letter stimuli, shifting the letter spectrum to low

spatial frequencies with large equivalent noise. Thus, to reconcile our results with theirs would require one to calculate the CSF-ideal observer's predictions for their specific conditions. However, Majaj et al. do not provide the specific values for these parameters, i.e., the noise spectral densities used with each letter size. In any case, our results argue against their conclusion that the letter "channel" is chosen bottom-up by the letter size and not affected by signal-to-noise-ratio considerations.

Experiment 2: Scale dependence

Why is letter identification scale dependent? Assume that the letter channel for identifying a 2-deg letter is found to have peak sensitivity at 8 cycles/letter (i.e., 4 cycles/deg). When identifying a smaller, 0.5-deg letter, the same 8 cycles/letter corresponds to 16 cycles/deg. Thus, if observers processed letters in a scale-invariant manner, then for this smaller letter their performance would suffer due to the relatively poor contrast sensitivity at such high retinal spatial frequencies. For this reason the observer would do well to use a lower frequency channel when identifying smaller letters and, by the same logic, to use a higher frequency channel (relative to letter size) for larger letters, all due to the band-pass CSF. This is the explanation Chung et al. (2002) gave for the scale dependence they found, and by implication that found by Majaj et al. (2002). Chung et al. suggested their experimental data supported this conclusion, but these data were based on the identification of band-pass-filtered letters and thus were collected under conditions that likely forced observers to switch channels from one condition to another.

In Experiment 2 we explicitly tested this idea. We retested the scale-dependence result by measuring human observers' letter channels at three different letter sizes. We show that the CSF-ideal observer shows a similar pattern of scale dependence. It is well known that the CSF is flatter when measured on a high-contrast white-noise pedestal (e.g., Rovamo et al., 1992). We demonstrate this in Appendix A in which the CSF is measured both with and without added white noise. In Experiment 2, we also estimated letter channels for the same three letter sizes in the presence of a white-noise pedestal. By the above logic, scale dependence should disappear or at least become less prominent in the presence of high-contrast white noise, if that noise was effective in flattening the CSF.

Methods

Subjects

Two subjects participated in this experiment: KD and IO. Both had normal or corrected-to-normal vision. IO was an author.

Procedure

[Experiment 2](#) was similar in procedure to [Experiment 1](#). Letter channels were estimated using the same procedure as [Experiment 1](#) either without or with an additional white-noise pedestal.

Stimuli

The medium-sized letter stimuli were identical to those in [Experiment 1](#) (4.7×4.7 deg at the standard viewing distance). Letter size was varied by having observers view the same stimuli from 3, 0.5, and 0.33 times the standard viewing distance.

The noise power spectral density of the low- and high-pass masking noises was chosen to be low enough to minimize channel switching and to leave room within the dynamic range of our display for the additional white-noise pedestal but also high enough to produce detectable levels of threshold elevation. The low- and high-pass masking noise was generated by filtering binary white noise (independent pixels randomly selected to be either dark or light). The RMS contrast of the noise was 0.5 before filtering, resulting in a noise power spectral density of 9.54×10^{-5} deg² at the standard viewing distance of 91.4 cm. A new white-noise pedestal was generated for each trial. The contrast of the white pedestal noise needed to be high enough to produce significant flattening of the CSF. Therefore, we chose the highest prefiltering contrast (0.5) allowed by the dynamic range of our display (the stimulus, the masking noise, and the pedestal white noise were summed), resulting in a noise power spectral density of 9.54×10^{-5} deg² at the standard viewing distance of 91.4 cm. Binary noise was used to minimize clipping. Again, our nominal white-noise condition was low-pass filtered with an 80 cycles/letter cutoff. Note that an identical noise pedestal was used for the CSF measurements described in [Appendix A](#).

Noise spectral density for a fixed stimulus on the monitor depends on viewing distance. We varied letter size by varying viewing distance. At the longest viewing distance ($3d$), noise spectral density was reduced by a factor of 9. Thus, the flattening of the CSF required for this experiment (and demonstrated in [Appendix A](#)) may no longer hold for the longest viewing distance. Therefore, we also included a condition at the standard viewing distance with reduced letter size on the monitor (and noise spectral density equal to other conditions run at that standard viewing distance). This is not an issue for the closer viewing distances, at which noise spectral density is stronger than at the standard viewing distance.

Data analysis

Data analysis was similar to that of [Experiment 1](#). However, for each condition the cumulative Gaussian curves fit to the high- and low-pass data were constrained

to yield the same channel center frequency. This was because we were specifically interested in determining the peak channel sensitivity and chose conditions to minimize channel switching.

Results

[Figure 10](#) shows the results of [Experiment 2](#). The two upper panels show the data for the two subjects, and the two lower panels show the results for the corresponding CSF-ideal observers. Letter-channel center frequency (in cycle/letter) is plotted as a function of letter size. In the absence of the white-noise pedestal (open symbols), the data are similar to the results of Majaj et al. (2002; gray curve): larger letters were identified using channels with higher peak spatial frequency (in cycle/letter). That is, letter identification was not scale invariant. We expected the white-noise pedestal to result in performance that was more scale invariant. Although the CSF-ideal results flatten completely with the addition of the white-noise pedestal, the human data from the conditions with the white-noise pedestal (dashed line) show, if anything, an exaggerated scale-dependent pattern.

The letter size was changed by varying viewing distance while leaving the stimuli on the display unchanged. As noted above, changing the viewing distance not only changes the retinal size of the letters but also varies the noise spectral density of noise maskers, including the noise pedestal used to flatten the CSF. This is not a problem at the close viewing distance/large letter size, as that viewing distance results in an increased noise spectral density, which, if anything, should have been more effective at flattening the CSF. However, one could argue that the pedestal was ineffective at flattening the CSF at the far distance/small letter size. To control for this, we measured the channel frequency a second time for the small letter-size condition by using the standard viewing distance, the same types of noise, but displaying physically smaller letters on the screen ([Figure 10](#), gray triangles). In other words, we remeasured that condition using a noise pedestal with higher noise spectral density. This had no effect on the measured channel frequency for one observer (IO) and produced an even less scale-invariant result for the other observer (KD).

Discussion

Is human scale dependence in letter identification caused by the band-pass shape of CSF? It is interesting to note that the CSF-ideal observer shows a qualitatively similar pattern of scale dependence to the human observer as measured using low- and high-pass noise maskers. Clearly, a truly ideal observer (i.e., an ideal observer unhampered by the equivalent noise used in the CSF-ideal simulations) is, by definition, scale invariant. The ideal

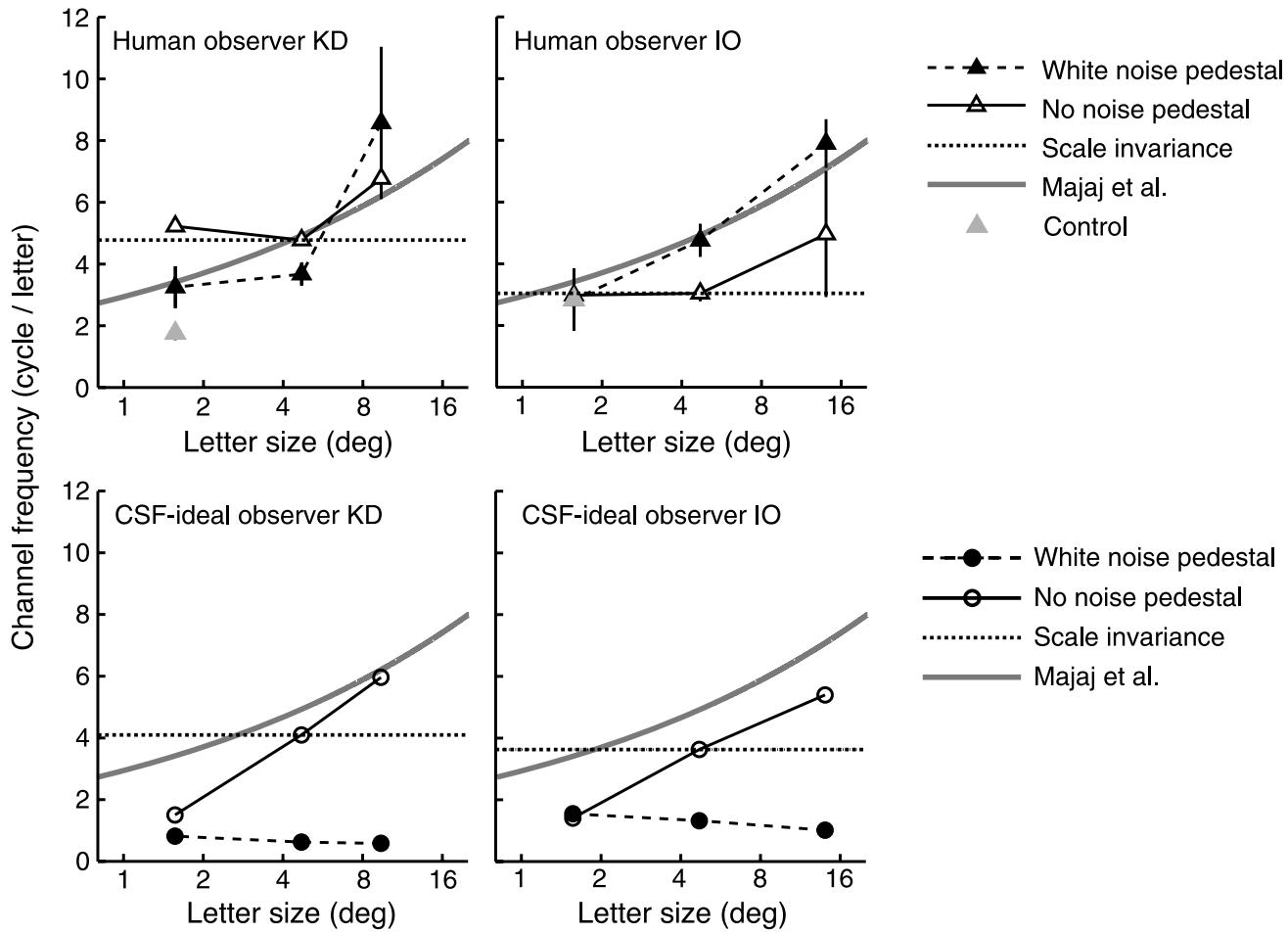


Figure 10. Scale dependence. Letter-channel center frequency is plotted as a function of letter size. (Top panels) Human data. (Bottom panels) Results of CSF-ideal observer simulations. With no added white noise, the human data (open triangles) are similar to the scale-dependence result of Majaj et al. (2002; solid gray curve). The results of the CSF-ideal observer (open circles) are similar. With an added white-noise pedestal, we predict scale invariance (e.g., horizontal dotted lines). The CSF-ideal observer (filled circles) did become more scale invariant; the human observers (filled triangles) did not. The gray triangles represent a control condition in which the smallest letter size condition with an added white-noise pedestal was carried out using a smaller letter on the display at the viewing distance corresponding to the medium-sized letters. Error bars represent standard errors across the 2–6 channel estimates for each condition.

observer simply uses all spatial frequencies weighted by their signal-to-noise ratio, and the amount of signal at each spatial frequency (in cycle/letter) is independent of letter size. The CSF-ideal observer is simply an ideal observer confronted with equivalent noise. The CSF-ideal observer displays scale dependence. Thus, the human no-pedestal results are at least qualitatively consistent with the hypothesis that scale dependence results from the shape of the CSF.

However, following this rationale, one should be able to eliminate scale dependence for both the human and the CSF-ideal observers by flattening the CSF. We attempted to do this by the addition of a white-noise pedestal, which does indeed flatten the CSF considerably (Appendix A). As expected, the white-noise pedestal resulted in scale invariance for the CSF-ideal observer. Surprisingly, the flattening of the CSF did not have this effect on human

behavior in Experiment 2; scale dependence was still evident in the presence of the white-noise pedestal.

With a white-noise pedestal, the channel frequency (in cycle/letter) increased substantially for the human observers as letter size was increased, while the CSF-ideal prediction changed little and, if anything, decreased slightly. This result is in conflict with the claim of Chung et al. (2002) that letter discrimination performance can be explained solely by the information content of the stimuli as masked by external and equivalent noise.

Why were we unable to eliminate scale dependence with the addition of a white-noise pedestal? We speculate that scale dependence is the result of long-term adaptation to the CSF. Thus, the visual system uses an optimal choice of visual channels for the identification of letters (and perhaps of other, well-trained visual stimuli) on the basis of the large amount of experience it has with such stimuli.

The visual system may not always monitor changes in the total effective noise spectrum and alter its behavior on a short-term basis. As a result, the same channels are used in the presence of the white-noise pedestal as without it. It remains to be seen whether observers become scale invariant in the presence of white noise if given sufficient training.

General discussion

How are letters identified visually? The experiments reported in this paper analyze the spatial frequency information used to discriminate letters at discrimination threshold. These experiments do not tell us what computation is used to perform detection leading to the spatial channels we estimate, nor how letters are processed under supra-threshold conditions. Indeed, an analysis of such mechanisms might well be more clearly described in the spatial, rather than spatial-frequency domain, resulting in a description of the visual “features” that are computed to represent and discriminate letters. New techniques are being developed to estimate such visual features (e.g., Fiset et al., 2008). Here, we concentrate on the spatial-frequency content of such features as estimated through masking experiments.

Are there specialized channels for the identification of letters, or are letters identified using the same first-stage filters used for the detection and identification of other spatial patterns? Nearly 40 years of vision research have led to a model of spatial vision that begins with a set of channels tuned for spatial frequency, orientation, and several other visual features (De Valois & De Valois, 1988; Graham, 1989). Such a model can account for the detection of a wide variety of visual patterns including bars, edges, checkerboards, and so on (Graham, 1980; Watson & Ahumada, 2005). Letter identification appears to be carried out using a spatial channel with a bandwidth (1.6 octaves according to Majaj et al., 2002) somewhat larger than those measured using sine-wave-detection tasks (0.5–1 octave, as summarized by Graham, 1989; although ranging from 0.5 to as high as 2.4 in the summary provided by De Valois & Devalois, 1988), leading one to the conclusion that letter identification is carried out using either a different mechanism or a combination of several of the spatial channels used for the detection of gratings and other simple patterns.

Some visual stimuli are clearly special in the sense either that they are particularly important for the organism or are stimuli with which the observer has a huge amount of experience. For example, faces are clearly important for humans and other primates, and we have a lot of experience with face discrimination. Moreover, there is evidence that face identification has special properties behaviorally (Bruce, Doyle, Dench, & Burton, 1991;

Tanaka & Farah, 1993; Yin, 1969) and that there are brain areas specific to face identification or at least for the identification of well-learned visual patterns (Kanwisher, McDermott, & Chun, 1997). There may be a brain area specific to the analysis of visual word forms independent of font as well (McCandliss, Cohen, & Dehaene, 2003). Thus, there may well be specific mechanisms dedicated to a well-learned task like letter identification.

On the other hand, Chung et al. (2002) conclude that there is no “letter channel” per se, and that the results of letter-identification experiments are merely the result of observers using different spatial frequencies optimally in the face of externally imposed and equivalent noise. They suggest that this ideal behavior is the explanation for the scale dependence found by Majaj et al. (2002). In their implementation of the CSF-ideal observer, they simulated an observer identifying band-pass-filtered letters in white noise, which is not the experimental conditions that led Majaj et al. (2002) to the conclusion that letter identification is scale dependent. Our study extends that of Chung et al. (2002) by comparing a CSF-ideal model with human behavioral data under precisely the same conditions as those of Majaj et al. (2002). Our simulations confirm the conclusion of Chung et al. (2002) that scale dependence would result from an ideal observer hampered by equivalent noise. Our behavioral data are qualitatively similar to that of the CSF-ideal but differ quantitatively in many details. There is a mismatch of estimated filters for the human observers and the CSF-ideal in **Experiment 1**. A white-noise pedestal results in scale invariance for the CSF-ideal but not the human observers in **Experiment 2**. Thus, we disagree with the conclusion of Chung et al. (2002) that the CSF-ideal explains letter discrimination with noise maskers.

Scale dependence is consistent with an observer using the stimulus information optimally. That is, the simulations are consistent with a model (the CSF-ideal) that is scale invariant *except* for the scale dependence caused by the equivalent noise. However, our measurements in the presence of a white-noise pedestal argue strongly against this. The pedestal effectively flattened the CSF but did *not* eliminate scale dependence.

We conclude from our results that the estimated letter channel is the spatial frequency sensitivity of an internal mechanism used for letter identification and not merely an epiphenomenon of the information available to make the discrimination. It may be constructed out of the filters revealed in sine-wave-detection experiments (each narrower in bandwidth than the letter channel). Its sensitivity profile may be determined by evolution or by previous training. More than one such mechanism is available to the observer, and observers can switch channels under some circumstances (e.g., when letters are masked by strong low- or high-pass noise, or for band-pass-filtered letter stimuli). However, observers are not capable of reacting perfectly to changes in the dependence of signal-to-noise ratio on spatial frequency. In particular, scale

dependence results from using letter size to determine the choice of channel, and observers stick with that choice even in the face of a noise pedestal that causes that choice to be suboptimal.

independent of the spatial frequency). The spatial frequencies were 0.5, 1, 2, 4, 8, and 16 cycles/deg. In the white-noise conditions, the same white-noise pedestal was used as in [Experiment 2](#): binary noise with 0.5 contrast, low-pass filtered with a cutoff of 80 cycles/letter (i.e., 17 cycles/deg).

Conclusions

1. Letter channels are not fixed. Observers can switch channels when it is useful to do so but do not always switch by the optimal amount.

2. Humans process letters in a scale-dependent manner consistent with a long-term adaptation to the shape of the CSF. Observers are not able to alter this behavior on a short-term basis.

3. Letter discrimination uses a mechanism with a relatively narrow bandwidth. The sensitivity profile of this mechanism is not merely an indication of the stimulus information available to the observer to perform the task.

Procedure

The task was 2-interval, forced-choice (2IFC) detection. In each trial there were two intervals. A Gabor patch was presented in one interval; the other interval contained a blank stimulus. In the white-noise conditions, an independent white-noise pedestal was present in both intervals. Each trial started with a 150-ms fixation cross, followed by 150-ms blank, and then the first stimulus was displayed for 150 ms. This was followed by another 150-ms display of the fixation cross, a 150-ms blank, and then the second stimulus was displayed for 150 ms. Finally a blank screen was displayed until the subject indicated by key press whether the Gabor was in the first or the second interval.

Appendix A

CSF experiments

We measured the contrast sensitivity function (CSF) for each subject using Gabor patches (sine-wave gratings windowed by a Gaussian envelope). This was done both with and without the addition of a white-noise masker. The resulting contrast sensitivities were then transformed into equivalent noise values for use in the CSF-ideal observer simulations described in [Appendix B](#).

Methods

Subjects

The CSF was measured for the five subjects who participated in [Experiment 1](#) (CO, IO, JC, KD, OY) without added noise. In addition, the CSF was measured in the presence of an added white-noise pedestal for two subjects (IO and KD). All subjects had normal or corrected-to-normal vision. IO was one of the authors.

Stimuli

Vertical sine-phase Gabor stimuli were used. The Gabors had 1.2 octaves full bandwidth at half height (i.e., the Gaussian envelope had an SD of 0.5 cycles

Data analysis

Threshold signal energy at each spatial frequency was computed as squared contrast integrated over the stimulus area. This value was then converted into an estimate of equivalent noise at that spatial frequency (see [Appendix B](#); Ahumada & Watson, 1985; Green & Swets, 1988).

Results

[Figure A1](#) shows threshold signal energy as a function of spatial frequency without (solid curve) and with (dashed curve) added white noise for two observers. The results are consistent with the literature (Kukkonen, Rovamo, Tiippana, & Näsänen, 1993; Rovamo et al., 1992): we find U-shaped curves when no external visual noise is added. However, when Gabors are detected in the presence of white noise, the curves become significantly flatter as required for [Experiment 2](#). Note that at the highest Gabor frequency (16 cpd) subjects did not reach threshold performance in the presence of added white noise. As an aside, the observer's calculation efficiency may be calculated as $d'^2 N/E_{\text{ped}}^+$, where d' is 1.3 corresponding to our 82% correct criterion and 2IFC task, N is the noise power spectral density, and E_{ped}^+ is the increase in threshold signal energy due to the white-noise pedestal (Pelli & Farell, 1999). This results in estimated

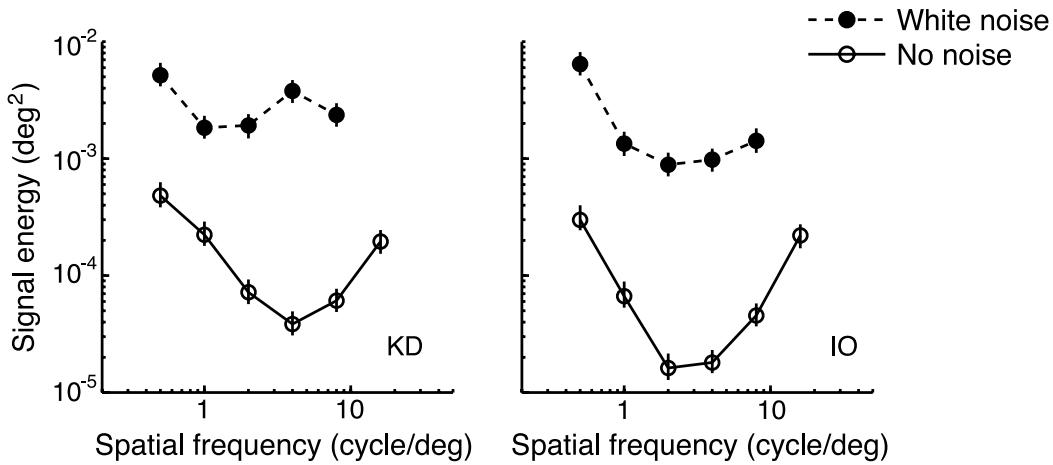


Figure A1. Gabor detection. Threshold signal energy for detecting vertical Gabor patches of varying spatial frequency is shown for two observers as a function of spatial frequency. Solid lines: No external noise. Dashed lines: Added white-noise pedestal. The no-noise result is as expected, with highest sensitivity around 2–4 cycles/deg. The white-noise pedestal increases thresholds and flattens the curve. At 16 cycles/deg subjects did not reach threshold in the white-noise condition. Error bars indicate 68% bootstrap confidence intervals.

calculation efficiencies on the order of 10% here, relatively independent of spatial frequency, which is a value consistent with the literature for this task.

energy was measured, we determined the equivalent noise power spectrum as

$$N_{eq}(f) = \frac{c^2(f)}{d'^2} \sum_{x,y} C_f^2(x,y), \quad (\text{B1})$$

where $c^2(f)$ is the squared contrast threshold for spatial frequency f , $C_f(x, y) = (L_f(x, y) - L_0)/L_0$ is the local contrast of the Gabor patch, $L_f(x, y)$, when it is at full contrast, d' is the value of d' corresponding to the 82% threshold criterion for the 2IFC contrast detection task in [Appendix A](#), and L_0 is the mean luminance of the Gabor patch.

An equivalent input noise curve $N_{eq}(f)$ was computed by first fitting a smooth curve given by

$$c^2(f) = p_1(1 + (p_2 f)^{p_3})(1 + (p_4 f)^{-p_5}), \quad (\text{B2})$$

to the measured values where f is spatial frequency and p_1, \dots, p_5 are free parameters constrained to be non-negative. We extrapolated noise power from 0 to 0.5 cycle/deg and from 16 to 25.6 cycles/deg (the Nyquist frequency for our viewing conditions) based on the best fitting curve (for comparison with an alternative method, see Nandy & Tjan, [2008](#)). For simplicity, we assumed sensitivities to be independent of orientation and retinal eccentricity, so we defined $N_{eq}(f_x, f_y) = N_{eq}(\sqrt{f_x^2 + f_y^2})$, redefining f_x and f_y in units of cycle/image, and N_{eq} in units of noise spectral density (noise contrast variance per unit bandwidth, in units of pixels²).

Appendix B

Equivalent noise and the CSF-ideal observer

For [Experiments 1](#) and [2](#), human performance was compared to the performance of a simulated observer performing the same task under the same conditions. In this appendix, we describe the simulations in detail. The calculation of the CSF-ideal observer involved two steps. First, based on the measurements of the human observer's CSF ([Appendix A](#)), we calculated the equivalent noise, i.e., the noise spectrum that, when added to the Gabor patch stimuli, would lead an ideal observer to perform identically to the human observer in the Gabor-detection task. Second, for each simulated trial of the letter-discrimination task, the letter stimulus was corrupted by both the external noise that the human observers confronted as well as the equivalent noise. The CSF-ideal observer made the optimal letter-identification decision given that noisy stimulus.

The equivalent noise spectrum

The power spectrum of the equivalent noise (Ahumada & Watson, [1985](#)) for each human observer was computed based on the contrast thresholds described in [Appendix A](#). For each spatial frequency f for which threshold signal

The model

On each trial, the CSF-ideal observer was given a stimulus that consisted of one of the five letters at a

contrast determined by the staircase procedure, with added samples of external noise N_{ext} and N_{eq} . The calculations are most easily described in the Fourier domain. The stimulus is $S(f_x, f_y)$. At the current contrast level, there were five possible letter templates, $T_1(f_x, f_y), \dots, T_5(f_x, f_y)$. In the Fourier domain, the noise consisted of independent Gaussian-distributed Fourier components, with variance of the real and imaginary components at each spatial frequency equal to $N(f_x, f_y) = N_{ext}(f_x, f_y) + N_{eq}(f_x, f_y)$.

The ideal observer chose the letter template T_i such that the likelihood $P(T_i|S)$ was maximized. Since the letters occurred with equal frequency, this was equivalent to a Bayesian, maximum a posteriori decision rule. We assume that the CSF-ideal observer knew the noise statistics as well as the current letter contrast. The maximum likelihood decision is easiest to describe by prewhitening the stimulus (Barrett & Myers, 2004, p. 839; Chao, Tai, Dymek, & Yu, 1980; Cook & Bernfeld, 1967). That is, we compute the whitened stimulus $S'(f_x, f_y) = S(f_x, f_y)/\sqrt{N(f_x, f_y)}$ for which the noise for each Fourier component is now a standard normal and determine the maximum likelihood prewhitened letter $T'_i(f_x, f_y) = T_i(f_x, f_y)/\sqrt{N(f_x, f_y)}$. The maximum likelihood decision was to choose the letter i that minimized the squared distance $d(T'_i, S')^2 = \sum_{f_x, f_y} (T'_i(f_x, f_y) - S'(f_x, f_y))^2$.

Acknowledgments

This research was supported in part by NIH Grant EY16165 to MSL and by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada. We thank Chris Oriet and Jen Corbett for volunteering their time to participate in our experiments, and Walter Bischof, Alan Kingstone, and Jason Barton for making the experimental equipment available and for their support. We thank Benjamin Backus, David Brainard, Denis Pelli, Bosco Tjan, and two anonymous reviewers for valuable comments on early drafts of this paper.

Commercial relationships: none.

Corresponding author: Ipek Oruç.

Email: ipek@psych.ubc.ca.

Address: Vancouver General Hospital, Eye Care Centre, 2550 Willow Street, Vancouver, BC V5Z 3N9, Canada.

References

- Ahumada, A. J., Jr., & Watson, A. B. (1985). Equivalent-noise model for contrast detection and discrimination. *Journal of the Optical Society of America A, Optics and Image Science*, 2, 1133–1139. [PubMed]
- Banks, M. S., Geisler, W. S., & Bennett, P. J. (1987). The physical limits of grating visibility. *Vision Research*, 27, 1915–1924. [PubMed]
- Barrett, H. H., & Myers, K. (2004). *Foundations of image science*. Hoboken, NJ: John Wiley & Sons.
- Biederman, I., & Cooper, E. E. (1992). Size invariance in visual object priming. *Journal of Experimental Psychology—Human Perception and Performance*, 18, 121–133.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10, 433–436. [PubMed]
- Bruce, V., Doyle, T., Dench, N., & Burton, M. (1991). Remembering facial configurations. *Cognition*, 38, 109–144. [PubMed]
- Chao, T. H., Tai, A. M., Dymek, M. S., & Yu, F. T. S. (1980). Optimum correlation detection by pre-whitening. *Applied Optics*, 19, 2461–2464.
- Chung, S. T., Legge, G. E., & Tjan, B. S. (2002). Spatial-frequency characteristics of letter identification in central and peripheral vision. *Vision Research*, 42, 2137–2152. [PubMed]
- Cook, C. E., & Bernfeld, M. (1967). *Radar signals: An introduction to theory and applications*. New York: Academic.
- De Valois, R., & De Valois, K. (1988). *Spatial vision*. New York: Oxford University Press.
- Fiset, D., Blais, C., Ethier-Majcher, C., Arguin, M., Bub, D., & Gosselin, F. (2008). Features for identification of uppercase and lowercase letters. *Psychological Science*, 19, 1161–1168. [PubMed]
- Furmanski, C. S., & Engel, S. A. (2000). Perceptual learning in object recognition: Object specificity and size invariance. *Vision Research*, 40, 473–484. [PubMed]
- Gold, J., Bennett, P. J., & Sekuler, A. B. (1999). Identification of band-pass filtered letters and faces by human and ideal observers. *Vision Research*, 39, 3537–3560. [PubMed]
- Graham, N. (1980). Spatial frequency channels in human vision: Detecting edges without edge detectors. In C. Harris (Ed.), *Visual coding and adaptability* (pp. 215–262). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Graham, N. (1989). *Visual pattern analyzers*. New York: Oxford University Press.
- Green, D. M., & Swets, J. A. (1988). *Signal detection theory and psychophysics*. Los Altos, CA: Peninsula Publishing.
- Jamar, J. H., & Koenderink, J. J. (1985). Contrast detection and detection of contrast modulation for noise gratings. *Vision Research*, 25, 511–521. [PubMed]

- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, 17, 4302–4311. [PubMed] [Article]
- Kingdom, F. A., Keeble, D., & Moulden, B. (1995). Sensitivity to orientation modulation in micropattern-based textures. *Vision Research*, 35, 79–91. [PubMed]
- Kukkonen, H., Rovamo, J., Tiippana, K., & Näätänen, R. (1993). Michelson contrast, RMS contrast and energy of various spatial stimuli at threshold. *Vision Research*, 33, 1431–1436. [PubMed]
- Landy, M. S., & Bergen, J. R. (1991). Texture segregation and orientation gradient. *Vision Research*, 31, 679–691. [PubMed]
- Landy, M. S., & Oruç, I. (2002). Properties of second-order spatial frequency channels. *Vision Research*, 42, 2311–2329. [PubMed]
- Legge, G. E., Pelli, D. G., Rubin, G. S., & Schleske, M. M. (1985). Psychophysics of reading—I. Normal vision. *Vision Research*, 25, 239–252. [PubMed]
- Majaj, N. J., Pelli, D. G., Kurshan, P., & Palomares, M. (2002). The role of spatial frequency channels in letter identification. *Vision Research*, 42, 1165–1184. [PubMed]
- McCandliss, B. D., Cohen, L., & Dehaene, S. (2003). The visual word form area: Expertise for reading in the fusiform gyrus. *Trends in Cognitive Sciences*, 7, 293–299. [PubMed]
- Nandy, A. S., & Tjan, B. S. (2008). Efficient integration across spatial frequencies for letter identification in foveal and peripheral vision. *Journal of Vision*, 8(13):3, 1–20, <http://journalofvision.org/8/13/3/>, doi:10.1167/8.13.3. [PubMed] [Article]
- Oruç, I., Landy, M. S., & Pelli, D. G. (2006). Noise masking reveals channels for second-order letters. *Vision Research*, 46, 1493–1506. [PubMed]
- Parish, D. H., & Sperling, G. (1991). Object spatial frequencies, retinal spatial frequencies, noise, and the efficiency of letter discrimination. *Vision Research*, 31, 1399–1415. [PubMed]
- Patterson, R. D. (1974). Auditory filter shape. *Journal of the Optical Society of America A, Optics and Image Science*, 55, 802–809. [PubMed]
- Pelli, D. G. (1981). *Effects of visual noise*. Unpublished Ph.D. thesis, Cambridge, England: Cambridge University.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437–442. [PubMed]
- Pelli, D. G., Burns, C. W., Farell, B., & Moore-Page, D. C. (2006). Feature detection and letter identification. *Vision Research*, 46, 4646–4674. [PubMed]
- Pelli, D. G., & Farell, B. (1999). Why use noise? *Journal of the Optical Society of America A, Optics, Image Science, and Vision*, 16, 647–653. [PubMed]
- Pelli, D. G., Robson, J. G., & Wilkins, A. J. (1988). The design of a new letter chart for measuring contrast sensitivity. *Clinical Vision Sciences*, 2, 187–199.
- Perkins, M. E., & Landy, M. S. (1991). Nonadditivity of masking by narrow-band noises. *Vision Research*, 31, 1053–1065. [PubMed]
- Rovamo, J., Franssila, R., & Näätänen, R. (1992). Contrast sensitivity as a function of spatial frequency, viewing distance and eccentricity with and without spatial noise. *Vision Research*, 32, 631–637. [PubMed]
- Solomon, J. A. (2000). Channel selection with non-white-noise masks. *Journal of the Optical Society of America A, Optics, Image Science, and Vision*, 17, 986–993. [PubMed]
- Solomon, J. A., & Pelli, D. G. (1994). The visual filter mediating letter identification. *Nature*, 369, 395–397. [PubMed]
- Sutter, A., Sperling, G., & Chubb, C. (1995). Measuring the spatial frequency selectivity of second-order texture mechanisms. *Vision Research*, 35, 915–924. [PubMed]
- Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology Section A—Human Experimental Psychology*, 46, 225–245. [PubMed]
- Watson, A. B., & Ahumada, A. J., Jr. (2005). A standard model for foveal detection of spatial contrast. *Journal of Vision*, 5(9):6, 717–740, <http://journalofvision.org/5/9/6/>, doi:10.1167/5.9.6. [PubMed] [Article]
- Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, 33, 113–120. [PubMed]
- Wiskott, L. (2006). How does our visual system achieve shift and size invariance? In J. L. v. Hemmen & T. J. Sejnowski (Eds.), *23 Problems in systems neuroscience* (pp. 322–340). New York: Oxford University Press.
- Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, 81, 141–145.