Spatial Vision, Vol. 17, No. 1-2, pp. 5–15 (2004)

Also available online - www.vsppub.com

The filtered Fourier difference spectrum predicts psychophysical letter discrimination in the peripheral retina

ROGER S. ANDERSON [1,*] and LARRY N. THIBOS [2]

[1] *Vision Science Research Group, School of Biomedical Sciences, University of Ulster at Coleraine, BT52 1SA, Northern Ireland, UK*
[2] *School of Optometry, Indiana University, Bloomington, IN 47405, USA*

Received 29 July 2002; revised 10 November 2002; accepted 11 November 2002

Abstract—We wished to devise a measure of dissimilarity (D) which could predict psychophysical discrimination performance for Snellen letter pairs in peripheral vision. Threshold size for discriminating 33 pairs of Snellen letters was measured at 30 degrees eccentricity in the nasal retina for two subjects. D was computed for each pair by performing an overlap subtraction in the spatial domain, followed by a Fast Fourier Transform on this difference image, and dividing the total power in the resultant 'difference spectrum' by the sum of the powers of the individual letter spectra. A plot of D *vs.* psychophysical threshold letter size gave a mean correlation of $R = -0.81$. When D was calculated for letters that were low-pass filtered at different cut-off frequencies, the correlation with psychophysical performance was greatest when cut-off was between 1.25–1.9 cycles/letter ($R = -0.85$). Conversely, when the difference spectrum was high-pass filtered at different cut-off frequencies, the correlation decreased continuously as the cut-off increased. These results imply that the band of frequencies between zero and 1.25 cycles/letter are most important for letter discrimination in peripheral vision.

Keywords: Visual acuity; letters; peripheral vision; Fourier analysis.

INTRODUCTION

One of the goals of vision research is to develop computational models that predict visual performance on various tasks. One such task is letter acuity, for which a variety of models have been developed for foveal vision (Ginsburg, 1980; Legge *et al.*, 1985; Parish and Sperling, 1991; Solomon and Pelli, 1994). However, models of letter acuity for peripheral vision are underdeveloped by comparison. This despite the fact that peripheral viewing is sometimes employed by people with

*To whom correspondence should be addressed. E-mail: rs.anderson@ulster.ac.uk

visual impairment. Two possible reasons for this state of affairs is that (1) peripheral vision has not been studied as much as foveal vision, and (2) the added complication of neural undersampling of the peripheral retinal image which cannot be neglected when modeling peripheral spatial vision (Thibos *et al.*, 1987; Anderson and Thibos, 1999). Gervais *et al.* (1984) examined three methods for predicting psychophysical similarity of uppercase letters in foveal vision. These included models based on template overlap, geometric features and spatial frequency content. They concluded that this latter model gave the best correlation with psychophysically determined confusion matrices for different letter pairs ($r = 0.70$). Their model computed the Fourier Transform of each letter and multiplied the amplitude spectrum by the human contrast sensitivity function. They then used this spectrum to calculate a measure of dissimilarity (D) between letter pairs using the equation:

$$D_{\mathrm{AB}} = \left\{ \sum_{ij} \left[\left(\log X_{ij\mathrm{A}} - \log X_{ij\mathrm{B}} \right) * \left(1 + Y_{ij\mathrm{A}} - Y_{ij\mathrm{B}} \right) \right]^2 \right\}^{0.5}, \tag{1}$$

where $X_{ij\mathrm{A}}$ is the amplitude of letter A at spatial frequency i, j, in the filtered amplitude spectrum, $X_{ij\mathrm{B}}$ is the corresponding amplitude of letter B, $Y_{ij\mathrm{A}}$ is the phase of letter A at cell i, j, in the phase spectrum, and $Y_{ij\mathrm{B}}$ is the corresponding phase of letter B.

This model produced good correlation with psychophysical measurements of similarity, but has been suggested to have a significant flaw (Higgins *et al.*, 1996). Since it subtracts corresponding amplitudes before taking account of phase, any letter pair that has the same amplitude spectrum and differs only in phase (i.e. any mirror symmetric pair (e.g. b and d) should be indistinguishable by this predictor. Vol and Pavlovskaya (1990), in commenting on this model, also concluded that it could not be realised by known neurophysiological mechanisms. Instead, they used the following equation to calculate 'Euclidian distance' between various visual objects:

$$S_{jk}^2 = \int\int \left\{ \mathrm{Re}^2[J(u,v,) - K(u,v,)]^2 + \mathrm{Im}^2[J(u,v,) - K(u,v,)] \right\} \mathrm{d}u\, \mathrm{d}v, \tag{2}$$

where S_{jk} is the distance between the jth and kth objects, J and K are their Fourier transforms, and Re and Im are the real and imaginary parts of the complex spectrum. The Euclidean distances defined by equation (2) were well correlated with psychophysical discrimination performance and, they believed, was an improvement on the model of Gervais *et al.* because it could be realised by the visual receptive fields. However, this model weights all spatial frequencies equally in the calculation of Euclidian distance without considering that some spatial frequencies might contribute more to visual performance than others.

Our present aim was to extend this earlier work into peripheral vision, but restricting the model to just those frequencies which are known to be important for letter resolution (Ginsburg, 1980; Legge *et al.*, 1985; Parish and Sperling, 1991; Solomon and Pelli, 1994) at the endpoint of an acuity paradigm.

METHODS

Psychophysical

To select appropriate letter pairs for the discrimination task, an initial computerised analysis was performed on an alphabet of 26 uppercase letters. From all possible combinations, 33 pairs were selected using the criterion that the total signal power in the two letters differed by less than 10%. This was done in order to preclude psychophysical discrimination simply on the basis of the difference in mean luminance, and meant that three letters of the uppercase alphabet (B, I, J) could not be paired with any other and therefore were not used at all. We measured monocular discrimination size thresholds for each of these 33 letter pairs at 30 degrees eccentricity in the horizontal temporal field (nasal retina) of two experienced subjects (the authors). These subjects (both near emmetropes foveally) were refracted at the location being tested using retinoscopy and the appropriate correction was placed in front of the eye in line with the peripheral target which was displayed on a computer monitor at a 3 m viewing distance. Peripheral refractive error was thus corrected at all times. This distance, and the fact that peripheral acuity is significantly lower than in the fovea, meant that letters remained sufficiently large to minimise pixelation effects on the monitor at all times. To fix gaze and accommodation, the subject fixated a high contrast illuminated cross at a distance of 3 m (fixation axis did not pass through the peripheral correcting lens).

We used a 2AFC staircase procedure to measure minimum angle of resolution (MAR). For each trial, one or other of the pair of letters in question was presented and the subject's task was to indicate which of the pair was present. The initial presentation size was well above threshold so subjects were aware of which letters were being discriminated. A correct response three times in a row caused the letter size to be decremented. Any error in response caused an increment in size. Increment or decrement size was 25% for the first two reversals and 10% for a further seven. The first two reversals of the staircase were discarded and threshold letter size was determined as the average of the last seven reversal sizes (see Anderson and Thibos, 1999b for further method details).

Model

Using the same 23-letter alphabet as in the experimental study reported above, 23 different images were created on the computer, each consisting of a different black uppercase letter in the centre of a white background (1-bit pixel contrast). The background was square with each side 256 pixels long. The letter in the centre was 80 pixels in both height and width.

The same 33 pairs of these letter targets used in the psychophysical experiment were chosen for computer analysis as follows. First, we subtracted the image of one letter pair from the other (i.e. pixel by pixel subtraction) and so obtained a 'difference image' for that pair. A Fast Fourier Transform (FFT) was then performed on each difference image to produce a 'difference spectrum' in the

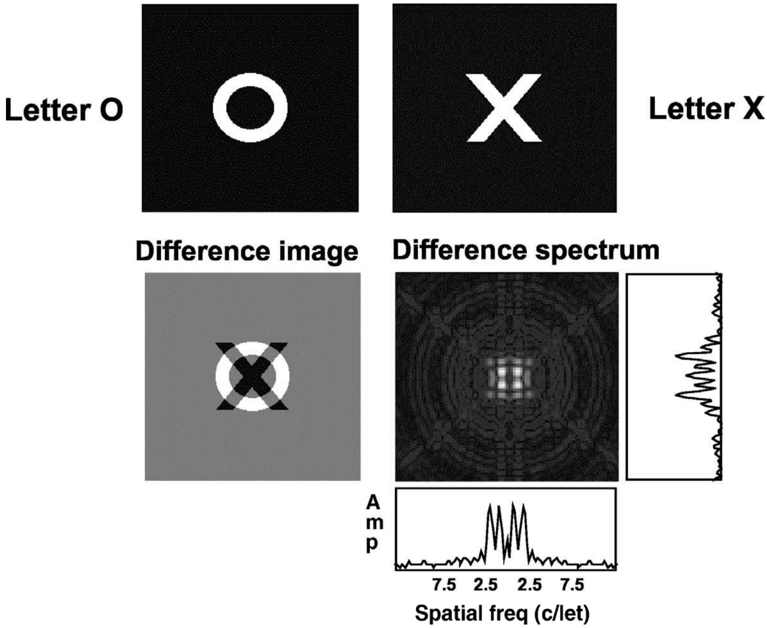

Figure 1. Difference image (in spatial domain) and difference amplitude spectrum (in frequency domain) of letters O and X.

frequency domain (see Fig. 1). The ‘dissimilarity’ of the two letters was calculated for each pair using the equation:

$$D = \frac{\{\sum_{ij} AB_{ij}^2\}^{0.5}}{\{\sum_{ij} A_{ij}^2\}^{0.5} + \{\sum_{ij} B_{ij}^2\}^{0.5}}, \tag{3}$$

where D is the dissimilarity of the letters, AB_{ij} is the complex amplitude of element ij in the Fourier difference spectrum of the letters A and B, A_{ij} is the complex amplitude of element ij in the Fourier spectrum of the letter A, and B_{ij} is the complex amplitude of element ij in the Fourier spectrum of the letter B. This equation is similar to that of Vol and Pavlovskaja (1990) in that it does not separate the phase component before calculating the power in the difference spectrum of the two letters. It differs from that of Vol and Pavlovskaya in that the power in the difference spectrum is normalized by the sum of the powers in the individual letter spectra. This means that D is a unitless metric which does not vary with the size of the letter pair.

Equation (3) uses all frequencies present in the spectrum to calculate power. To produce a family of models implementing the concept of a critical frequency band, we recalculated the dissimilarities (D) after filtering the difference spectra using low and high-pass ideal filters with cut-off frequencies of 2.5, 1.8, 1.6, 1.25, 0.9 and

0.6 cycles/letter using the equation:

$$D = \frac{\{\sum_{ij}(F_{ij} \cdot AB_{ij})^2\}^{0.5}}{\{\sum_{ij}(F_{ij} \cdot A_{ij})^2\}^{0.5} + \{\sum_{ij}(F_{ij} \cdot B_{ij})^2\}^{0.5}}, \tag{4}$$

where F_{ij} in this case is the amplitude of element ij of the low or high-pass filter with cutoff frequency C (the choice of values of C was limited by pixel quantisation).

For low-pass filtering: $F = 1$ for $i^2 + j^2 \leqslant C^2$,
$F = 0 \ldots$ if not

For high-pass filtering: $F = 0$ for $i^2 + j^2 \leqslant C^2$,
$F = 1 \ldots$ if not

This dissimilarity index (D) for each letter pair and for each of the low and high-pass filtered models was then plotted against the corresponding psychophysically measured threshold value for the same letter pair. The correlation coefficient (R) of the resultant scatter plot was taken as a measure of the ability of each model to predict psychophysical discriminability.

We note that D can be equivalently calculated in the spatial domain and is essentially a measure of the average spatial contrast contained in the difference image (see Appendix A). Our reasons for performing the calculation in the frequency domain were twofold: first, to more easily compare with previous studies which used the frequency domain, and second, to simplify the calculation of D for filtered letters (filtering was easier to perform in the frequency domain).

RESULTS

The premise of our experiments is that letters which are highly dissimilar, as quantified by the dissimilarity metric D, should be highly discriminable and therefore the minimum discriminable size will be relatively small. The results of a test of this hypothesis are displayed in Fig. 2 which indicates how threshold letter size (i.e. acuity) varies with dissimilarity D for unfiltered letter pairs for both subjects. The best fitting straight line has been drawn and the correlation coefficient (R) calculated ($p < 0.01$ for R for both subjects). Of the letter pairs tested, the most dissimilar was A *vs.* U, for which the average threshold letter size was 23 arcmin for our two subjects. By comparison, the least dissimilar pair was O *vs.* Q, for which the average threshold letter size was 92 arcmin. This result demonstrates that visual acuity for discriminating letter pairs can vary over a 4-fold range depending on the particular letter pair being discriminated.

To determine the importance of excluding some spatial frequency components when computing the dissimilarity metric D, the preceding analysis was repeated for the various filtering schemes described in Methods. For example, Fig. 3 is the same as Fig. 2 except that letter dissimilarity was computed for letter pairs that were high-pass filtered at 2.5 cycles/letter in order to exclude the lower spatial frequencies. The

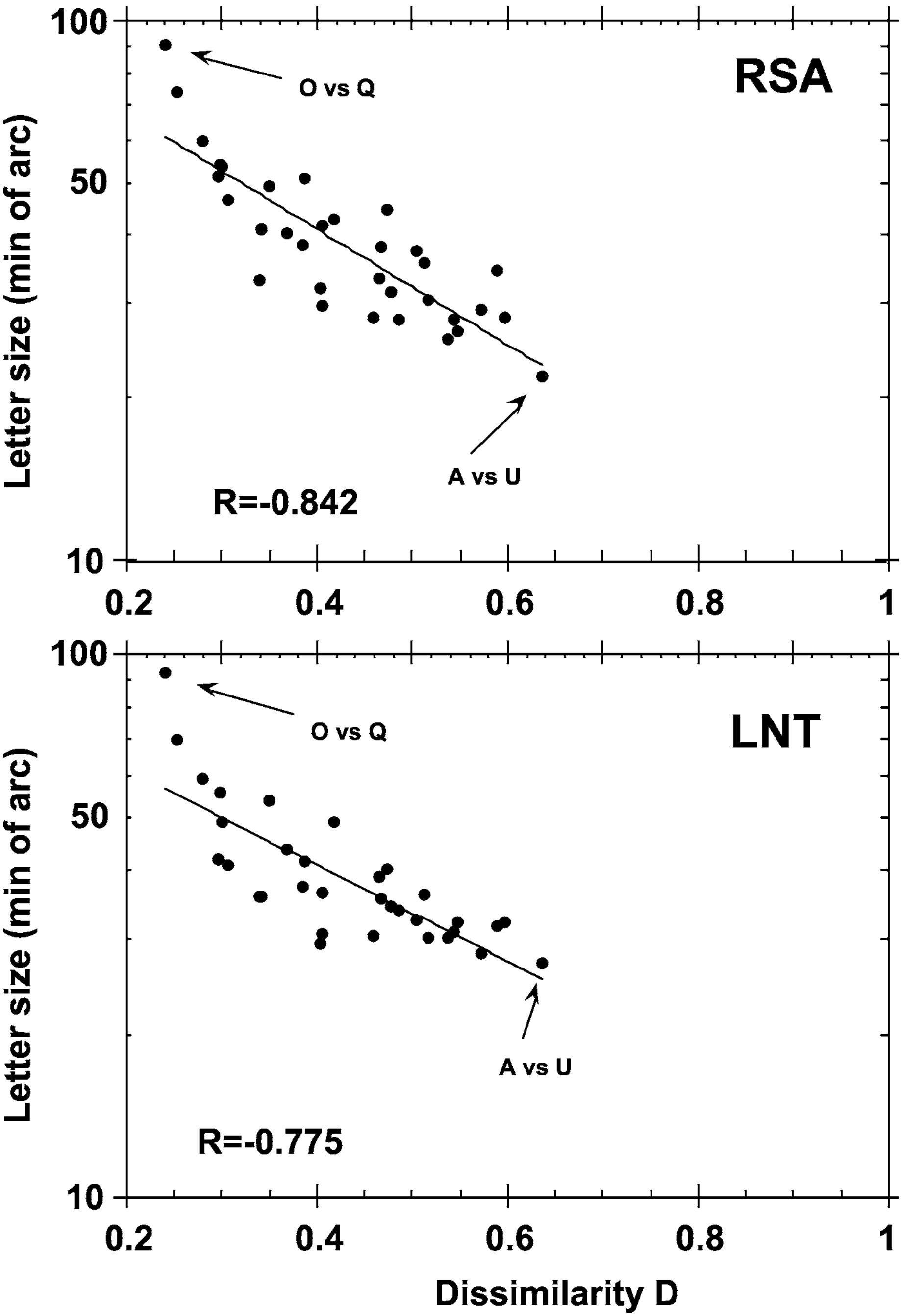

Figure 2. Threshold letter discrimination size *vs*. dissimilarity (D) for unfiltered letter pairs (subjects RSA and LNT).

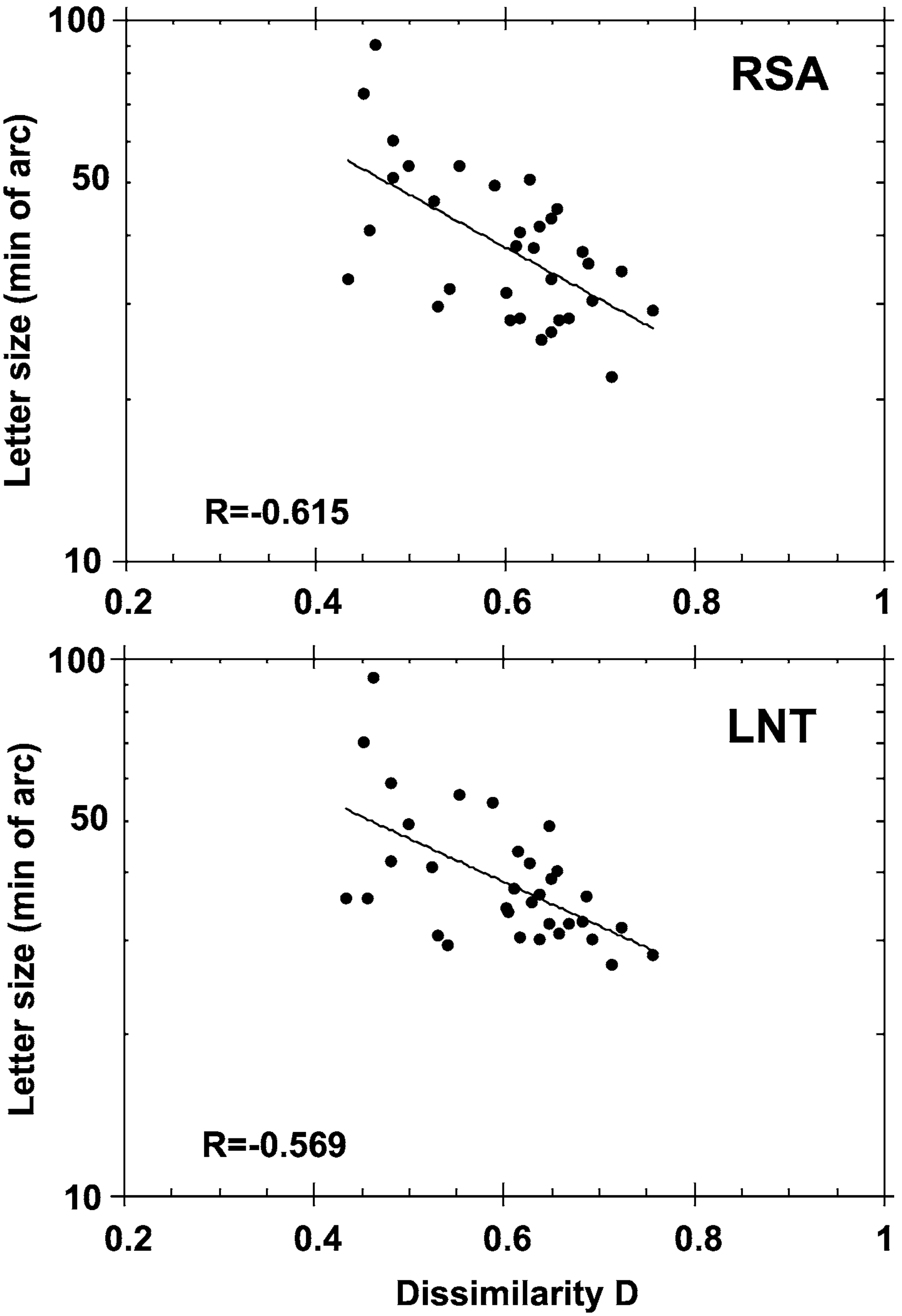

Figure 3. Threshold letter discrimination size *vs.* dissimilarity (D) for letter pairs high-pass filtered at 2.5 cycles/letter (subjects RSA and LNT).

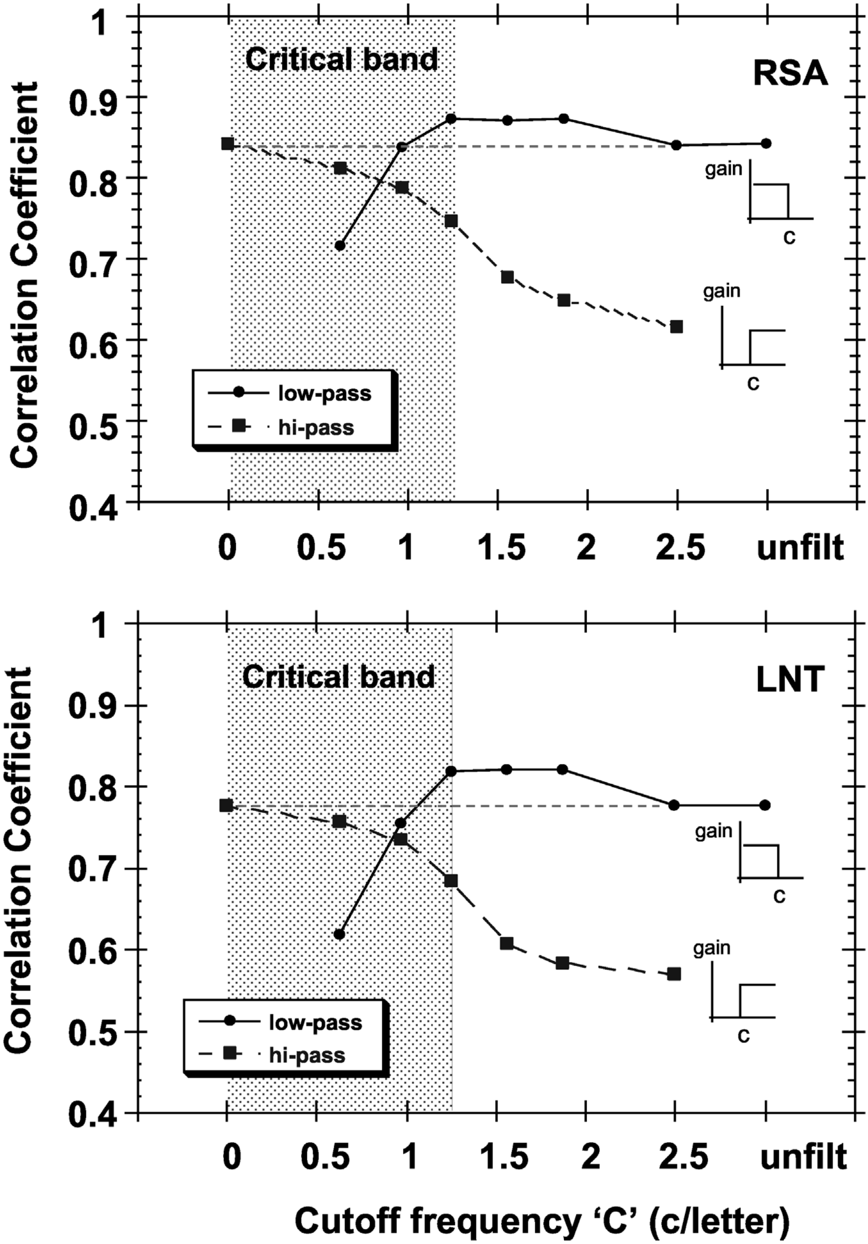

Figure 4. Plot showing how correlation of discrimination model with psychophysical performance changes with filter cut-off frequency for both low and high-pass filtered pairs. Filter characteristics displayed in schematic form beside relevant curve; C is cut-off (subjects RSA and LNT).

correlation coefficients in this case are now much lower ($p < 0.05$), indicating that the predictive power of D is less when only the high spatial frequencies are included in the computation.

Correlation as a function of cut-off spatial frequency of the model is plotted in Fig. 4 for both the low-pass and high-pass models for each subject. For the low-pass models, it can be seen that as the filter cut-off frequency decreases, the correlation increases slightly until the cut-off reaches 1.9 cycles/letter (peak $R = -0.871$ for subject RSA, -0.821 for subject LNT), after which correlation remains constant, and then deteriorates below 1.25 cycles/letter ($p < 0.01$ for all correlations).

For the high-pass models, correlation decreases continuously as the cut-off increases beyond zero cycles/letter for both subjects ($p < 0.05$ for all correlations) and reaches a plateau level at about 1.9 cycles/degree.

DISCUSSION

Our results demonstrate that the power in the unfiltered difference spectrum is a strong predictor of the minimum discriminable letter size with a correlation coefficient of -0.842 for subject RSA and -0.775 for subject LNT ($p < 0.01$ for R for both subjects). The fact that predictive power increased slightly if the higher spatial frequencies beyond 1.9 cycles/letter are removed by low-pass filtering, and remained steady until cut-off reached 1.25 cycles/letter implies that frequencies beyond 1.25 cycles/letter are not useful visually for discriminating letters at the endpoint of an acuity paradigm. Further removal of spatial frequency components below 1.25 cycles/letter begins to reduce the ability of the model to predict psychophysical performance, which implies that an upper bound on the critical band of frequencies for peripheral letter discrimination is in the range 0 to 1.25 cycles/letter. Although the slight improvement in correlation with low-pass filtering down to 1.9 cycles/letter is not significant, it should be noted that the pattern is identical for both subjects. For the high-pass models, correlation decreases continuously as the cut-off increases beyond zero cycles/letter for both subjects ($p < 0.05$ for all correlations). We interpret this to mean that even the very lowest spatial frequencies contribute significantly to performance and that the lower limit to the critical band is close to zero.

Taken together, these results indicate that letter discrimination performance at 30 degrees eccentricity in the nasal retina is well predicted by a stimulus-based model of letter discrimination using only the power in the band of frequencies between 0 and 1.25 cycles/letter (indicated by shaded areas in Fig. 4). This is slightly lower than our previous studies which indicated that the band of frequencies between 0.9–2.2 cycles/character was most important for discrimination of tumbling E stimuli in peripheral vision (Anderson and Thibos, 1999a, b). However, our results do not necessarily mean that all letter pairs are best discriminated using exactly the 0–1.25 cycles/letter band of frequencies since our assessment is a statistical summary across letter pairs and some of the scatter observed in the data may actually reflect real inter-pair variation in the critical band for discrimination. It is probable that the critical band of frequencies is slightly different for the tumbling-E letter set specifically compared to all letters generally.

We noted above in connection with Fig. 2 that threshold letter size for discrimination varies by as much as a factor of 4 depending on which letter pair is being discriminated. This represents a difference in size of six lines on a logMAR acuity chart. Some letter charts have been designed using letters that are supposedly equally discriminable, e.g. the Sloan set of letters. This is an attempt to make the variability in performance within a line less than that between lines. A method of determining letter discriminability more accurately for this purpose would be useful. It should also be noted, however, that our experiment was conducted only in the periphery and used only a two-alternative forced choice procedure. Increasing the number of choices, e.g. to 4AFC, would introduce more uncertainty and may change performance for individual letter recognition. This is a topic for further research.

REFERENCES

Anderson, R. S. and Thibos, L. N. (1999a). Relationship between acuity for gratings and for tumbling-E letters in peripheral vision, *J. Opt. Soc. Am. A* **16**, 2321–2333.

Anderson, R. S. and Thibos, L. N. (1999b). Sampling limits and critical bandwidth for letter discrimination in peripheral vision, *J. Opt. Soc. Am. A* **16**, 2334–2342.

Bracewell, R. N. (1978). *The Fourier Transform and its applications*. New York, McGraw-Hill.

Gervais, M. J., Harvey, L. O. and Roberts, J. O. (1984). Identification confusions among letters of the alphabet, *J. Exper. Psych.: Hum. Percep. Perform.* **10** (5), 655–666.

Ginsburg, A. P. (1980). Specifying relevant spatial information for image evaluation and display designs: an explanation of how we see certain objects, *Proc. SID.* **21**, 219–227.

Higgins, K. E., Arditi, A. and Knoblauch, K. (1996). Detection and identification or mirror-image letter pairs in central and peripheral vision, *Vision Res.* **36**, 331–337.

Legge, G. E., Pelli, D. G., Rubin, G. S. and Schleske, M. M. (1985). Psychophysics of reading-1. Normal vision, *Vision Res.* **25**, 239–252.

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

Peli, E. (1990). Contrast in complex images, *J. Opt. Soc. Am. A* **7**, 2032–2040.

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

Thibos, L. N., Cheney, F. E. and Walsh, D. J. (1987). Retinal limits to the detection and resolution of gratings, *J. Opt. Soc. Am. A* **4**, 1524–1529.

Vol, I. A. and Pavlovskaja, M. B. (1990). Similarity between Fourier transforms of objects predicts their experimental confusions, *Percept. Psychophys.* **47**, 12–21.

APPENDIX: ENERGY CONSERVATION BETWEEN SPATIAL AND FREQUENCY DOMAIN

By Parseval's theorem (Bracewell, 1978):

$$\sum_{x,y} m_{xy}^2 = N\left(L_0^2 + 0.5\sum_{i,j} X_{ij}^2\right), \tag{A1}$$

where m_{xy} is the intensity of pixel xy in the spatial domain, N is the total number of pixels, L_0 is the amplitude of the d.c. component and X_{ij} is the amplitude of spectral elements ij in the frequency domain. In our case, the quantity on the left involves the sum of the squared pixel intensities in the *difference image* and the quantity on the right involves the sum of the squared pixel intensities in the *difference spectrum*. By this interpretation, Parseval's theorem is a statement of energy conservation in that, assuming pixel intensity is linear, the signal contains a given amount of energy regardless of whether the energy is computed in the spatial or frequency domain.

Dividing both sides of equation (A1) by N, the left side becomes the mean energy per pixel, and the right side becomes the total power in the spectrum. Then, rearranging terms gives

$$\frac{1}{N}\sum_{x,y} m_{xy}^2 - L_0^2 = 0.5 \sum_{i,j} X_{ij}^2. \qquad \text{(A2)}$$

The left hand side is recognized as the variance of the pixel luminance values in the spatial domain. Therefore, the variance of the pixel luminance values in the difference image equals the total amount of power in the harmonic components of the difference spectrum (excluding the d.c. component).

Since the standard deviation of the pixel values may be interpreted as a measure of the average spatial contrast in an image (Peli, 1990), this result demonstrates that our measure of dissimilarity may also be considered a measure of contrast in the difference image.