**Title**: A Spatial Frequency Analysis Model for Predicting Human Performance at Visual Pattern Matching Tasks

**Author(s)**: Mark W. Cannon, Jr.

**Performing Organization Name and Address**: Aerospace Medical Research Laboratory, Aerospace Medical Division, Air Force Systems Command, Wright-Patterson Air Force Base, Ohio 45433

**Report Date**: June 1977

**Type of Report & Period Covered**: Proceedings

**Distribution Statement (of this Report)**: Approved for public release; distribution unlimited

**Abstract**: A model for simulating human performance in visual pattern matching tasks is presented. The model is based on evidence of spatial frequency processing in the visual system, and on the hypothesis that shape recognition is determined only by the low spatial frequency harmonics of the image. Two psychophysical pattern matching experiments are described that demonstrate a clear functional relationship between the "similarity" of two patterns as judged by human observers and the Euclidean distance between spatially filtered Fourier transforms of the patterns.

**Supplementary Notes**: Prepared for presentation at the meeting and publication in the Proceedings of Systems, Man and Cybernetics Society, sponsored by IEEE, to be held in Washington, D. C., 19-21 September 1977.
A SPATIAL FREQUENCY ANALYSIS MODEL FOR PREDICTING HUMAN PERFORMANCE AT VISUAL PATTERN MATCHING TASKS

Mark W. Cannon, Jr.
Aerospace Medical Research Laboratory
Wright-Patterson Air Force Base, Ohio

Summary

A model for simulating human performance in visual pattern matching tasks is presented. The model is based on evidence of spatial frequency processing in the visual system, and on the hypothesis that shape recognition is determined only by the low spatial frequency harmonics of the image. Two psychophysical pattern matching experiments are described that demonstrate a clear functional relationship between the 'similarity' of two patterns as judged by human observers and the Euclidean distance between spatially filtered Fourier transforms of the patterns.

Introduction

The research described in this report represents a portion of the work being conducted at the Aerospace Medical Research Laboratory to develop quantitative models for observer-display interactions. These models will lead to design of displays optimally matched to human information processing capabilities under a variety of conditions. This paper addresses, in particular, the problem of predicting the confusability of symbols of the type that may be used in a graphic display. Alphanumeric symbols are used for the tests discussed here but the technique is not limited to these. Any two-dimensional display symbols can be analyzed by this technique.

Two bodies of research have helped to lead our work in its present direction. The first of these is the increasing amount of psychophysical and neurophysiological literature demonstrating the organization of the visual system as a spatial frequency analyzer. These works range from the early reports of Campbell et al.1,2,3 to some of the more recent works by Hamerly, Quick and Reichert4 and by Carlson, Cohen and Gorog5. These latter papers demonstrate clearly that from threshold to contrasts and computing results which provide a strong partial validation of the model.

Description of Experiments

Computer Analysis of Symbols

A description of the computer processing performed on the symbol sets is in order at this point, since essentially the same model predictions are used to analyze both psychophysical experiments. The symbols used were digitized as ones on a background of zeros for computer analysis. Each symbol, except for a few in set 1, has a maximum size of 10 x 14 points, and is located in a 32 x 32 background window of zeros. The four symbol sets used in these experiments are shown in the appendix in the same form in which they were presented to the human observers. The smooth effect of digitized diagonal lines was therefore presented to both humans and computer.

In both experiments, the same symbol set was used as both input and prototype. As we will see, this is fully justified in experiment 1, and is the best we can do at present for experiment 2. The 36 symbols were Fourier transformed using a two-dimensional Cooley-Tukey FFT algorithm and filtered using a square low-pass filter. The dc term and all harmonic components up to the maximum desired were saved while all higher terms were set equal to zero. Next, a 36 x 36 correlation matrix was generated. Each entry of this matrix contained the maximum value of the cross-correlation functions computed between the filtered input symbol and each of the other symbols of the set. The correlation function for a symbol pair (i,j) was determined by multiplying the filtered spectrum of symbol i with the complex conjugate of the filtered spectrum of symbol j and taking the inverse transform of the product. The maximum amplitude of this inverse transform was the maximum correlation coefficient. Finally, a distance matrix D was computed from the correlation matrix by determining the Euclidean distances at which maximum correlation occurred. This distance is the minimum mean square distance between the two patterns, and can be derived from the maximum correlation coefficient, as we now show. Let X and Y be the position vectors of
two patterns in the spatial domain and let \( d \) be the distance between \( X \) and \( Y \).

\[
d = |X - Y| = \sqrt{(X - Y) \cdot (X - Y)} \tag{1}
\]

\[
d = \sqrt{(X \cdot X - 2X \cdot Y + Y \cdot Y)} \tag{2}
\]

The vectors have been energy-normalized after filtering so the dot products are equal to 1 and the distance is

\[
d = \sqrt{2 - 2p} \tag{3}
\]

where \( p \) is the cross-correlation between \( X \) and \( Y \).

If we let \( p \) be the maximum of the cross-correlation function, the distance \( d \) is minimized. The value of \( d \) can range from zero to 1, since \( p \) ranges from zero to 1. The \( 36 \times 36 \) distance matrix \( D \) contains these minimum distances between each input symbol and all other symbols of the set. Diagonal elements were zero, since these represent the distance from each symbol to itself.

**Experiment 1: Shape Matching**

The subjects were seated before a chart containing symbols of font 1, 2 or 3, as shown in the appendix, but the symbols on the experimental charts were arrayed in a random order. The relative distances between the symbols was larger than that shown on the charts in the appendix and the symbols themselves were 2.5 cm wide. The subject was seated at a distance from the chart such that the symbol width was one degree of visual angle. Subjects were given a randomized list naming all 36 symbols in the set. The subjects were instructed to locate on the chart a symbol named in the random list and then to report which other symbol on the chart matched it most closely in shape.

In the account below, we will refer to the symbol to be matched as the test symbol and the symbol chosen by a subject as the comparison symbol. Set 1 was viewed by 35 subjects; set 2 by 23 subjects and set 3 by 27 subjects. The results showed that for each symbol of the test set, there was a good deal of agreement about which comparison symbol was closest. If the comparison symbols for each test symbol are arranged in rank order by the number of times it was chosen as closest, and the number in each rank is summed over all three sets as in Fig. 1, we see that approximately 50% of the choices are in rank 1. In fact, the curve is nearly perfectly exponential by rank. In rank 2, 25% of the people agree on the closest match and in rank 3 approximately 12% agree that this is the closest match.

Since 75% of the subjects agree that the rank 1 or 2 choices are the closest matches, we decided that our computer model would be given credit for a correct choice if the comparison symbol picked by the model as closest to each test symbol agreed with either the human rank 1 or 2 choice.

The results of the computer choices for a range of filter bandwidths is shown in Fig. 2. The ordinate gives the number of times the model agreed with human predictions. The best score across all three sets occurs at the eighth harmonic of the \( 32 \times 32 \) viewing window, which is between the second and third harmonic of the symbol. The score at the sixth harmonic of the window, which is equal to the second harmonic of the symbol, is just marginally lower, so maximum performance occurs in the sixth to eighth harmonic range. The overall score is an 80% correct match with the human data. Note also that performance deteriorates at both larger and smaller filter bandwidths. We see that those symbol pairs with smallest intersymbol distances as determined by our model correspond quite well with human judgments of symbol pairs most similar in shape. However, if shape similarity is related to intersymbol distance, there should be a functional relationship between the number of times a comparison symbol is chosen as most similar to a test symbol and the Euclidean distance between test and comparison symbols. This relationship is derived in the following way.
The human matching results for a particular symbol set are arranged into a 36 x 36 choice matrix C. Each entry \( C_{ij} \) is the number of times that subjects chose a comparison symbol \( j \) as most similar to test symbol \( i \). The intersymbol distances corresponding to each \( i,j \) pair are contained in previously computed distance matrix \( D \) for this symbol set. A filter corresponding to the sixth harmonic of the window was used to compute \( D \). Let us now divide up the distance axis into bins of width .05. Using matrices \( C \) and \( D \), we add up all choices that fall in a given distance bin and divide by the number of test symbols that generated choices in that bin. This gives us the average number of choices per test symbol at a given distance between test and comparison symbols. Further normalization was accomplished for each set by dividing these averages by the number of subjects who took part for that set. The final averages for all three sets are plotted as a function of distance in Fig. 3. The points from the three sets overlap in a very satisfactory manner, almost as if all were derived from the same function. Thus Fig. 3 demonstrates, as we had hoped, that the relative number of times a particular comparison symbol is chosen to be a best match for a given test symbol decreases with increasing distance between comparison and test symbols. Apparently, similarity and Euclidean distance are functionally related for our spatially filtered symbols.

MATCHES VS DISTANCE

![Graph showing matches vs distance](image)

Fig. 3. The sample points represent the average number of times per test symbol that a given comparison symbol was picked as a closest match in shape to that test symbol. The abscissa is the model-predicted Euclidean distance between spatially filtered test and comparison symbols. The solid line is the average of the sample points, and shows a definite decrease with distance between symbols.

**Experiment 2: Symbol Recognition**

In a series of psychophysical experiments conducted at the Air Force Flight Dynamics Laboratory, Dr. Larry Goble has evaluated the confusability of three of the symbol sets shown in the appendix. Sets 2, 3, and 4 were used in his experiment. The subjects viewed the sets statically to become familiar with the shapes. They were then asked to identify the symbols when they were flashed on a screen and partially masked by a preceding and following uniform field of the same intensity as the symbol. The paradigm proceeded as follows: uniform field 10 msec, blank screen 5 msec, symbol 10 msec, blank screen 5 msec, and uniform field 10 msec. This paradigm involves some significant differences from the matching experiment covered in the previous section. First, the short duration of the symbol presentation and the masking involve some temporal input parameters not yet treated in the model. We cannot yet say what effects stimulus duration would have on the complicated spectrum of a symbol, and we cannot adequately define the effect of pre- and post-stimulus masking. We will just assume these effects are small when we apply the model to analyze the results of these psychophysical experiments. The second problem is the prototype to which the test symbols are compared. We assume that the subject can learn the symbol shapes well enough to compare the inputs to them. However, the prototype may be some combination of a wide variety of fonts to which the subject has been exposed. The fact that the subject can change his prototype set indicates that the subject can learn the symbol shapes well enough to compare the inputs to them. The other two sets showed trial-to-trial variability, but the average number of correct responses remained relatively constant over all trials.

The first comparison of model performance to human data was very similar to the comparison made in Fig. 2. The symbol with the smallest intersymbol distance in the ith row of the distance matrix \( D \) was picked as the model's choice for the symbol most confusable with the ith test symbol. We asked, "In how many cases does this choice agree with the symbol that produced either the largest or second largest number of human errors in each row?" The results are plotted in Fig. 4. The plot shows the number of test symbols for which model predictions agreed with human data as a function of the filter bandwidth. The striking difference between these plots and those of Fig. 2 is that they reach a maximum percentage of agreement at a filter bandwidth equal to the fifth harmonic of the window and are essentially flat above that bandwidth. Note also that the average curve has a maximum of only 47% compared to 80% for the static matching experiment. However, scanning the data showed us that there was still considerable correlation between the distribution of errors and Euclidean distance. In Fig. 5 we derive a curve giving the average number of human errors per test symbol as a function of the model-generated distance between the test symbol and a comparison symbol. We again divided the distance axis into bins of width .05. We summed the number of errors in each bin and divided by the number of test symbols for which errors were generated in that bin and by the total number of errors for all symbols of that set. The total errors varied considerably over the three sets. There were 1639 errors for set 2, 3184 errors for set 3 and 2893 errors for set 4. The averages for each set were plotted as the sample points in Fig. 5, and we see that there is a definite decrease in number of errors per test symbol as the distance between test and comparison symbol increases. Again, all three symbol sets give results which fall along the same curve, independent of the symbol set, depending only on Euclidean distance.
we'll in attempting to predict which particular symbol pairs will produce the most errors (Fig. 4) when the symbol images are dynamic. The general trends which were shown for the static experiment do hold, however, and we may say that on the average those symbol pairs which are shown to be close together by the static analysis model will produce more confusions in the dynamic case than those symbol pairs shown to be far apart statically. It may be possible to treat the uniform masking fields as some sort of noise in the system, since it is their presence that causes most of the errors. A number of researchers have developed models of temporary visual memory storage which may be applicable. Sperling, in particular, has demonstrated that presentation of a uniform pre- and post-stimulus masking field interferes with the storage of symbols in this temporary memory area. However, a much more general model of spatio-temporal interactions in the visual system is required before we can apply this promising approach to the analysis of dynamic display imagery.

The results of experiments by Arend have demonstrated that at least at threshold, there is a greater decline in sensitivity to high spatial frequencies than to low spatial frequencies as the stimulus duration is decreased from continuous to 20 msec. Thus the distortion in the spectrum of the input symbol image is due to the temporal response variation among the spatial frequency channels as well as the interference of the uniform field. Similar spatio-temporal effects are being investigated by our laboratory, and should supply a firm base on which to build a model of human performance that can be applied to dynamic as well as static imagery.

Fig. 4. The sample points represent the percentage of times that the symbol pairs with smallest model-derived intersymbol distances were the same as those symbol pairs having the largest or second largest number of confusions in a symbol identification experiment. The solid line is the average of the points. The abscissa gives the bandwidth of the model spatial frequency filter as explained in the text.
References


