The visual filter mediating letter identification

Joshua A. Solomon* & Denis G. Pelli†

Institute for Sensory Research, Syracuse University, Merrill Lane, Syracuse, New York 13244–5290, USA
* Present address: NASA Ames Research Center, Mail Stop 262-2, Moffett Field, California 94035–1000, USA
† To whom correspondence should be addressed

We hear periodic sounds, or tones, by means of parallel auditory filters, each tuned to a band of temporal frequency[1], and we see periodic patterns, or gratings, by means of parallel visual filters, each tuned to a band of spatial frequency[2]. Beyond helping us to see gratings, do these visual filters participate in everyday tasks such as reading and object recognition? After all, grating visibility only requires the distinguishing of pattern from blank, whereas object recognition, for example letter identification, requires classification by the observer into one of many learned categories. Here we make use of results from hearing research[3], applying to vision a noise-masking paradigm that reveals the filter(s) mediating any threshold task. We find that letter-identification and grating-detection filters are identical, showing that the recognition of these objects at one size is mediated and constrained by a single visual filter, or 'channel'.

Visual science has made great advances using tasks that de-emphasize cognitive processing in order to isolate 'low-level' physiological processes, such as the spatial frequency channels that mediate grating detection[2,4]. However, everyday visual tasks like reading[5] and object recognition[6] are cognitive, and it is unclear what role, if any, these channels play in such 'high-level' tasks. Letters, being over-learned, are good for testing the limits of object recognition. As letters—unlike gratings—are spatially compact and spectrally broad, we expected their identification to be mediated by multiple channels, or a much broader channel than the channel that mediates grating detection. Surprisingly, our noise-masking paradigm shows that the same channel performs both low-level detection of narrow-band gratings and high-level identification of broad-band letters.

Figure 1 demonstrates the unexpected bandpass nature of letter identification. Noise at 3 cycles per letter masks letter identification more effectively than noise at lower or higher frequencies. The contrast sensitivity function traced out by the faintest identifiable letters has a bandwidth similar to that of filters for grating detection.

A first step towards understanding an object recognition process is to determine which data are used by the process. Figure 2 shows how the high- and low-pass noise-masking paradigm measures the tuning of the filter mediating any threshold task. We measure threshold contrast energy (that is, integrated squared contrast) for letter identification as a function of the cut-off frequency of the noise. Data from two observers are shown in Fig. 3.

We express these results in terms of a minimal channel model: a linear filter followed by a nonlinear decision. It has been demonstrated that threshold signal energy E is linearly related to the power spectral density of a white noise mask[7–10] but what about non-white noise? If we assume that E is linearly related to the total noise power passed by the linear filter, then the frequency dependence of the filter gain can be derived directly from either our high- or our low-pass masking results[3,11–13]. More generally, this total-filtered noise-power assumption allows one to derive the tuning of the observer's filter for any task as a function of any dimension, such as spatial frequency, orientation or space, directly from appropriate noise-masked threshold measurements. The derived filters for letter identification are shown in Fig. 4.

A remarkable feature of these results is that the filters estimated from high- and low-pass data are so similar. By analogy with off-frequency listening[14], if an observer could choose one of many filters, then the ideal strategy would be to look off-frequency, choosing the filter that yields the best signal-to-noise ratio[11]. The observer has every incentive to avoid low-pass noise by choosing a high-frequency filter, and to avoid the high-pass noise by using a low-frequency filter. Yet the filters revealed by high- and low-pass masking are both centred at spatial frequen-

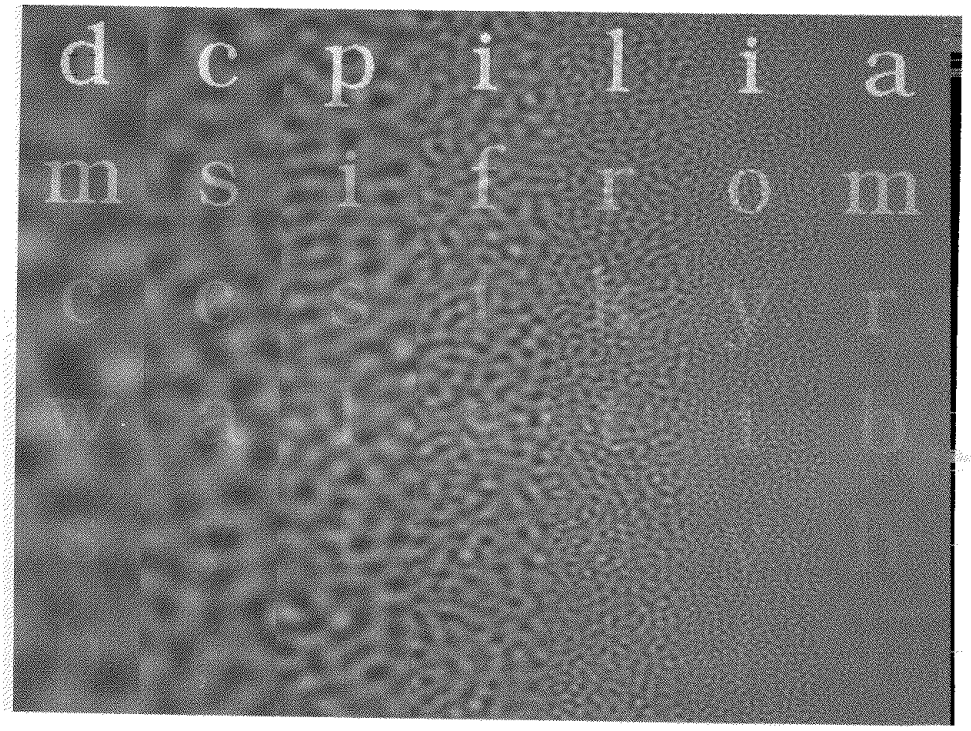

FIG. 1 Letters in noise, demonstrating that, despite the broad $1/f^2$ spectra of letters, noise at 3 cycles per letter (middle column) masks letter identification more effectively than noise at lower or higher frequencies. Read down each column as far as you can; the positions of the faintest identifiable letters trace out the sensitivity of your eye's letter-identification channel as a function of spatial frequency. Note that changes in viewing distance, from 3 to 60 cm, hardly affect the visibility of any given letter, indicating that the channel scales with letter size. (Identification of letters in noise is independent of size for letters subtending 0.5 to 15.8 degrees of visual angle[22].) Letter contrast (luminance increment divided by background luminance) decreases by factors of ×0.588 from 0.89 in the top row to 0.063 in the bottom row. The noise contrast c_{rms} is 0.22. Centre frequency (cycles per letter) of the noise's octave-wide pass band is: 0.8, d, m, c, w, s, h; 1.3, c, s, c, a, y, o; 2.0, p, i, s, l, f, t; 3.2, i, f, i, s, r, w; 5.1, l, r, k, t, o, s; 8.1, i, o, y, i, y, f; 13, a, m, r, b, h, t.

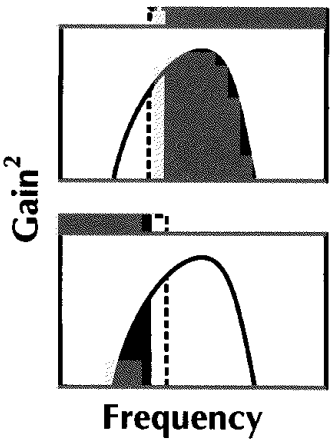

FIG. 2 The high- and low-pass noise masking paradigm. The curves represent the power gain (that is, gain squared) of the observer's hypothetical filter. The bar along the top of each graph represents the noise spectrum. Increasing the bandwidth of the noise adds noise in a new band (outlined by dashes). The effect of the noise in that band can be gauged by the resultant change in threshold, and independent estimates are provided by the high- and low-pass experiments. If we assume that the threshold energy is proportional to total filtered noise power, which is the black area under the filter's power gain curve, then the change in threshold specifies the filter's gain in the new band.

cies of ~3 cycles per letter, indicating that higher- and lower-frequency filters must be far less efficient or incapable of mediating letter identification.

The bandpass nature of these filters is characteristic of spatial frequency channels, whose one-to-two octave-wide tuning functions (at half height) have classically been estimated by grating summation[15], masking[12,16-18] and identification at threshold[19]. For a direct comparison, we applied our high- and low-pass noise-masking paradigm to a grating identification task, measuring the filter used by an observer to identify the orientation of a grating that was tilted right or left by 45°. This task is known to produce thresholds similar to those obtained in simple detection experiments[20]. Although the $\sim 1/f^2$ letter spectrum is utterly different from the narrow-grating spectrum, Fig. 4*c* and *d* shows that letter and grating identification are mediated by very similar filters, and off-frequency looking is no greater with the broad-band letters than it is with narrow-band gratings. The simplest explanation of this result is that one channel mediates and constrains object recognition; that is, the filter reflects the architecture of the visual system.

Of course, at least in principle, the (broad) average letter spectrum is irrelevant to letter identification; what matters is the differences between letters, and no single spectrum can summarize the many pairwise differences. Might the letter-derived filter merely reflect the demands of the letter-identification task, rather than a constraint of human vision? To determine what filter would reflect the task, we implemented the ideal classifier for letters in white noise, and applied our masking paradigm to it. The ideal classifier maximizes the probability of a correct response by choosing the hypothesis (letter identity) with maximum a posteriori probability. The white-noise ideal classifier chooses the letter that minimizes total squared difference between letter and stimulus, normalized by twice the noise variance, minus the letter's log prior probability[21]. The white-noise ideal classifier received the same information, as numbers, that was presented to the human observers as visual stimuli. The white-noise ideal classifier is ideal only when the noise is white.

The ratio of the ideal's threshold energy to that of a particular observer defines the observer's efficiency at that task. Observer L.L.'s and J.S.'s efficiencies at high levels of white noise, when $E^* \gg E_0$, are 13% and 10% for letter identification, and 6% and 8% for grating identification. Higher efficiencies have been reported: up to 42% for letter identification[22] and up to 70% for grating discrimination[7]. However, in the former case, stimuli were band-pass-filtered, which may boost efficiency[23], and in the latter, the task was discrimination of near-threshold contrasts of otherwise identical gratings, which increases efficiency as much as 9-fold over detection[24,25].

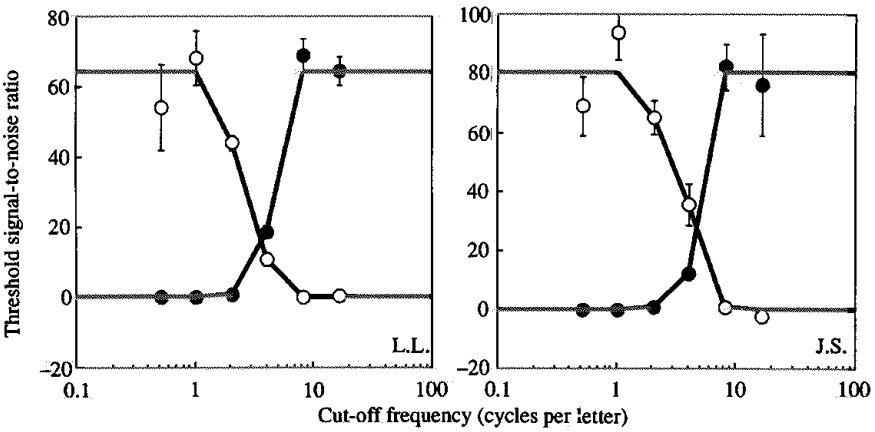

FIG. 3 Threshold signal-to-noise ratio, E^*/N, for letter identification as a function of cut-off frequency of high pass (open symbols) or low pass (filled symbols) noise, where N is the spectral density of the noise mask and $E^* = E - E_0$ denotes the difference in threshold contrast energy for identification between noise-masked and non-masked letters. Letter size is specified by the font's x height, which was 0.95°. As rendered in the Adobe Bookman font, 14 of the 26 lower-case letters in the English alphabet—acemnorsuvwxz—have the same height (top to bottom) as an x, and the rest—bfghijklpqty—are higher. When occurring at English language frequency (that is, many e and few j characters)[28], as in our experiments, the average letter height and width are 1.13 and 1.08 x heights, respectively. Letters were presented randomly, for 195 ms, at their English language frequencies. As shown in Fig. 2, the noise, which was static and isotropic, had power spectral densities of zero and N on either side of the cut-off frequency. All stimuli were displayed with gamma correction[29] on a monochrome CRT with a background luminance of 71 cd m^{-2}. The VideoToolbox software[29] used in these experiments is available by anonymous ftp of 'info-mac/dev/src/video-toolbox-. . .' from sumex-aim.stanford.edu. The viewing distance was 120 cm. The spatial extent of each noise mask was 3.9° × 3.9°. Each plotted symbol represents a maximum likelihood estimate of threshold (63% correct) based on at least 4 sessions of 30 trials each[30]. Standard error of each threshold measurement is indicated. Subsequent analysis of each observer's thresholds used maximum-likelihood, monotonic fits (solid curves) constrained to be positive with equal white-noise asymptotes.

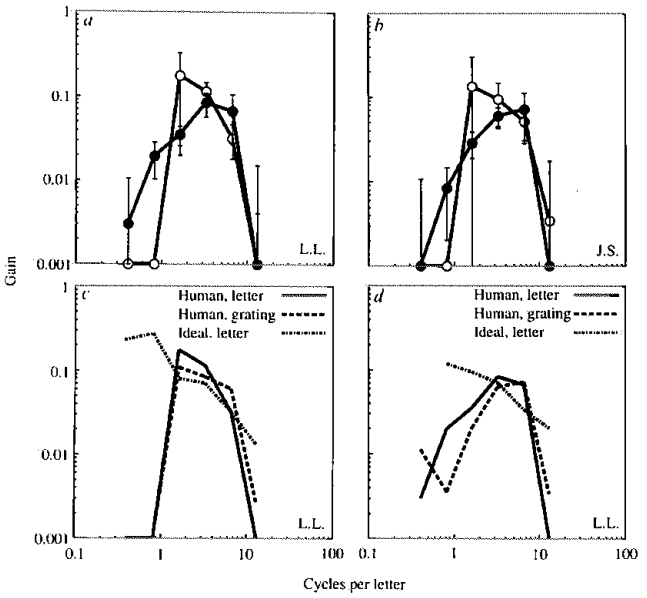

FIG. 4 *a* and *b*, Letter-identification filters for two subjects L.L. and J.S., as derived from Fig. 3. Error bars contain the 68% confidence interval about each plotted gain, as determined below. Zero gain is indicated by plotting the symbol on the horizontal axis. Filters plotted in *a* are replotted in *c* (high pass) and *d* (low pass) for comparison with similarly derived filters for grating identification by the same observer and letter identification by the white-noise ideal. (The other observer, J.S., produced similar results, which are not shown.) Each sinusoidal luminance grating had a spatial frequency of 3 cycles per letter, or 3.2 cycles per degree, tilted 45° to the left or right, and its contrast was windowed by a spatial gaussian envelope with a space constant of one grating period, centred midway between a peak and trough. Filter derivation: We assume that the threshold E is linearly related to the total noise power passed by some filter $G(f, \theta)$,

$$E = E_0 + a\int_0^{\infty} df \int_0^{2\pi} d\theta f G^2(f, \theta) N(f, \theta) \quad (1)$$

where E_0 is the threshold measured without display noise, a is a proportionality constant, f is radial spatial frequency, θ is orientation, $G^2(f, \theta)$ is the filter's power gain, and $N(f, \theta)$ is the power spectral density of the displayed noise. (This is equivalent to assuming a constant signal-to-noise ratio at the output of the filter at threshold.) E_0 may be due to intrinsic visual noise[8]; we subtract if from all of our data, leaving just the threshold elevation $E^* = E - E_0$ due to the displayed noise. The low- and high-pass noise spectra are

$$N_{low}(f, \theta) = \begin{cases} N & \text{for } f < f_c \\ 0 & \text{for } f \geq f_c \end{cases}$$

$$N_{high}(f, \theta) = \begin{cases} 0 & \text{for } f < f_c \\ N & \text{for } f \geq f_c \end{cases}$$

where f_c is the cut-off frequency. Substituting each of these noise spectra into equation (1), differentiating with respect to cut-off frequency, and solving for the filter's average power gain across all orientations yields

$$G^2(f) = \frac{1}{2\pi a f N}\frac{dE^*_{low}}{df} \quad (2)$$

and

$$G^2(f) = \frac{-1}{2\pi a f N}\frac{dE^*_{high}}{df} \quad (3)$$

where E^*_{low} and E^*_{high} are the threshold elevations produced by low- and high-pass noise, as a function of cut-off frequency. Comparisons among filters are facilitated by normalizing the gain so that all the filters pass equal power from a white noise input. This is achieved by setting the equation (1) proportionality constant to $a = E^*_{all}/N$, where E^*_{all} is the threshold elevation in white 'all pass' noise with power spectral density N. In our experiments we only used discrete values of cut-off frequency, f_i, $f_{i+1}, \ldots$, so we use discrete approximations to equations (2) and (3),

$$G^2\left(\sqrt{\frac{f_{i+1}^2 + f_i^2}{2}}\right) \approx \frac{1}{\pi E^*_{all}}\frac{E^*_{low}(f_{i+1}) - E^*_{low}(f_i)}{f_{i+1}^2 - f_i^2}$$

and

$$G^2\left(\sqrt{\frac{f_{i+1}^2 + f_i^2}{2}}\right) \approx \frac{-1}{\pi E^*_{all}}\frac{E^*_{high}(f_{i+1}) - E^*_{high}(f_i)}{f_{i+1}^2 - f_i^2}$$

which are exact if $G^2(f)$ is constant over the interval f_i to f_{i+1}.

Figure 4*c* and *d* shows that the white-noise-ideal-derived filter—which is a pure reflection of the demands of the task—is low-pass, unlike the human-derived filters, which are bandpass. The high gain of the ideal filter at low frequencies indicates the presence of information relevant to letter identification at these frequencies. The rapid fall-off of the gain of the human filters at frequencies below 1.5 and above 6 cycles per letter indicates that the human identifying letters is insensitive outside this two-octave band. The high-frequency fall-off parallels the finding that adding noise between 10 and 40 cycles per picture height is much more effective in suppressing face recognition than adding noise between 40 and 70 cycles per picture height[26]. The low-frequency fall-off confirms the human observers' inability to identify severely low-passed noisy letters, which are still identifiable by an ideal observer[22], and severely impaired character legibility[27] and reading rate[5] for text optically low-pass filtered to less than 2 cycles per character width. Thus the channel mediating letter identification is very different from what one would expect solely from a consideration of the task, and instead represents a visual constraint upon object recognition. □

Received 9 February; accepted 3 May 1994.

1. Fletcher, H. *Rev. Mod. Phys.* **12,** 47–65 (1940).
2. Campbell, F. W. & Robson, J. G. *J. Physiol.* **197,** 551–556 (1968).
3. Patterson, R. D. *J. Acoust. Soc. Am.* **55,** 802–809 (1974).
4. Graham, N. V. S. *Visual Pattern Analyzers* (Oxford University Press, Oxford, 1989).
5. Legge, G. E., Pelli, D. G., Rubin, G. S. & Schleske, M. M. *Vision Res.* **25,** 239–252 (1985).
6. Farah, M. J. *Visual Agnosia: Disorders of Object Recognition and What They Tell Us About Normal Vision* (MIT Press, Cambridge, MA, 1990).
7. Burgess, A. E., Wagner, R. F., Jennings, R. J. & Barlow, H. B. *Science* **214,** 93–94 (1981).
8. Pelli, D. G. thesis, Univ. Cambridge (1981).
9. Pelli, D. G. in *Vision: Coding and Efficiency* (ed. Blakemore, C.) 3–24 (Cambridge University Press, 1990).
10. Pavel, M., Sperling, G., Riedl, T. & Vanderbeek, A. *J. Opt. Soc. Am.* A**4,** 2355–2365 (1987).
11. Pelli, D. G. *Invest. Ophthal. Vis. Sci.* suppl. **19,** 44A (1980).
12. Henning, G. B., Hertz, B. G. & Hinton, J. L. *J. Opt. Soc. Am.* **71,** 574–581 (1981).
13. Pelli, D. G. & Watson, A. B. *Perception* **12,** A6–A7 (1983)
14. Patterson, R. D. & Nimmo-Smith, I. *J. Acoust. Soc. Am.* **67,** 229–245 (1980).
15. Graham, N. & Robson, J. G. *Vision Res.* **27,** 1997–2007 (1987).
16. Pantle, A. & Sekuler, R. *Science* **162,** 1146–1148 (1968).
17. Stromeyer, C. F. & Julesz, B. *J. Opt. Soc. Am.* **62,** 1221–1232 (1972).
18. Wilson, H R., McFarlane, D. K. & Phillips, G. C. *Vis. Res.* **23,** 873–882 (1983).
19. Watson, A. B. & Robson, J. G. *Vis. Res.* **21,** 1115–1122 (1981).
20. Thomas, J. P. & Gille, J. *J. Opt. Soc. Am.* **69,** 652–660 (1979).
21. Van Trees, H. L. *Detection, Estimation and Modulation Theory* (Wiley, New York, 1968).
22. Parish, D. H. & Sperling, G. *Vision Res.* **31,** 1399–1416 (1991).
23. Pelli, D. G., Jacknowski, M. M. & Hoepner, J. A. *Noninvasive Assessment of the Visual System, 1990 Tech. Digest Ser.* **3,** 144–146 (Opt. Soc. Am., Washington DC, 1990).
24. Nachmias, J. & Sansbury, R. V. *Vis. Res.* **14,** 1039–1042 (1974).
25. Barlow, H. B. *J. Physiol.* **160,** 155–168 (1962).
26. Harmon, L. D. & Julesz, B. *Science* **180,** 1194–1197 (1973).
27. Ginsburg, A. P. *Proc. of the Society for Information Display* **21,** 219–227 (1980).
28. Kucera, H. & Francis, W. N. *Computational Analysis of Present-Day American English* (Brown University Press, Providence, RI, 1967).
29. Pelli, D. G. & Zhang, L. *Vis. Res.* **31,** 1337–1350 (1991).
30. Watson, A. B. & Pelli, D. G. *Percept. Psychophys.* **33,** 113–120 (1983).

ACKNOWLEDGEMENTS. We thank B. Farell and G. Sperling for refining these ideas, and D. Balmori, R. B. Barlow, A. Bradley, C. W. Burns, S. Codner, S. Khan, W. Makous, C. Pelli, M. Raghavan, J. Robson, G. S. Rubin and D. Solomon for clarifying the presentation. Supported by a National Eye Institute grant to D.G.P.