Watson, A. B. & Fitzhugh, A. E. (1989). Modelling character legibility. *Society for Information Display Digest of Technical Papers 20*, 360-363.

MODELLING CHARACTER LEGIBILITY

Andrew B. Watson
NASA Ames Research Center, Moffett Field, CA
beau@vision.arc.nasa.gov

Andrew E. Fitzhugh
Hewlett-Packard Co., Pal Alto, CA

Introduction

The process by which humans distinguish visually between letters of the alphabet is a matter of considerable theoretical and practical interest. On the theoretical side, it is a classical problem in psychology, and has served as a test case for numerous theories of feature analysis and pattern recognition. On the practical side, it is relevant to the effective design of any textual display. Furthermore, principles discovered in the study of letter recognition may generalize to the perception of other types of displayed information.

In the psychological literature, a number of theories of letter discrimination have been proposed. These may be roughly divided into those that are feature-based, and those that are image-based. The former predict letter similarity based on the sharing of particular features, such as vertical lines, concave right curves, etc [1]. A defect of these models is that they do not specify the process by which the letter image is transformed into features. Consequently they are of little use in font design.

The image-based models predict similarity based on some measure of the luminance images of the letters, such as their overlap[2], or the overlap of their Fourier spectra[3, 4]. A defect of these models is that they have not been well motivated by basic principles of pattern recognition. A general problem with existing feature and image-based models is that they do not work very well[5].

Our goal was to construct a model of letter recognition that remedied these flaws. First, we sought an image-based model that could be applied to font design. Second, we sought a "principled" model, that is, one which assumed that the human observer employed a sensible and efficient recognition process. Finally, we sought a "minimal" model that incorporated only processes that could not be avoided. In this way we can test whether this simplest model is adequate, or whether other more complex processes must be considered.

There are many possible measures of legibility, including reading rates, letter discrimination, and letter recognition. We chose a recognition procedure in which we collected letter confusion matrices for low contrast letters of one font. Matrices could then be compared to those generated by the model.

Model

Before we enter into a detailed description of our model it is useful to note how it differs from most prior models. First, rather than operating on some abstract features ours operates directly on the intensity image. This has the advantage of obviating the difficult step of converting the image to features, and also tests whether such a conversion is necessary to understand letter discrimination. Second, rather than adopting some arbitrary recognition strategy, we adopt an ideal observer as the basic mechanism. This is an instance of the minimalist principle enunciated above, since the ideal observer is the only model which uses all available information and which therefore assumes no arbitrary losses of information[6].

Components

The overall structure of the model is pictured in Figure 1. The components are a spatial filter, a noise source, spatial position uncertainty, and an ideal observer.

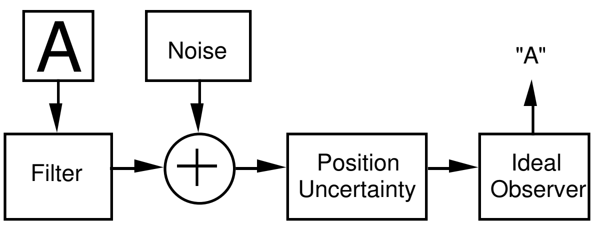

Figure 1. Letter recognition model.

The spatial filter represents the limited resolution of the visual system. We have modeled this filter as a Gaussian, with a scale parameter *s* controlling the amount of blur

$$\exp\left[\pi (x/s)^2\right] \quad (1)$$

In the frequency domain, this is a Gaussian with a scale of 1/s, and a half-amplitude half-width of 0.47/s. Figure 2 shows this filter, with a scales of 1.25, 2.5, and 5 pixels, superimposed on contrast sensitivity data collected in our lab [7]. The best fit is a scale of 2.5 pixels.

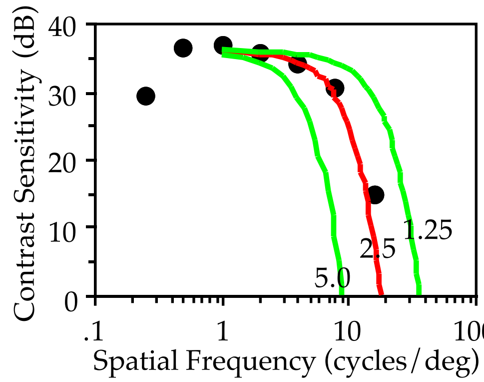

Figure 2. Contrast sensitivities and three spatial filters.

The noise element represents noise in both the signal and in the neurons in the early visual system. We model it by a spatially uncorrelated Gaussian process. By varying the standard deviation of the noise we can control the overall performance of the model.

The ideal observer maintains a memory image (template) of each possible letter, as it would appear after blurring. It examines the blurred, noisy sample image, and computes which letter is most probable[8]. However, because the ideal observer does not know exactly where the test image was, because of eye movements and the like, it must consider multiple templates for each letter, each consisting of the same template shifted by differing amounts horizontally or vertically. Each possible shift has a certain probability, which may be represented collectively by a prior density function. We represent this *uncertainty function* $p(x,y)$ as a Gaussian density with a particular scale. A large scale means high uncertainty, a small scale means little uncertainty. We considered scales of 0, 1, 4, and ∞ pixels. The last value corresponds to a uniform probability over the 32x32 pixel image.

The final output of the model is a letter name. Data are collected from the model by repeated Monte Carlo trials, and compared to results from human observers. Note that the human and model observers are presented with the identical letter stimuli.

Computations

Let the letter presented be indexed by *s*, and the candidate letter by *k*. Then the signal received is given by $m_s + n$, where m_s is the actual letter image and n is a noise image. For each candidate letter image m_k, we evaluate the discriminant function

$$d(k,s) = \sum_x \sum_y p \ \exp\left[\frac{1}{\sigma^2}\left[(m_s + n) \otimes m_k - \frac{1}{2}\|m_k\|^2\right]\right] \quad (2)$$

where p is the uncertainty function, σ is the noise standard deviation, $\otimes$ indicates discrete correlation, and $\|m_k\|$ is the norm of the candidate image. This function is monotonic with the posterior probability that candidate k was presented, given a signal $m_s + n$, such as would be produced by the sample letter image m_s. The model observer selects the candidate letter *k* for which this discriminant is largest.

General Methods

Some of our methods are common to both simulations and human experiments, and they are described here. We used a font (gacha.r.7) drawn from the font library on a SUN workstation (Fig. 3). Only upper case letters were used, with negative contrast. Each character was defined on a raster of 5 pixels wide by 9 pixels tall. At the viewing distance of 114 cm this corresponds to 7.5 by 13.5 min arc. This was centered in a raster of 32 by 32 pixels to prevent wrap-around during digital filtering.

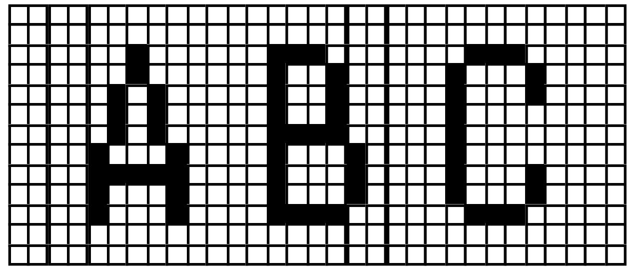

Figure 3. Three letters from the font gacha.r.7 used in the experiments and simulations.

Data were collected in two phases. In Experiment 1 we determined a *font recognition contrast threshold*. This is the letter luminance contrast threshold that yields approximately 82% correct identifications. The threshold was determined by means of the QUEST adaptive staircase procedure [9]. On each trial a letter was randomly selected (with replacement) and presented to the observer, either real or simulated. The observer reported the apparent identity of the letter. A psychometric function for the complete alphabet was maintained, describing the probability of recognition as a function of contrast, independent of letter. After each trial these data were fit with a Weibull function, and the next trial was placed at the 82% point of this function. At the conclusion of the experiment, threshold is estimated as the 82% point of the best fitting Weibull curve [10]. Figure 4 shows typical simulation results from Experiment 1.

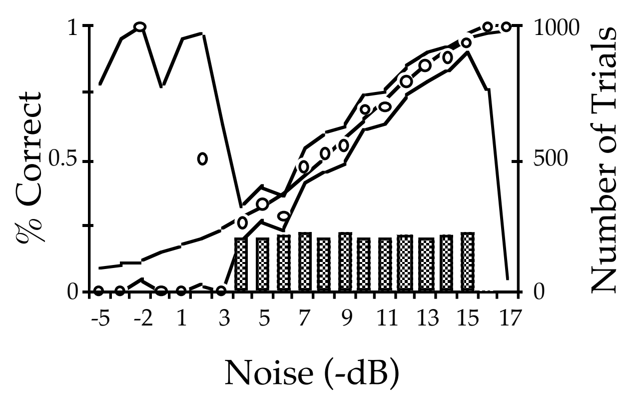

Figure 4. Simulated results from Experiment 1.

Circles are data, heavy line is the best fitting Weibull function, light lines are confidence limits, and bars are numbers of trials.

Experiment 2 is the collection of a confusion matrix. Contrast was set 1 dB below the font recognition contrast threshold. This contrast was chosen to permit generation of a useful confusion matrix, which requires a substantial number of wrong answers. With contrast fixed at this value, the complete set of letters was presented a number of times in random order. The number of each letter response to each letter presented was recorded.

Simulation Methods

Simulations were carried out on a SUN workstation augmented with an array processor. The blur filtering, and calculation of posterior probabilities were done in the frequency domain. In Experiment 1, 2600 trials were used. In Experiment 2, each confusion matrix contained 26000 trials (1000 trials/letter).

We simulated filter scales of 0, 1.25, 2.5, and 5 pixels. For a viewing distance of 114 cm, these correspond to frequency half-widths of ∞, 15.02, 7.51, 3.75 cycles/degree. The uncertainty function was simulated by a Gaussian with scales of 0, 1, 4, and ∞ pixels. The simulation proceeds at a rate of about 1.5 second/trial, or about 11 hours for a complete confusion matrix of 1000 trials/letter.

Psychophysical Methods

Letters were stored in an Adage RDS-3000 framebuffer and displayed on a monochrome monitor with a resolution of 20 pixels/cm. Viewing distance was 114 cm, providing an effective resolution of 40 pixels/degree. Letters were displayed with negative contrast on a background of 100 cd/m^2. Display was viewed binocularly with natural pupils. A small (one pixel) fixation point was provided between trials.

Frame rate of the display was 60 Hz, non-interlaced. The contrast during each presentation followed a Gaussian time course with a scale of 87 msec and a total duration of 200 msec. Calibration and other procedures are described elsewhere[11].

Responses were collected verbally and typed in by the experimenter. Feedback was provided in Experiment 1, but not in Experiment 2. Four observers served in both experiments. All were male between 17 and 33 years of age with corrected acuity.

Results and Data Analysis

Experiment 1

Letter recognition contrast thresholds for four observers were -15.79, -13.73, -14.61, -15.54 dB (average = 14.91, sd = 0.94). This mean corresponds to about 18% contrast. Each threshold was estimated from 128 trials.

Experiment 2

Confusion matrices with 60 trials/letter were collected from the same four observers used in Experiment 1. Overall percent correct for the four observers were: aef, 75.8; abw, 80.8; abp, 64.7; ejl, 78.3. The average matrix is shown in greyscale in Figure 5. Expressed relative to trials/letter, off diagonal values range from zero to 27%.

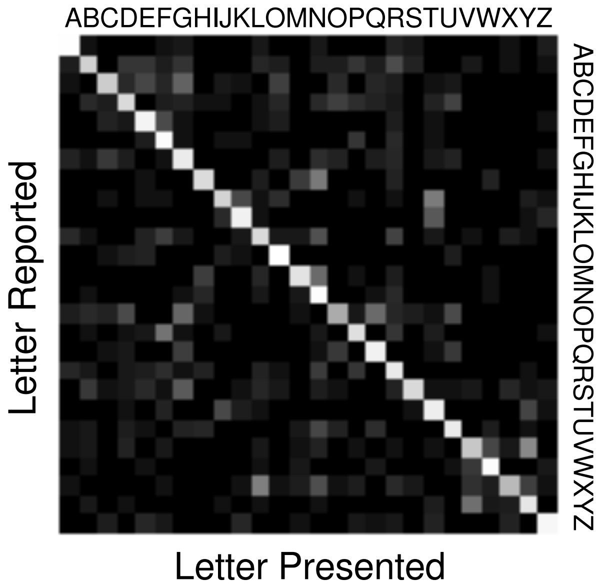

Figure 5. Average confusion matrix for four observers.

A simple comparison that can be made between empirical and simulated confusion matrices is the correlation. However it has been pointed out that this exaggerates the agreement since it tends to be dominated by the fact that both matrices have a large main diagonal (correct responses) [5]. Accordingly we consider separately the correlations between the main diagonals and the off-diagonals of the two matrices.

As a standard of comparison, we first generated a predicted confusion matrix using the "physical overlap" method described by Townsend [2]. This consists of counting the "on" pixels common to two letters when superimposed, and assigning matrix probabilities in proportion to this count. We used this as a standard because it appears to have produced the best published performance. The off-diagonal correlation was 0.49, and the on-diagonal was 0.12.

The correlations for our model with various parameters are shown in Table 1. For the filter spread of 2.5 pixels and the prior density of 1 pixel, the on- and off-diagonal correlations are 0.67 and 0.72. These are much higher than those of the overlap model. The correlation is somewhat sensitive to filter scale, but less sensitive to uncertainty scale.

Filter	Un-certainty	On	Off	Off, Unbiased
0	0	0.24	0.62	
0	1	0.25	0.62	0.65
0	4	0.16	0.55	
0	∞	0.13	0.51	
1.25	0			
1.25	1	0.39	0.67	0.74
1.25	4	0.32	0.65	
1.25	∞			
2.5	0	0.67	0.71	0.81
2.5	1	0.67	0.72	0.86
2.5	4	0.65	0.72	0.84
2.5	∞	0.65	0.71	0.86
5	0			
5	1	0.63	0.53	0.62
5	4			
5	∞			
overlap model	0.12	0.49	0.54	

Table 1. Correlations between empirical and model confusion matrices.

Bias

Although the correlations produced by the model are substantially better than those in the literature and than the overlap model, considerable variance remains unaccounted for. One probable source for this variance is bias. This may be seen in Fig. 6 which shows the frequency with which each letter was reported by each of the four observers. Since each of the letters was presented equally often, this is a rough indicator of bias.

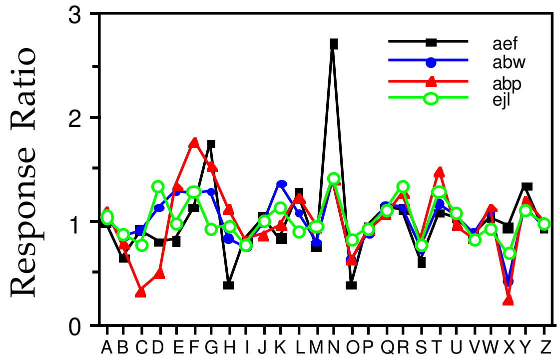

Figure 6. Ratio of response totals to presentation totals for each letter.

To determine how well our model accounts for the underlying similarity structure we transformed the matrix in a manner that removes bias [2, 12]. The off-diagonal correlations that result are shown in the last column of Table 1. With the best fitting parameter values (filter scale = 2.5, uncertainty scale = 1 or infinite), a correlation of 0.86 is obtained.

Conclusions

We have shown that a very simple model of letter recognition is capable of generating confusion matrices that correlate well with empirical matrices. After bias has been removed, little variance remains unaccounted for. This suggests that little additional predictive power will be gained by making the model more complex.

We must acknowledge that these conclusions apply only to the particular conditions we explored. It will be important to generalize these results to other letter sizes, fonts, conditions of presentation, and measures of performance. Indeed, a demonstration that such predictions could be made, without change in model parameters (noise level, filter scale, uncertainty scale) would greatly increase the value of the model.

One purpose of this study is the development of a robust, easily-calculated metric for the legibility of letters and fonts. When predictions depend upon Monte Carlo simulation, such a metric is difficult to specify. These simulations are necessary because position uncertainty complicates the discriminant function in such a way that a closed form solution is unavailable. However, the results in Table 1 show that uncertainty scale has only a modest effect on the quality of predictions. Consequently it may prove possible to use the zero-uncertainty closed-form solution as an approximation, and as the basis of a legibility metric.

Acknowledgements

Brief reports of this work have appeared elsewhere[13].

References

1. L.H. Geyer and C.G. DeWald,"Feature lists and confusion matrices," Perception and Psychophysics, 14(3), 471-482 (1973).

2. J.T. Townsend,"Theoretical analysis of an alphabetic confusion matrix," Perception and Psychophysics, 9(1A), 40-50 (1971).

3. S. Coffin,"Spatial frequency analysis of block letters does not predict experimental confusions," Perception and Psychophysics, 23(1), 69-74 (1978).

4. M.J. Gervais, J. Lewis O. Harvey and J.O. Roberts,"Identification confusions among letters of the alphabet," J. of Exp. Psychology, 10(5), 655-666 (1984).

5. R.S. leBlanc and J.G. Muise,"Alphabetic confusion: a clarification," Perception and Psychophysics, 37, 588-591 (1985).

6. A.B. Watson,"The ideal observer as a

modeling tool," In Frontiers of visual science: Proceedings of the 1985 Symposium (pp. 32-37). Washington D.C.: National Academy Press (1987).

7. A.B. Watson,"On the estimation of local scale," Journal of the Optical Society of America A, 4, 1579-1582 (1987).

8. R.O. Duda and P.E. Hart,Pattern classification and scene analysis, John Wiley, New York (1973).

9. A.B. Watson and D.G. Pelli,"QUEST: A Bayesian adaptive psychometric method," Perception and Psychophysics, 33(2), 113-120 (1983).

10. A.B. Watson,"Probability summation over time," Vision Research, 19 , 515-522 (1979).

11. A.B. Watson, K.R.K. Nielsen, A. Poirson, A. Fitzhugh, A. Bilson, K. Ngugen and J. A. J. Ahumada,"Use of a raster framebuffer in vision research," Behavioral Research Methods, Instruments, & Computers, 18(6), 587-594 (1986).

12. R.D. Luce,"Detection and recognition," (1963). In R. D. Luce, R. R. Bush, & E. Galanter (Ed.), Handbook of Mathematical Psychology, New York: Wiley.

13. A.B. Watson and A.E. Fitzhugh,"A new look at letter recognition," Perception, 15(1), A31 (1986).