PERGAMON

Available online at www.sciencedirect.com
SCIENCE DIRECT®

Vision Research 43 (2003) 1507–1511

Vision Research
www.elsevier.com/locate/visres

Rapid communication

Spatial-frequency spectra of printed characters and human visual perception

Endel Põder *

Laboratory of Cognitive Neuroscience, Tallinn Pedagogical University, 25 Narva Road, 10120, Tallinn, Estonia

Received 23 July 2002; received in revised form 25 March 2003

Abstract

It is well known that certain spatial frequency (SF) bands are more important than others for character recognition. Solomon and Pelli [Nature 369 (1994) 395–397] have concluded that human pattern recognition mechanism is able to use only a narrow band from available SF spectrum of letters. However, the SF spectra of letters themselves have not been studied carefully. Here I report the results of an analysis of SF spectra of printed characters and discuss their relationship to the observed band-pass nature of letter recognition.

Keywords: Spatial-frequency channels; Character recognition; Noise masking

1. Introduction

Many studies have found that certain spatial frequency (SF) bands are more important than others for character recognition (Alexander, Xie, & Derlacki, 1994; Legge, Pelli, Rubin, & Schleske, 1985; Parish & Sperling, 1991). Most impressively it has been demonstrated with critical-band noise masking by Solomon and Pelli (1994). These authors presented letters in either low-pass (LP) or high-pass (HP) filtered noise with different cut-off frequencies. The masking effect of noise was observed in a quite limited range of SFs around 3 cycles per letter height for both HP and LP noise. It was concluded that human pattern recognition mechanism is able to use only a narrow band from available SF spectrum. Tuning of the mechanism seemed to scale approximately with letter size (3 cycles per letter, independent of letter size).

More recent studies have revealed that exact proportionality with letter size does not hold (Chung, Legge, & Tjan, 2002; Majaj, Pelli, Kurshan, & Palomares, 2002). Majaj et al. (2002) used many different fonts and letter sizes and found that, in logarithmic coordinates, peak frequency of "letter channel" increases with slope 2/3 as dependent on letter "stroke frequency" (defined as average number of lines through letter slice divided by the letter width), or decreases with slope 2/3 as dependent on letter size for any fixed font. The bandwidth of revealed letter channel was approximately 1.5–1.6 octaves (full bandwidth at half of peak power gain), regardless of fonts, letter sizes, and task (identification or detection).

Recently, Tjan, Chung, and Legge (2002) have questioned the reality of "letter channels" in human vision, and argued that previous findings can be explained by a combination of the human contrast sensitivity function (CSF) across SFs and the distribution of letter identity information across SFs.

In the present study, I report the results of an analysis of spatial frequency spectra of printed characters and discuss the possible consequences of these results for the controversial issue of letter channels. It is well known that spectra of letters are broadband, consisting of all SFs from zero to the acuity limit. But surprisingly, more detailed information seems to have been absent. Both Majaj et al. (2002) and Solomon and Pelli (1994) mentioned about approximately $1/f$ fall-off of the spectra, which is usual for natural images, but without any reference to actual data.

The widely accepted standard model of human low-level vision consists of an array of mechanisms (channels) with more or less constant spatial frequency bandwidth (1–2 octaves) across a range of SFs (Blakemore & Campbell, 1969; Watson, 1983). Such architecture has

* Tel.: +372-662-3749.
E-mail address: ep@tpu.ee (E. Põder).

0042-6989/03/$ - see front matter
doi:10.1016/S0042-6989(03)00216-5

been found well adapted to the properties of natural images (Field, 1987). Here I attempt to characterise the spectra of letters also in a similar (octave) format.

2. Methods

At first, I analysed the SF spectra of printed numerals (0–9, Arial font, image size 256×256 pixels, letter height 96 pixels). An image of a character was Fourier transformed, the two-dimensional frequency plane was divided into the concentric octave-wide rings, and squared amplitude of Fourier components was integrated within each of them (Fig. 1a–c). This is equivalent to filtering an image with a set of octave-wide filters and calculating contrast energy (sum of squared deviations of pixel luminance values from the mean luminance) for each filtered image.

With the same methods, 26 lowercase letters of Bookman Old Style font (similar to Bookman font used in Solomon & Pelli (1994) and Majaj et al. (2002)) were analysed.

3. Results and discussion

The results depicted in Fig. 1d and e show that the spectra of printed characters are very different from these of "natural" images (the last have been found to have nearly equal energy across octave-wide bands (e.g. Field, 1987). The energy maximum is at 2–3 cycles per letter height, close to the location of supposed "perceptual filter" of Solomon & Pelli (1994). Average bandwidth (at half height energy, estimated with Gaussian approximation) is about 2.7 octaves for lovercase letters.

The results should not look too mysterious. Letters consist of strokes with rectangular luminance profile. The 1D spectrum of rectangle (in octave format) has maximum at 1/2 cycle per rectangle width. Further, many characters have some periodicity in their pattern (e.g. "*E*" can be seen as a grating with 2.5 cycles in vertical dimension). Both factors can contribute to the total spectrum. (Analysis with larger support in spatial domain that increases the resolution in low frequency octaves can reveal secondary mode near 0.5 cycles per letter height. However, this seems to be not very important for the following discussion).

The present results support the intuition of Majaj et al. (2002) that number of lines may be important measure for letter-like stimuli, and suggest more general and theoretically sound equivalent for their "stroke frequency"—peak (or median) of octave-scale energy spectrum.

The spectrum calculated here is very similar to the ideal letter sensitivity function (LSF) calculated by Chung et al. (2002). Chung et al. presented band-pass

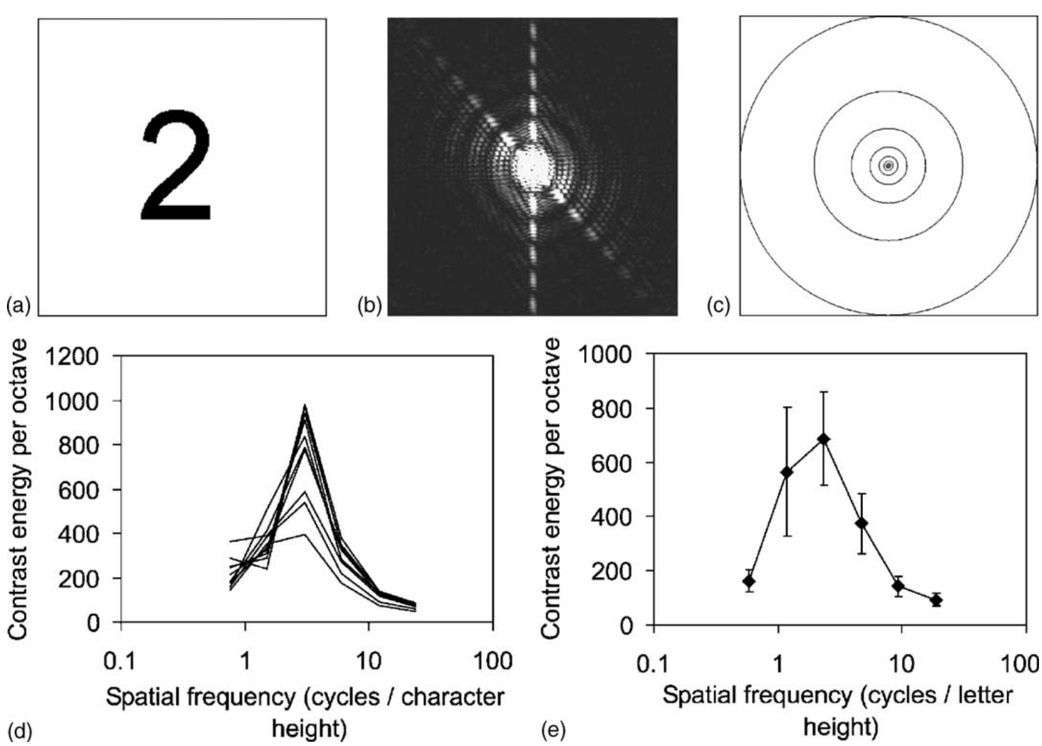

Fig. 1. An example image of a numeral used in the analysis (a), an image of its SF power spectrum (b), frequency plane divided into the octave-wide rings (c), and the distribution of contrast energy across octave-wide bands for 10 numerals (d) and for 26 lower-case letters (e). For lower-case letters, letter x height was used as measure of letter height. Error bars in the last graph indicate standard deviation.

filtered letters in white noise to the simulated ideal observer and calculated contrast sensitivity for letter-identification as dependent on filter centre frequency (contrast defined as contrast of unfiltered letter). For 26 lower-case Times-Roman letters, this curve is band-pass, with bandwidth approximately 4 octaves (at half height contrast sensitivity, that is 2.8 octaves in energy terms), and with peak at 2.1 cycles per letter height. LSF should measure the amount of letter-identity information across octave-wide SF bands. I suspect that LSF is in a large extent determined simply by contrast energy distribution across SF bands. If these functions play an important role in critical-band masking, then this similarity would explain why Majaj et al. (2002) have found very similar masking functions for detection and identification of letters.

The most interesting is the question about the nature of "letter channels". It is improbable that band-pass properties of letter spectrum reported here have nothing to do with channels observed in critical-band masking. But their exact relationship is not obvious.

Majaj et al. (2002) and Solomon & Pelli (1994) were surprised to find that observers seemed to use the same relatively narrow SF band for identification of broad-band letters in both LP and HP noise. An ideal observer would find SF bands outside of noise spectrum more useful, and consequently select different filters in different noise conditions. This argument seems to be valid even if letter spectrum is "not-so-broadband". Ideal observer, with no internal noise, with perfect SF resolution, and sensitive across broad SF spectrum, would use signal bands not covered by noise even if their energy is an infinitesimal part of masked signal energy. But what about more realistic observer with internal noise, SF resolution blurred by SF channels, and SF range constrained by CSF?

I used simple models of low-level vision to simulate the effect of LP and HP noise on the contrast threshold of signals with given SF (octave) spectrum (Appendix A). Similar models have been used by Perkins & Landy (1991) and Solomon (2000).

The basic model consists of a filter corresponding to human CSF and a dense array of SF channels with a Gaussian tuning function and constant bandwidth on the log frequency (octave) scale. Each channel transmits a part of input (signal + external noise) corresponding to (weighted by) its tuning function, and adds to the result a fixed amount of internal noise.

I considered three versions of combining information and calculation of thresholds: (1) best-channel model that selects the channel with highest signal-to-noise ratio (SNR); (2) summation model that integrates energy SNR across all SF channels; and (3) "matched filter" model that integrates noise energy by the filter matched to the signal energy distribution, regardless of SNR (and uses no pre-wired channels). Third model is nearly identical to the one suggested by Tjan et al. (2002).

Fig. 2 shows the typical results of critical-band masking experiments (following Majaj et al., 2002), and predictions of the three models with parameters supposedly close to these of representative human observers and experimental conditions (see Appendix A).

In spite of factors that make channel switching less advantageous, the best-channel and summation models

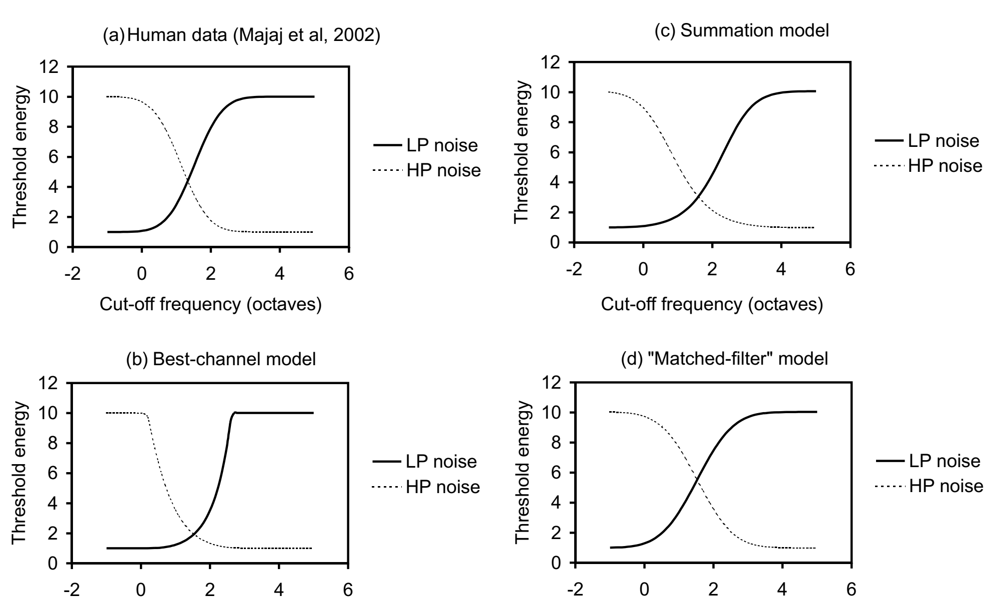

Fig. 2. Idealized results of critical-band masking experiments (Gaussian approximations according to Majaj et al., 2002) (a), and predictions of three models: best-channel model (b), summation model (c), and "matched filter" model (d).

still use quite different channels with LP and HP noises (e.g. difference about 1.8 octaves for best-channel and 1.2 octaves for summation model in conditions depicted in Fig. 2) that is not observed in experiments with human observers. Human data are better in accord with the third model that uses the same channel regardless of noise spectrum. This parallels with the claim that human observers are unable to pre-whiten visual noise (e.g. Myers, Barrett, Borgstrom, Patton, & Seeley, 1985). However, this inability does not need to be absolute. Majaj et al. (2002) report average shift about 0.5 octaves between filters measured with LP and HP noises, and Solomon & Pelli (1994) graphs exhibit some resemblance to my best-channel simulation: asymmetric masking curves rising more steeply in their noise-free side.

Majaj et al. (2002) and Solomon & Pelli (1994) assume that observed letter channel is one selected from classical SF channels. Chung et al. (2002) and Tjan et al. (2002), and the results of the present study suggest another basis for letter channel: the spectrum of object itself. One may ask which of them is correct. Interestingly, both ideas can be essentially correct.

Tjan et al. (2002) suggest that ideal LSF $\times$ human CSF model can account for critical-band-masking experiments, without any channel in human head. However, their model can be viewed a bit differently. It uses the channel that is matched to signal spectrum. It is ideal for white noise. In filtered noise, observer is assumed to use the same channel (that is not optimal now). This channel reflects the (useful) spectrum of object to perceive, but exists still as a filter in human head. Also, this channel is not necessarily exact matched filter but some approximation within constraints of brain mechanisms. At limit, it can be selected from available well-known SF channels.

The role of CSF is largely the same regardless of following pattern recognition mechanism. Multiplication by CSF shifts the centre frequency of any band-pass mechanism towards the peak of CSF and makes the observed bandwidth narrower (Chung et al., 2002). This can explain nicely the variance of letter channel frequency as dependent on letter size. But slope 2/3 found by Majaj et al. (2002), implies (at least for Gaussian tuning functions) ratio 1:2 for LSF/CSF bandwidths, while actual ratio is near 1:1. This may indicate that some additional factor is narrowing human letter channel. The results of Parish & Sperling (1991) point to the same direction. They show that human observers use information at lowest and highest object SF bands less efficiently as compared with centre bands, contrary to simple LSF $\times$ CSF model.

Thus, the spectra of printed characters can explain where the "letter channels" come from, but this explanation does not eliminate the constraints of human pattern recognition mechanism. Our visual system may have little ability to adapt to different spectra of external noise, and may be less efficient in using information far from the centre of object SF spectrum.

Appendix A. Modelling of critical-band masking

All calculations were done in logarithmic (octave) SF scale. All band-pass functions (human CSF, letter energy spectrum, and SF channels) were assumed to be Gaussian (parabolas in log–log coordinates).

Human CSF is approximated by

$$F(f) = 2^{-4(f-f_F)^2/b_F^2},$$

where f is SF in octaves (0 octave placed at 1 cpd), f_F—location of peak sensitivity, and b_F—bandwidth at half height.

Signal (letter) energy per octave

$$E_1(f) = 2^{-4(f-f_E)^2/b_E^2},$$

where f_E—location of energy peak, inversely proportional to letter size, and b_E—bandwidth of letter energy spectrum.

Spectral density for LP masking noise

$$N_1(f) = N, \quad \text{if } f < f_C$$
$$N_1(f) = 0, \quad \text{otherwise;}$$

and for HP noise

$$N_1(f) = N, \quad \text{if } f \geqslant f_C$$
$$N_1(f) = 0, \quad \text{otherwise,}$$

where f_C is cut-off frequency.

Both signal energy distribution and noise spectral density were multiplied by squared contrast sensitivity and convolved with SF channel tuning function, and constant inner noise N_i was added to the noise term.

$$E_2(f) = [E_1(f) \cdot F^2(f)] \otimes G^2(f),$$

$$N_2(f) = [N_1(f) \cdot F^2(f)] \otimes G^2(f) + N_i,$$

where $G(f) = 2^{-4(f-f_G)^2/b_G^2}$, with f_G—tuning SF, and b_G—bandwidth of SF channels.

While constant in octaves, linear bandwidths increase with SF. With white noise, the noise power transmitted by channels increases proportionally to f^2. But this is exactly compensated with increasing number of signal components. Noise term $N_2(f)$ can be interpreted as average spectral density within channel located at f.

Thresholds were calculated as follows.

$$T_1 = C_1 \cdot \min[N_2(f)/E_2(f)] \quad \text{(best-channel model)}$$

$$T_2 = C_2 \bigg/ \int E_2(f)/N_2(f)\,\mathrm{d}f \quad \text{(summation model)}$$

$$T_3 = C_3 \cdot \frac{\int E_1(f) \cdot [F^2(f) \cdot N_1(f) + N_i]\,\mathrm{d}f}{\int E_1(f) \cdot F^2(f)\,\mathrm{d}f}$$

("matched filter" model)

The following values of parameters were used for generating graphs in Fig. 2.

CSF: peak $f_F = 2$ oct (4cpd); bandwidth $b_F = 4$ oct (e.g. Chung et al., 2002; Watson, 2000).

Letter energy spectrum: peak $f_E = 1$ oct (corresponding the letter size 1 deg; bandwidth $b_E = 2.7$ oct (present study).

Bandwidth of SF channels $b_G = 1.5$ oct (e.g. De Valois, Albrecht, & Thorell, 1982).

Internal noise $N_i = 1$; constants C_1 –C_3 were adjusted to equalize thresholds without external noise; N was selected for representative tenfold rise of energy thresholds (e.g. Majaj et al., 2002) (values 11.3, 15.3 and 15.1 for models 1–3).

References

Alexander, K. R., Xie, W., & Derlacki, D. J. (1994). Spatial-frequency characteristics of letter-identification. *Journal of the Optical Society of America A, 11*, 2375–2382.

Blakemore, C., & Campbell, F. W. (1969). On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *Journal of Physiology, 203*, 237–260.

Chung, S. T. L., Legge, G. E., & Tjan, B. S. (2002). Spatial-frequency characteristics of letter identification in central and peripheral vision. *Vision Research, 42*, 2137–2152.

De Valois, R. L., Albrecht, D. G., & Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research, 22*, 545–559.

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A, 4*, 2379–2394.

Legge, G. E., Pelli, D. G., Rubin, G. S., & Schleske, M. M. (1985). Psychophysics of reading—I, Normal vision. *Vision Research, 25*, 239–252.

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.

Myers, K. J., Barrett, H. H., Borgstrom, M. C., Patton, D. D., & Seeley, G. W. (1985). Effect of noise correlation on detectability of disk signals in medical imaging. *Journal of the Optical Society of America A, 2*, 1752–1759.

Parish, D. H., & Sperling, G. (1991). Object spatial frequencies, retinal spatial frequencies, noise, and the efficiency of letter discrimination. *Vision Research., 31*, 1399–1415.

Perkins, M. E., & Landy, M. S. (1991). Nonadditivity of masking by narrow-band noises. *Vision Research, 31*, 1053–1065.

Solomon, J. A., & Pelli, D. G. (1994). The visual filter mediating letter identification. *Nature, 369*, 395–397.

Solomon, J. A. (2000). Channel selection with non-white-noise masks. *Journal of the Optical Society of America A, 17*, 986–993.

Tjan, B.S., Chung, S.T.L., & Legge, G.E. (2002). *O letter channels, where art thou?* Poster presented at Vision Sciences Society 2nd Annual Meeting, Sarasota, FL, May 10–15, 2002.

Watson, A. B. (1983). Detection and recognition of simple spatial forms. In O. J. Braddick, & A. C. Sleigh (Eds.), *Physical and biological processing of images* (pp. 110–114). Berlin: Springer-Verlag.

Watson, A. B. (2000). Visual detection of spatial contrast patterns: Evaluation of five simple models. *Optics Express, 6*, 12–33.