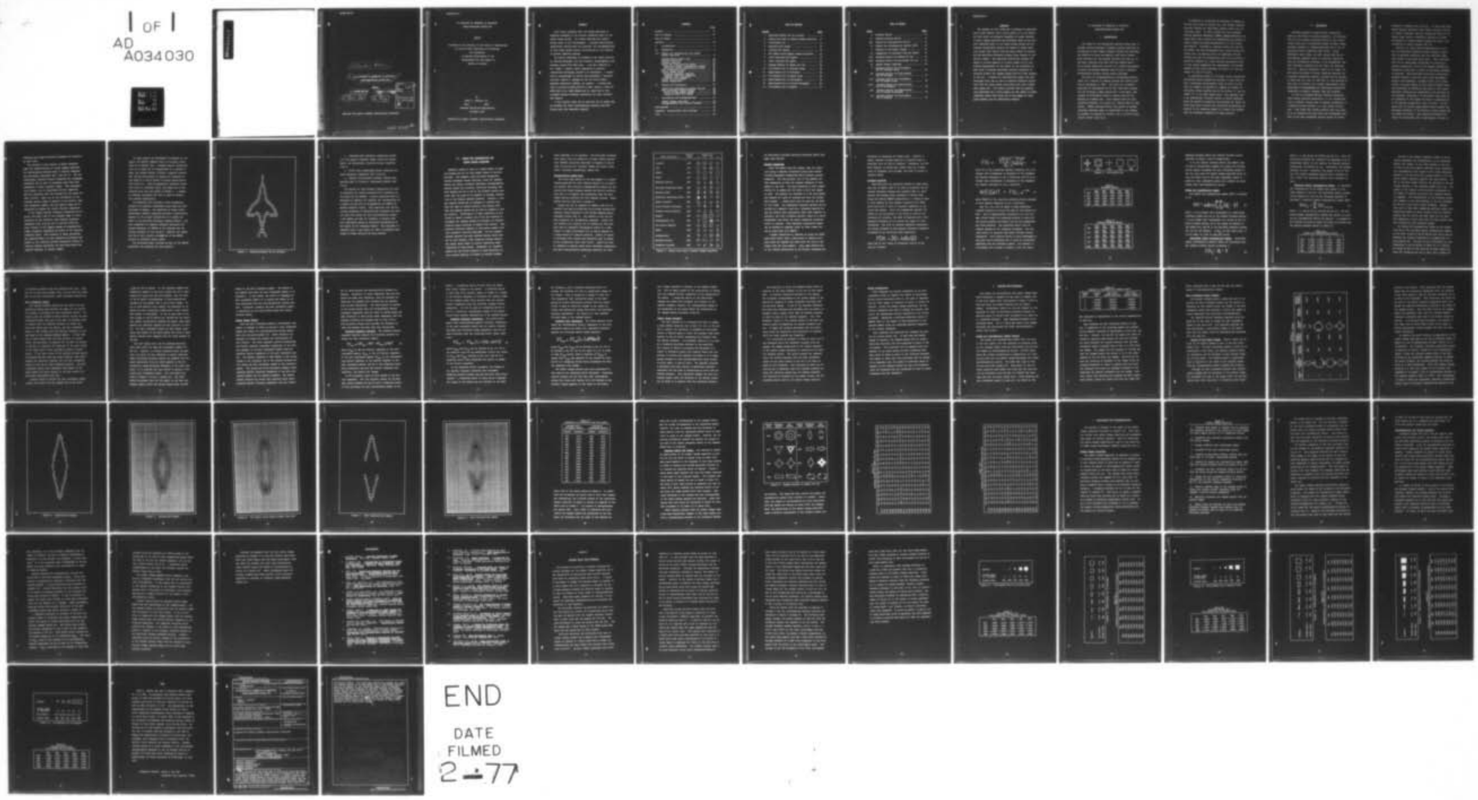
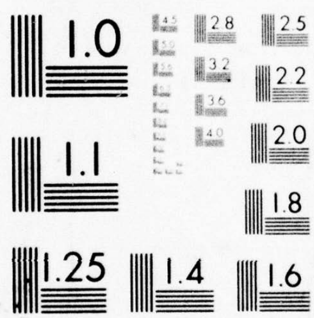
AD-A034 030 AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OHIO SCH--ETC F/G 6/4
AN ALGORITHM TO GENERATE AN OPTIMALLY HUMAN-SEPARABLE SYMBOL SE--ETC(U)
DEC 76 J A JOHNSON
UNCLASSIFIED GE/EE/76D-26 NL

1 OF 1
AD
A034030



END
DATE
FILMED
2-77



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

GE/EE/76D-26

(9) Masters thesis,

(6) AN ALGORITHM TO GENERATE AN OPTIMALLY HUMAN-SEPARABLE SYMBOL SET.

THESIS

(14) GE/EE/76D-26

(10) James A. Johnson
Capt USAF

(11) Dec 76

(12) 71p.

ACCESSION for	
NTIS	White Section ☑
DDC	Buff Section ☐
UNANNOUNCED	☐
JUSTIFICATION	
BY	
DISTRIBUTION/AVAILABILITY CODES	
Dist.	AVAIL. and/or SPECIAL
A	

Approved for public release; distribution unlimited.

012 225 dn

GE/EE/76D-26

AN ALGORITHM TO GENERATE AN OPTIMALLY HUMAN-SEPARABLE SYMBOL SET

THESIS

Presented to the Faculty of the School of Engineering
of the Air Force Institute of Technology
Air University
in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

by

James A. Johnson, B.S.
Capt USAF
Graduate Electrical Engineering

December 1976

Preface

This thesis presumes that the reader possesses a moderate knowledge of the Fourier transform model of the human visual system. The reader desiring more detail should refer to the bibliography. Although statistically significant results were not obtained, the recommendations of this study should allow a continuation of the research to provide improved results.

My sincere gratitude is extended to my thesis advisor, Dr. Matthew Kabrisky, for his interest, encouragement, and guidance during this past year. I am also indebted to Maj. Roger A. Gagnon, whose experience in pattern recognition techniques proved to be invaluable. I should like to acknowledge the advice and assistance I received from Maj. Robert P. Bateman of the Air Force Flight Dynamics Laboratory (AFFDL), my sponsor. I should also like to extend my appreciation to Capt. Larry G. Goble of AFFDL and to Dr. Mark Cannon and Lt. Gary Sims of the Aerospace Medical Research Laboratory for their interest and support.

A very special thank you is reserved for my typist and my children for their understanding, patience and love during some very memorable moments.

Contents

List of Figures

List of Tables

Abstract

The purpose of this study was to develop an algorithm which could produce from a given symbol set a new symbol set that would be optimally separable by human subjects. A symbol change algorithm was developed based on the Fourier transform model of the human visual system and the inverse relationship between the number of human confusion errors that result from the basic shapes of symbols and the Euclidian distances between these symbols in the Fourier domain. The algorithm effectively changes the shape of certain symbols of a given symbol set by controlled manipulation of the symbol in the transform domain until a minimum, Euclidian distance threshold is achieved between the changed symbol and all other symbols in the set. A symbol set confusability number was also developed which can be used to evaluate different symbol sets with the lower number identifying the least confusable symbol set. The author believes that by applying this algorithm for a given symbol set the number of human confusion errors that occur due to the shapes of these given symbols will be effectively reduced.

AN ALGORITHM TO GENERATE AN OPTIMALLY HUMAN-SEPARABLE SYMBOL SET

I. Introduction

The number of two dimensional symbolic codes used to display various meanings to weapons systems operators is increasing as more complex weapons systems are developed. The addition of each new symbolic code often creates some confusion for the systems operator who must be, or has been, familiar with more than one weapons system. This becomes especially critical in a combat environment in which the systems operator must be able to rapidly differentiate between various symbol meanings.

The need for standardization in developing symbolic codes for a variety of weapons systems was stated quite well by Honigfeld in 1964: "The need for a standard symbology is highlighted by the fact that each contractor who develops a radar system has, in the past, been allowed to arbitrarily select a symbol code and its meaning for display use. Since symbols have not been specified formally, the result is a unique code for each system. Symbol meanings differ from system to system; identical meanings might be represented on one display by numbers, on another by letters, and on a third by geometric forms" (Ref 12:1).

In addition to confusing the meanings of symbols in one set with those of another set, the systems operator may also confuse the individual symbols within one set with each other. In 1973 a group from the Aerospace Medical Research Laboratory (AMRL), Wright-Patterson Air Force Base, Ohio, studied the standardization problem of symbol coding and range presentation in a high threat environment. Although an "optimum" symbol set was not developed, the team recommended that a given symbol set should use a flickering circle to highlight critical threats and that the alphabetical form of the letter "A" should be replaced by a more distinct geometrical form. If this latter suggestion was implemented, the systems operator would avoid confusing the letter "A" with the number "4" (Ref 14:27-32).

An engineering model of the human visual information processing system proposed by Kabrisky in 1966 (Ref 13) and a subsequent psychological study by Goble in 1975 (Ref 10) suggest an approach to explain the degree of confusion between symbols such as the letter "A" and the number "4" in the AMRL symbol sets. This approach uses a two dimensional, discrete Fourier or Walsh transform model for pattern recognition and provides the basis for this study which involves the controlled manipulation of symbols in the transform domain to "create" new symbols that are optimally separable by human subjects.

II. Background

Kabrisky proposed a human pattern recognition model in 1966 based on Fourier analysis. He theorized that the rich neural interconnectivity of the visual cortex could save a significant number of basic computational elements by transferring data with a two dimensional Fourier transform (Ref 13:82). Using Kabrisky's work, Radoy, in 1967, showed that the entire English alphabet could be correctly classified from the information contained in the fundamental spatial frequency component and the first two harmonics (Ref 16). Based on Kabrisky's model of the human visual system and Radoy's Fourier transform approach for a pattern recognition machine, further work by Tallman in 1969 (Ref 17), Carl in 1969 (Ref 3), and Granlund in 1970 (Ref 11:195-201) more closely approximated human performance in processing and classifying handprinted characters by machine (greater than 95% accuracy).

In 1973 an Aerospace Medical Research Laboratory team studied problems associated with the standardization of symbolic coding used to display information to a weapons systems operator in a high threat environment. They found that certain symbols were more likely to be confused with each other and recommended that some of the more confusable symbols should be either replaced or changed (Ref 14:31,32). In that same year, Thomas applied Kabrisky's more highly refined, Fourier (or Walsh) transform, pattern recognition model of the human visual system to develop a predictor of human performance in certain symbol recognition tasks, concentrating on threat symbols (Ref 19).

However, not until Goble's dissertation in 1975 had the Euclidian distances between symbols, characters or patterns resulting from the transform model been linked to the human error rate. Goble indicated that "as an inadvertant concomitant to producing a pattern recognition device which handles handprinted alphabetic characters, a matrix of Euclidian distances is produced. Each prototype letter has an Euclidian distance between it and all other class prototypes. Using a third harmonic, low-pass filter (7x7) establishes a 49 space domain in which each filtered transform prototype is defined by one point in this space. Euclidian distances are normalized so the distance of, for instance, prototype A to prototype B may serve as an indication of the degree of error or confusability in the pattern recognition" (Ref 10:19-20). He proposed that an inverse relationship exists between the Euclidian distances in the Fourier or Walsh domain and the number of errors. Thus, smaller Euclidian distances are associated with a high degree of error or confusion and larger Euclidian distances are relatively error free.

The success of this Fourier or Walsh transform model for classification of visual imagery parallels the technological advances made in digital computers and processing techniques (Ref 1 and Ref 5). The two dimensional, discrete Fourier transform (an orthogonal transform) can be used to decompose input data that represents a visual (spatial) image. This decomposition yields a set of sinusoidal Fourier components that has preserved both the distance and the angle between vectors. Thus, all of the information in the original spatial image is present in the Fourier (spatial frequency) domain, and the spatial image can be reproduced by adding the correct Fourier components.

If the higher spatial frequency components are removed, the basic form of the original image is still preserved by the filtered, lower spatial frequency components (Ref 3:52-52, Ref 4, and Ref 9:92-103). Finer details of the imagery appear as progressively higher frequency components are added to the filtered transform space. This property of image processing and classification not only suggests an alternate approach to the classical pattern recognition theory of distinct feature extraction but also facilitates the computer analyses of spatial images.

To fully employ the advantages of analysis by computer, the spatial image(s) must be initially reproduced in a digital form. Although digital representation of an image can be accomplished in a variety of ways, the primary method utilizes a computer controlled, flying spot scanner to sample the intensity of the image at a finite number of evenly spaced points (Ref 18:12-13). With an appropriate threshold intensity level to eliminate noise, the image can be represented much like the symbol in Figure 1. A number of computer techniques can then be used to analyze such digital simulations.

One algorithm that performs a form recognition analysis on digitally represented input data has been developed by Gagnon. His predictor of human visual performance (PREVIP) algorithm provides a quantitative measure of the overall suitability of a symbol set based on the Euclidian distances between symbols. The full complement of computer analysis techniques and options available in PREVIP is not required for the initial purposes of this project. However, the reader can find a more complete description of the PREVIP options in reference number seven.

The following steps, included as part of the PREVIP algorithm, are required for this project:

```
             **
             **
             **
             **
             **
            ****
            ****
           **  **
           **  **
           **  **
          ***  ***
          **    **
          **    **
         ***    ***
       ****      ****
      ****        ****
    ****            ****
   ***                ***
  ***                  ***
 ***                    ***
***     ***      ***     ***
************    ************
*******   ***  ***   *******
           **  **
           **  **
           ******
            ****
            ****
             **
            ****
            ****
           ******
          **    **
          **    **
```

Figure 1. Digitized Symbol for an Aircraft

1. Transform each digitally represented symbol into the spatial frequency domain using the Cooley-Tukey, two dimensional, discrete Fourier transform (Ref 5).

2. Filter each transformed symbol retaining the lower frequency components (Ref 17:32-37).

3. Compute the Euclidian distance between each possible pair of filtered, transformed symbols (Ref 10:15).

The results of this distance computation are used to determine the overall suitability or confusion of the given set of symbols. Another algorithm which is described in Chapter III computes the confusability of a symbol set and can be incorporated into PREVIP as another analysis option. This algorithm uses the Euclidian distances to quantitatively determine the symbol which will cause the most discrimination errors. A new symbol can then be "created" to replace this most confusable symbol by controlled manipulation of the symbol in the transform domain. The algorithm is repeated until a new symbol set that is optimally separable by human subjects has been created.

III. Symbol Set Confusability and Symbol Change Algorithm

Symbolic coding is used in a large number of weapons systems and for an even larger number of mission profiles. For example, point-to-point navigation, terrain avoidance aerial delivery, close air support, search and rescue, airborne intercept, reconnaissance, missile launch and recovery, satellite tracking, and area defense are but a few of the broader mission profiles that use some set of symbols to display information to the weapons systems operator. However, a single weapons system may be designed to perform more than one type of mission profile, and usually more than one weapons system can be used to complete a single profile. Furthermore, if each contractor is permitted to select a symbol set of his own choosing to accompany the weapons system he develops, an extraordinarily large inventory of symbol sets results. This produces different symbols or even entire symbol sets that represent identical meanings. For the weapons systems operator who must perform various tasks with one or more systems, much confusion between symbols can result which leads to errors, especially under conditions of high stress. Therefore, an "optimum" symbol set (one which produces no discrimination errors between symbols) is needed to rapidly display exact meanings to the operator. The Euclidian distances that result from the predictor of human visual performance (PREVIP) algorithm described in Chapter II can be used to develop this "optimum" symbol set from a given input, digitally represented, symbol set.

Representative Symbol Sets

The first step needed in the development of a symbol change algorithm for this research project is to select the symbols that provide a representative symbol set for current and future weapons systems and mission profiles. Four sets of 18 symbols each adequately span the more common mission profiles for most weapons systems. These symbol sets are depicted in Figure 2.

Symbol set one is composed of symbols that are extracted from high altitude and low altitude enroute and terminal charts prepared by the Defense Mapping Agency Aerospace Center (Ref 6). The symbols of set one are used primarily for navigational purposes. Symbol set two consists of a typical set of cathode ray tube symbols used to represent cartographic points in a high threat or combat environment but is easily adapted to a large variety of mission profiles. Symbol set three is developed from symbols and modifiers used by Thomas in his recognition tests (Ref 19:20). Symbol set four is composed of symbols taken from a National Aeronautics and Space Administration symbology character set used

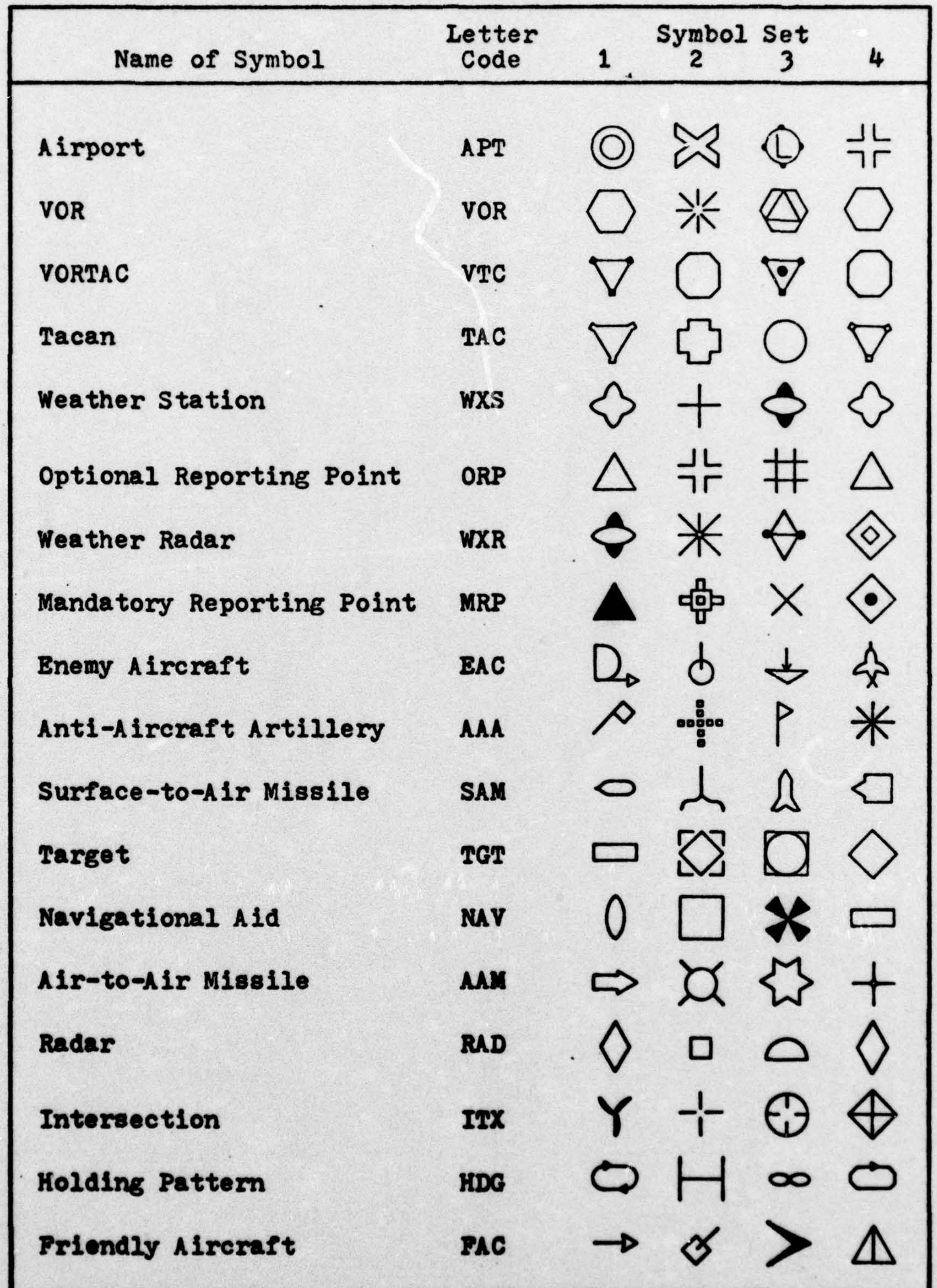


Figure 2. Symbol Sets Used in Symbol Change Algorithm

for Electronic Attitude Direction Indicator (EADI) displays (Ref 20:129).

Digital Simulation

After the symbol sets are chosen, they are digitized using a computer controlled flying spot scanner (Digital Equipment Corporation PDP-12 general purpose computer). For this project, all of the symbols are size normalized so that the largest dimension of each symbol is one inch. The only exception to this normalization is in symbol set two where a half inch square is used. The symbols are centered in a two inch square and scanned on a grid of varying size with the flying spot scanner. A 64x64 point grid is used for the data in this study. The subsequent intensity pattern can be written on a magnetic tape and a threshold applied using the Control Data Corporation (CDC) 6600 computer to provide a two level intensity pattern (black on a white background) from which the noise (undesired intensity level) is removed. Each digitally represented symbol can be punched on computer cards in octal format for use as input data with PREVIP.

A higher resolution is obtained by using the 64x64 point grid rather than a 32x32 point or smaller size grid since the symbols are about half the size of the window (the two inch square). Also, edge effects due to the induced periodicity of the Fourier transform are minimized by employing the larger grid. However, a larger computer storage capacity is required than is available with the PDP-12 computer. Therefore, the CDC 6600 computer is needed and, rather than use a large block of permanent file storage, the data is stored on computer cards.

Distance Matrices

Operating with the digitized symbols as input data, task one of PREVIP (Ref 7) is used to calculate the Euclidian distances between each possible pair of input symbols and arrange these distances in matrix form. Using the special PREVIP nomenclature, a "field of view" of four degrees by four degrees represents the 64x64 point grid with a "size of interest" parameter of two cycles per degree. Since the fundamental spatial frequency is the reciprocal of the field of view in the direction of interest, the data is analyzed with a fundamental frequency of 0.25 cycles per degree in both dimensions. The size of interest parameter determines the weight assigned to each spatial frequency component according to the following three equations:

$$FOA = 30 \times \ln(8 \times SI) \tag{1}$$

where FOA is the "field of attention" and SI is the size of interest.

$$FN = \frac{\sqrt{FX^2 + FY^2}}{8 \times e^{-FOA/30}} \quad (2)$$

where FN is the normalized spatial frequency, FX is an integer that corresponds to harmonics of the fundamental spatial frequency in the x direction, and FY is an integer that corresponds to harmonics of the fundamental spatial frequency in the y direction.

$$WEIGHT = FN \times e^{1-FN} \quad (3)$$

where WEIGHT is the resulting weighting factor assigned to each spatial frequency in the y direction.

Eqs (1), (2), and (3) apply a human modulation transfer function to filter the data in the Fourier domain. This is approximately equivalent to a 7x7 low-pass filter except for weighting factors of unity for the second harmonic and nearly one for the fundamental and third harmonic. The weighting factor then decreases rapidly as the frequency increases. The distance matrix is computed from these filtered Fourier components. This distance matrix is then rearranged so that each row of distances are in order of increasing magnitude from the reference symbol. For example, a five-symbol set is shown in Figure 3 with the corre-

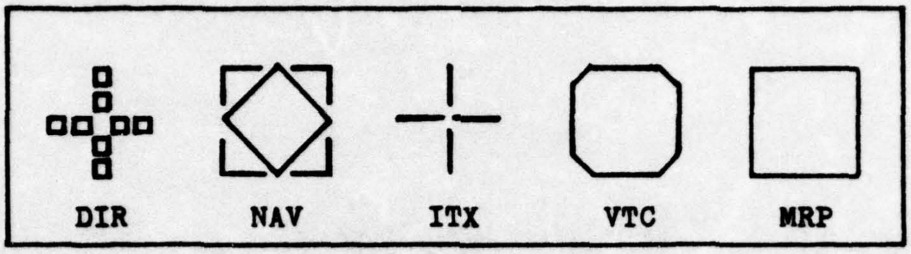


Figure 3. Five-symbol Set

Table I
Distance Matrix

	DIR	NAV	ITX	VTC	MRP
DIR	.0000	.7978	.2126	.8330	.8550
NAV	.7978	.0000	.7894	.3520	.2980
ITX	.2126	.7894	.0000	.8191	.8371
VTC	.8330	.3520	.8191	.0000	.3116
MRP	.8550	.2980	.8371	.3116	.0000

Table II
Ordered Distance Matrix

	DIR	ITX	NAV	VTC	MRP
DIR	.0000	.2126	.7978	.8330	.8550
	NAV	MRP	VTC	ITX	DIR
NAV	.0000	.2980	.3520	.7894	.7978
	ITX	DIR	NAV	VTC	MRP
ITX	.0000	.2126	.7894	.8191	.8371
	VTC	MRP	NAV	ITX	DIR
VTC	.0000	.3116	.3520	.8191	.8330
	MRP	NAV	VTC	ITX	DIR
MRP	.0000	.2980	.3116	.8371	.8550

sponding distance matrix and ordered distance matrix depicted in Tables I and II respectively.

It is the ordered distance matrix from PREVIP that is used to sequentially compute the symbol set confusability number, determine the most confusable symbol in the set, and enter the symbol change algorithm which alters the basic form of the most confusable symbol to achieve a less confusable symbol and a symbol set which causes fewer discriminability errors.

Symbol Set Confusability Number

The symbol set confusability number (SSC) is defined to be

$$SSC = \frac{1}{NS(NS-1)} \sum_{i=1}^{NS} \sum_{j=2}^{NS} \left(\frac{1}{DS_{ij}} - \frac{1}{2} \right) \qquad (4)$$

where i is an integer that corresponds to a particular reference symbol and row of the ordered distance matrix, j is an integer that corresponds to a column of the ordered distance matrix, NS is the number of symbols in the symbol set, and DS is the Euclidian distance between the ith and jth symbols. If DS_{ij} is less than 0.001, it is set equal to 0.001 to keep SSC finite.

Individual Symbol Confusability Number. Individual symbol confusability numbers (ISCs) are calculated from the ordered distance matrix according to

$$ISC_{ij} = \frac{1}{DS_{ij}} - \frac{1}{2} \qquad (5)$$

where i, j, and DS are as defined for Eq (4). Since the Euclidian distances are computed from mappings on an NS dimensional sphere of unit radius, the maximum distance between any two symbols is two. Conversely, the minimum distance is zero. Therefore, the range of ISC is from zero for ideally separated symbols with a distance of two to 999.5 for identical symbols with a distance of zero.

Reference Symbol Confusability Number. To represent the confusability of one symbol with respect to the other symbols in the set, the individual symbol confusability numbers are used in the following equation to calculate a reference symbol confusability number (RSC):

$$RSC_{ij} = \sum_{j=2}^{NS} ISC_{i,\, NS-j+2} \tag{6}$$

where i, j, and NS are as defined for Eq (4). The reference symbol confusability numbers are stored in a symbol set confusability matrix (SC matrix). An example of an SC matrix is shown in Table III which is derived from the ordered distance matrix in Table II.

Table III
Symbol Set Confusability Matrix

	4	3	2	1
DIR	6.3272	2.1235	1.3701	.6696
NAV	6.7168	3.8611	1.5202	.7534
ITX	6.3859	2.1822	1.4155	.6946
VTC	6.4715	3.7622	1.4213	.7005
MRP	6.9291	4.0734	1.3642	.6696

The RSC in the highest numbered column of the SC matrix represents the confusability of the reference symbol with respect to all other symbols in the set. The RSC in the next highest numbered column represents the confusability with respect to all other remaining symbols in the set after the closest symbol to the reference symbol (nearest neighbor) has been deleted. The RSC in the next column reflects the deletion of the two closest symbols to the reference symbol and so on. The number of the column in the SC matrix corresponds to the number of symbols with which the reference symbol is compared. The confusability of the reference symbol with respect to the other symbols of the set increases as the RSC in the first column increases. However, care must be taken when comparing reference symbol confusability numbers in other columns since the deleted symbols will not follow the same order for all rows. The comparison is only valid when the same symbols have been deleted for each reference symbol.

The symbol set confusability number is the normalized sum of the reference symbol confusability numbers in the first column of the SC matrix and can be used to evaluate the overall suitability or separability of a given symbol set. The SSC for the five-symbol set corresponding to the SC matrix in Table III is 1.641525. The range of SSC is from zero for an ideal symbol set where all distances are two to 999.5 for a symbol set of identical symbols where all distances are zero. Then for two or more given symbol sets, the set with the lower SSC is the most discernible, least confusable symbol set.

Most Confusable Symbol

The ordered distance matrix and the symbol set confusability matrix also provide the basis for two different methods of defining the most confusable symbol. The method using the ordered distance matrix selects the most confusable symbol from one of the pair (or perhaps more) of symbols with the smallest distance in column two. From this symbol pair (or group), the most confusable symbol is that symbol with the smallest distance in column three of the ordered distance matrix. If more than one of these symbols has the same distance in column three, the procedure is repeated for the pair (or group) of symbols by comparing the distances in the next column to the right until a most confusable symbol is determined. If the last column of the ordered distance matrix is checked and more than one symbol remains, then these symbols are identical symbols, and the most confusable symbol is any one of these remaining symbols. The most confusable symbol then represents that symbol of the given symbol set which results in the most errors in a pair-wise discriminability test.

Another method to define the most confusable symbol chooses the symbol with the highest RSC in the first column of the SC matrix. If the reference symbol confusability numbers in the first column for two or more symbols are equal and also the highest, then the rows of the SC matrix corresponding to those symbols are checked for the highest RSC in the second column. If two or more symbols still remain, the procedure is repeated for each succeeding column until a most confusable symbol is determined. If two or more rows of the last column of the SC matrix are checked and the reference symbol confusability numbers are equal, then these symbols are identical symbols and the solution is trivial. The most confusable symbol by this method represents that symbol of the given symbol set which is more easily confused when compared with all other symbols in the set.

For most symbol sets, the two methods described above will yield the same most confusable symbol for the set. However, in certain instances where two symbols in a given set are identical or nearly identical (small Euclidian distance between the symbols) and are quite discernible from all other symbols in the set (relatively large Euclidian distances to the other symbols), it is possible for each method to produce a different most confusable symbol. The five-symbol set shown in Figure 3 is such a symbol set. The first method determines that the ITX symbol is the most confusable symbol, while the second method shows the MRP symbol as the most confusable symbol. The method to use depends upon what the most confusable symbol is to represent. In most cases, the intent of changing the most confusable symbol is to reduce the number of errors associated with the discriminability between symbols. Therefore, normally the most confusable symbol is determined by the first method using the ordered distance matrix.

Symbol Change Theory

Once the most confusable symbol is determined, that symbol is changed in order to achieve a less confusable symbol and reduce the discriminability errors associated with the given symbol set. Three different approaches to change this symbol are investigated for this project. The first approach involves subtracting the Fourier (spatial frequency) components of the nearest neighbor to the most confusable symbol from the weighted Fourier components of the most confusable symbol. The nearest neighbor is that symbol in the second column of the reference row in the ordered distance matrix using the most confusable symbol as the reference symbol. The second and third procedures compare corresponding spatial frequency components of the most confusable symbol and its nearest neighbor. The second method reflects the theory that possibly those corresponding spatial frequency components that are different in value provide the discernibility between the symbols. Therefore, if those components that are different are made more different, then the distance between the two symbols will increase and the symbols will be more discernible. The third method is similar to the second except that the corresponding spatial frequency components that are equal or nearly equal are made different in value while those components that are already different are changed none or very little. This procedure should also increase the distance between the symbols and make them more discernible.

<u>Weighted Component Approach</u>. The weighted Fourier component approach applies the following equation to change the most confusable symbol:

$$FC_{new} = (FC_{old} \times WT - FC_{nbr})/WT \qquad (7)$$

where FC_{new} is the new Fourier component of the most confusable symbol, FC_{old} is the old Fourier component of the most confusable symbol, FC_{nbr} is the corresponding Fourier component of the nearest neighbor to the most confusable symbol, and WT is the weighting factor which determines how much the Fourier component and, therefore, the symbol will change.

As WT increases, there is less change in the Fourier component. For this project, values for WT have been varied between one and 20 with a weighting factor of five providing the most controllable change in the symbol. A weighting factor of more than five makes very little change in the symbol. A weighting factor of less than five, however, makes too large of a change in the Fourier component to interpret the spatial image of the changed symbol which results from the inverse Fourier transform using the filtered, changed Fourier components. If five is used as a weighting factor in Eq (7), a definite, controllable change in the symbol occurs which enables reproduction of the changed symbol.

Separate Different Components. The method which takes those corresponding spatial frequency components of the most confusable symbol and its nearest neighbor and makes more different those components that are already different uses the following symbol change equation:

$$FC_{new} = FC_{old}\,(1 + (FD \times WT)) \qquad (8)$$

where FC_{new} and FC_{old} are as defined in Eq (7), FD is the absolute value of the difference between the values of FC_{old} and FC_{nbr} (defined in Eq (7)), and WT is a weighting factor that determines the amount of change in the Fourier component.

As the weighting factor increases, the change in the spatial frequency component also increases. Weighting factors between one and five have been investigated. A weighting factor of unity shows a negligible change in the symbol and the distance to its nearest neighbor, while a greater weighting factor increases the distance but with no significant change in the symbol. This is caused by a change in those Fourier components that contribute energy to the background at points sufficiently distant from the symbol itself to make no effective change in the symbol although this additional energy does affect the Euclidian distance computation. The result is a very smeared, filtered image of the changed symbol.

<u>Separate Like Components</u>. The third approach which takes the corresponding Fourier components of the most confusable symbol and makes like components different applies the following symbol change equation:

$$FC_{new} = FC_{old}\left(1+\left(\frac{SIGN}{FD \times WT}\right)\right) \quad (9)$$

where FC_{new} and FC_{old} are as defined in Eq (7), FD is as defined in Eq (8) but is set equal to 0.01 in order to keep FC_{new} finite, SIGN is positive if FC_{old} is larger than FC_{nbr} and negative if the reverse is true, and WT is a weighting factor which determines how much the component will change.

The symbol change becomes much more pronounced in this case as the weighting factor decreases. Weighting factors between 100 and 1000 have been investigated during this study with factors over 150 showing no significant change apparent in the shape of the symbol, yet a large increase in distance to the nearest neighbor. This is again caused by the contribution of energy from the changed Fourier components to the background of the symbol. A weighting factor of 100 effectively changes the symbol and increases the distance to the nearest neighbor, however, the corresponding change to the background of the symbol makes the recognition of the changed symbol extremely difficult.

Symbol Vector Movement

The next problem is to determine how much to change a given symbol using any one of Eqs (7), (8), or (9) and how to control the direction of movement of the symbol vector in Fourier space or limit the distance from other symbols in the set while increasing the distance from the nearest neighbor. An arbitrary value of 0.5 is used as a threshold Euclidian distance in the algorithm. This figure is based on reviewing the distances in a number of different symbol sets (mostly those sets consisting of letters and numbers) and seems to be a large enough distance so that discriminability errors should be quite small, if any do exist. The appendix includes a discussion and some results of additional research conducted into the area of determining an error free Euclidian distance. This algorithm changes the most confusable symbol until the distance to its nearest neighbor is equal to or greater than the threshold distance.

The direction in which the changed symbol moves in relation to the other symbols in the set is also important. To provide controlled movement of the symbol, the distances corresponding to the latest change in the symbol are compared to those distances resulting from the previous change. If any of the previous distances are below the threshold and decrease when the symbol is changed, then a symbol other than the nearest neighbor is used in place of the nearest neighbor. This "new neighbor" symbol is the closest symbol to the most confusable symbol disregarding the nearest neighbor symbol. If again any distance decreases that was previously below the threshold, this procedure is repeated until all other symbols are exhausted.

If a direction of movement is found in which all of the distances below the threshold increase with the change, the symbol being used as the nearest neighbor in the symbol change equations is retained as the nearest neighbor symbol. The symbol changes are repeated until the minimum distance from the most confusable symbol to any other symbol in the set is at least as great as the threshold. This requires at least one iteration and is dependent upon the original distance between the reference symbol and its nearest neighbor, the direction of movement of the symbol vector, and the weighting factor used in the symbol change equation.

Symbol Recognition

After changing the Fourier components of the most confusable symbol to improve the number of discriminability errors associated with it, the task of reproducing the "new" symbol is required in order to verify the results. Within the symbol change algorithm, the new Fourier components are used in taking the inverse, discrete, two dimensional Fourier transform of the filtered symbol. Employing a printing subroutine option of PREVIP (Ref 7), the symbol is then reproduced in the spatial domain using the resulting spatial components of the inverse transform.

This completes the algorithm for changing one symbol of a given symbol set. If more than one symbol change is desired, the changed symbol must be redigitized and reentered into the algorithm in place of the old symbol. This procedure can be repeated until all symbols in the input symbol set have distances from each other at least as great as the threshold distance. The number of times that this procedure must be repeated is dependent upon the number of different pairs of symbols in the original symbol set that had distances below the threshold and the difference of each of these distances from the threshold.

IV. Results and Discussion

The symbol set confusability and symbol change algorithm discussed in Chapter III was used to compare each of the four symbol sets illustrated in Figure 2 both before and after one symbol deletion and one symbol change. Furthermore, symbol set one was changed by this algorithm so that the Euclidian distances between all possible pairs of symbols were increased to at least the 0.5 threshold distance. This threshold distance is an arbitrary value. The 0.5 value was chosen since larger distances are associated with fewer discriminability errors (Ref 10:20).

Symbol Set Confusability Number Results

The symbol set confusability number (SSC) for each of the four symbol sets was computed to determine which of the original symbol sets was associated with the fewest number of discriminability errors. The SSC represents the overall suitability or separability of a symbol set with the low numbers being associated with fewer discriminability errors. The ideal SSC would be zero for an error free symbol set. From Table IV, the results indicated that symbol set three was the best symbol set initially since the SSC of 0.822801 was lower than that for the other symbol sets. After deleting the most confusable symbol in each set, the effect on the

Table IV
Symbol Set Confusability Numbers (SSC)

Symbol Set	Initial SSC	SSC after Symbol Deletion	SSC after Symbol Change
1	.910207	.883114	.891442
2	1.102955	1.072673	1.056545
3	.822801	.816180	.842880
4	1.172562	1.144084	1.118844

SSC indicated an improvement in the overall separability of each set.

Upon changing the most confusable symbol in each set, a comparison of the resulting symbol set confusability numbers indicated an improvement in sets one, two and four. However, the new SSC for symbol set three indicated a slightly worse symbol set than the original set three although the minimum distance from the most confusable symbol increased from 0.3025 to 0.4074 after the change. The reason for the increase in the SSC was that the changed symbol was closer to 15 of the other symbols in the set after the change. This did not affect the direction of movement of the change since the original distances to these 15 symbols were all above the threshold and since the original distances to the remaining two symbols increased with the change. However, after both deleting and changing the most confusable symbol, symbol set three still had the lowest SSC which indicated that it was the set with the fewest number of discriminability errors.

Most Confusable Symbol Changes

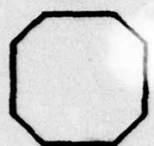
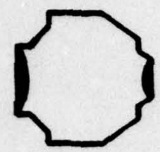
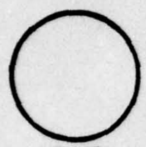
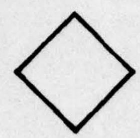
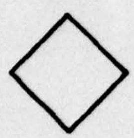
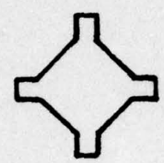
The ordered distance matrix method was used to determine the most confusable symbol for the results illustrated in this chapter since the change in the most confusable symbol was intended to reduce the number of errors associated with pair-wise discriminability.

The weighted Fourier component approach which employs Eq (7) was used to change the most confusable symbol in all four sets. This approach provided the most significant, controllable change in the symbol such that the symbol could be reproduced and reentered into the data set.

Effect of One Symbol Change. Table V shows the results of one symbol change on each of the four given symbol sets. The most confusable symbol of each set is illustrated in the table, its nearest neighbor symbol is identified, and the distance to that nearest neighbor is also shown. The changed symbol in reproduced form is depicted to the right of the original symbol with the corresponding nearest neighbor and distance.

Although the most confusable symbols were significantly changed and the distances to the nearest neighbor increased for each of the symbol sets, the new distances were still below the 0.5 threshold after repro-

Table V
Results of One Symbol Change

Symbol Set	Most Confusable Symbol	Nearest Neighbor	Nearest Neighbor Distance	Changed Symbol	Nearest Neighbor	Nearest Neighbor Distance
1	 NAV	RAD	.1892		RAD	.3418
2	 VTC	TAC	.1277		TAC	.4200
3	 TAC	ITX	.3025		APT	.4074
4	 TGT	VOR	.2668		AAA	.5192

duction of the symbol. This indicated that the changed symbol was not accurately reproduced since the distance to any other symbol in the set had been increased to 0.5 or more by the algorithm. This inaccuracy was caused by the lack of the higher frequency components which contained the detailed information relating to the shape of the changed symbol. Only the basic form of the changed symbol was available since only the filtered spatial frequency components were used in the symbol change algorithm. In reconstructing the symbol, the details in the shape such as sharp edges versus curved edges or pointed protrusions versus rounded ones were not accurately reproduced.

To explain this point further, an application of the symbol change algorithm on the most confusable symbol of symbol set one is demonstrated through a series of illustrations. The original, digitally represented symbol is depicted in Figure 4 (black on a white background). Figure 5 shows the original symbol after it had been filtered by the modulation transfer function of PREVIP. For easier viewing, the filtered symbols are shown as white on a black background. After five symbol changes using Eq (7), the "new" symbol in its filtered form appeared as in Figure 6 with a corresponding distance to its nearest neighbor (RAD) of 0.5110. The illustration in Figure 7 shows the reproduced, digitally represented symbol after an arbitrary threshold was applied to the

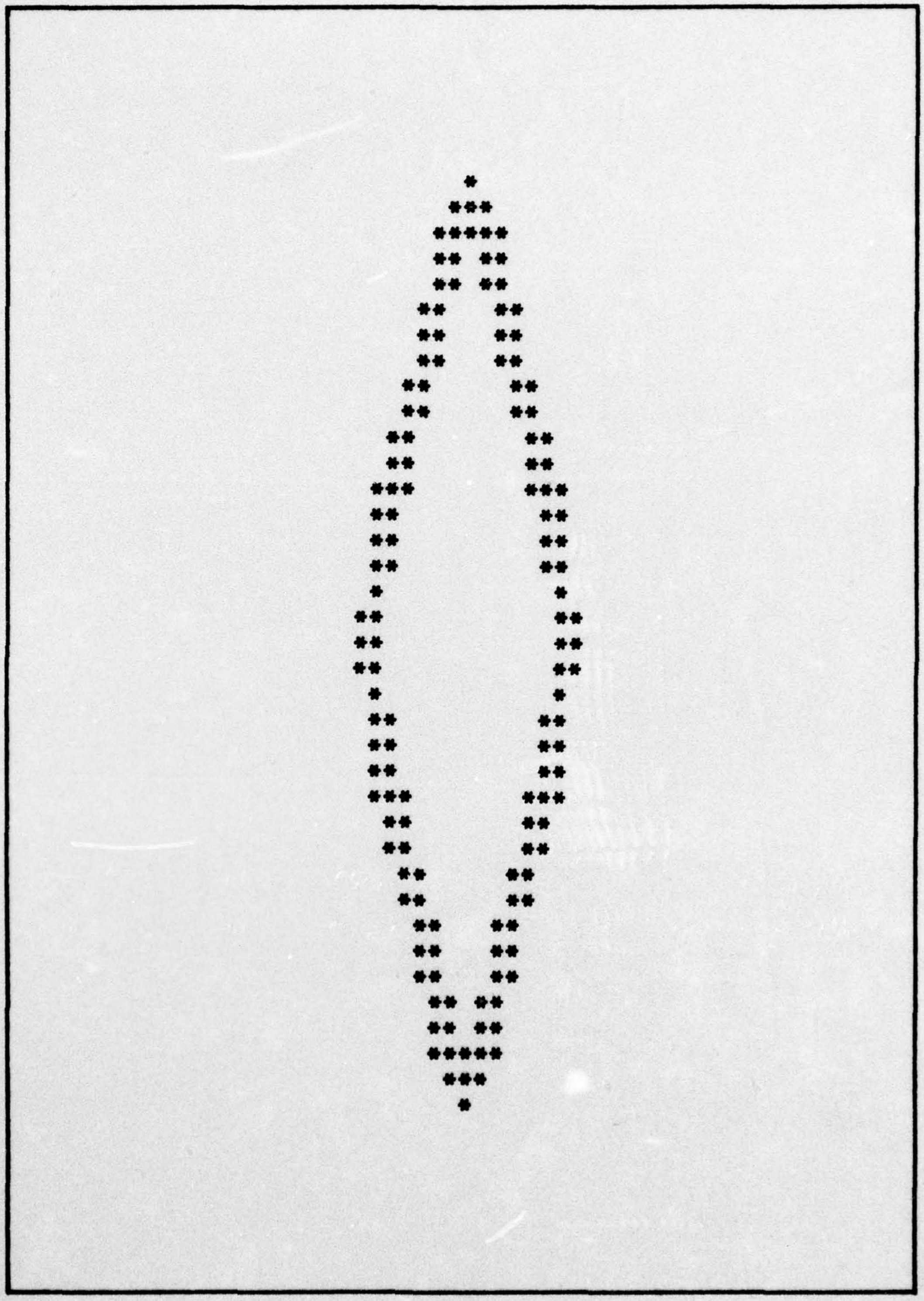


Figure 4. Digitized NAV Symbol

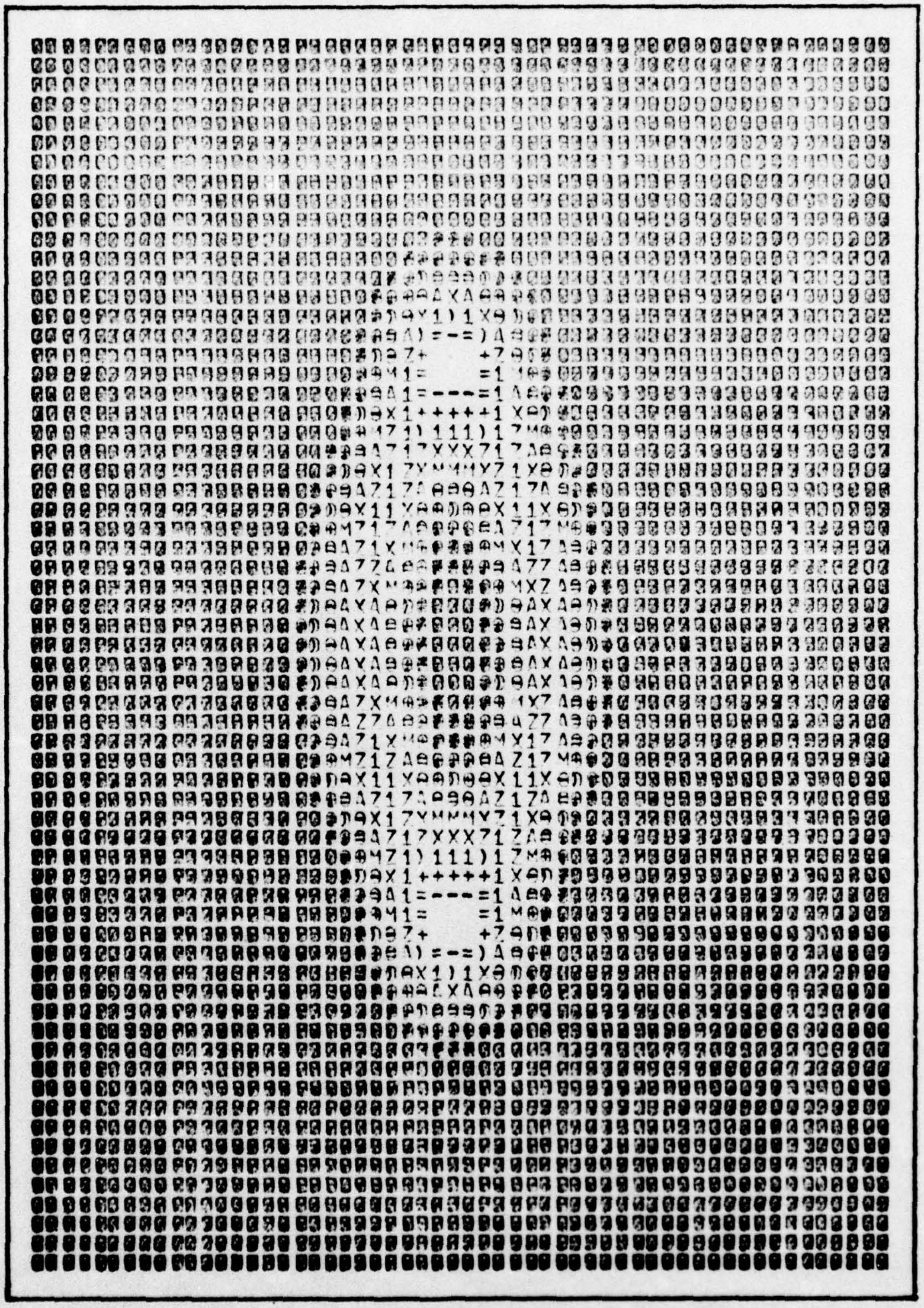


Figure 5. Filtered NAV Symbol

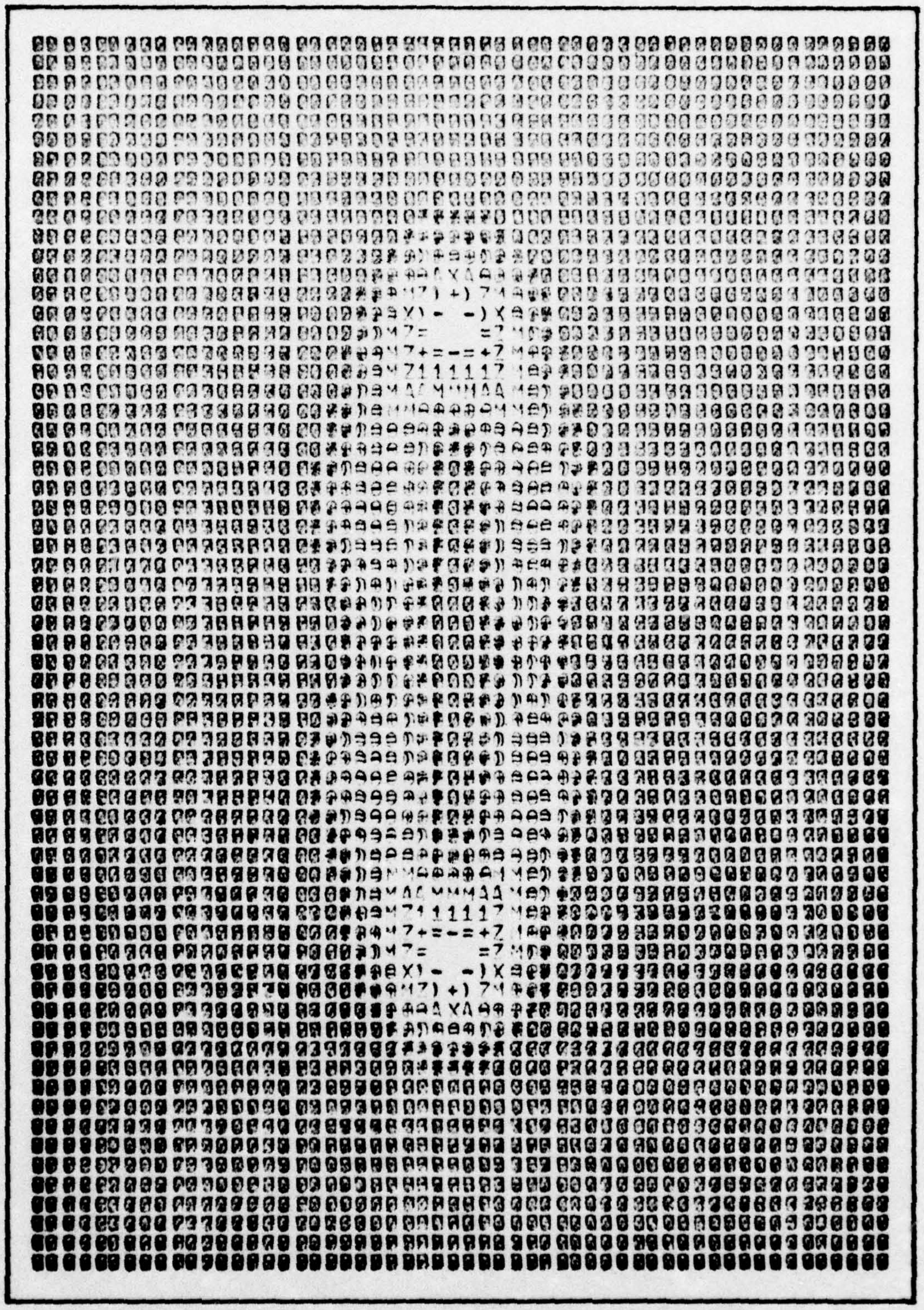


Figure 6. NAV Symbol after Symbol Change Algorithm

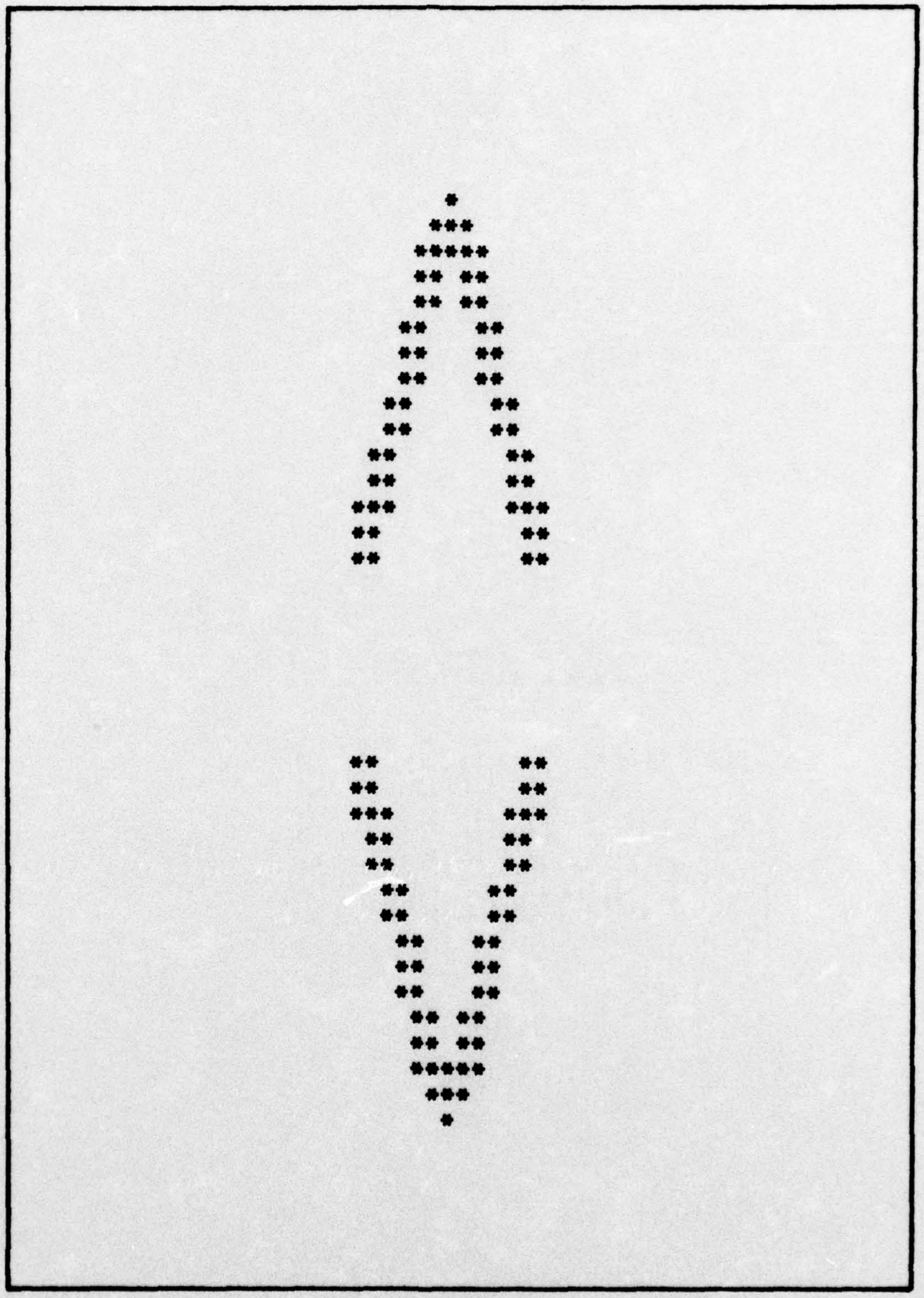


Figure 7. "New" Digitized NAV Symbol

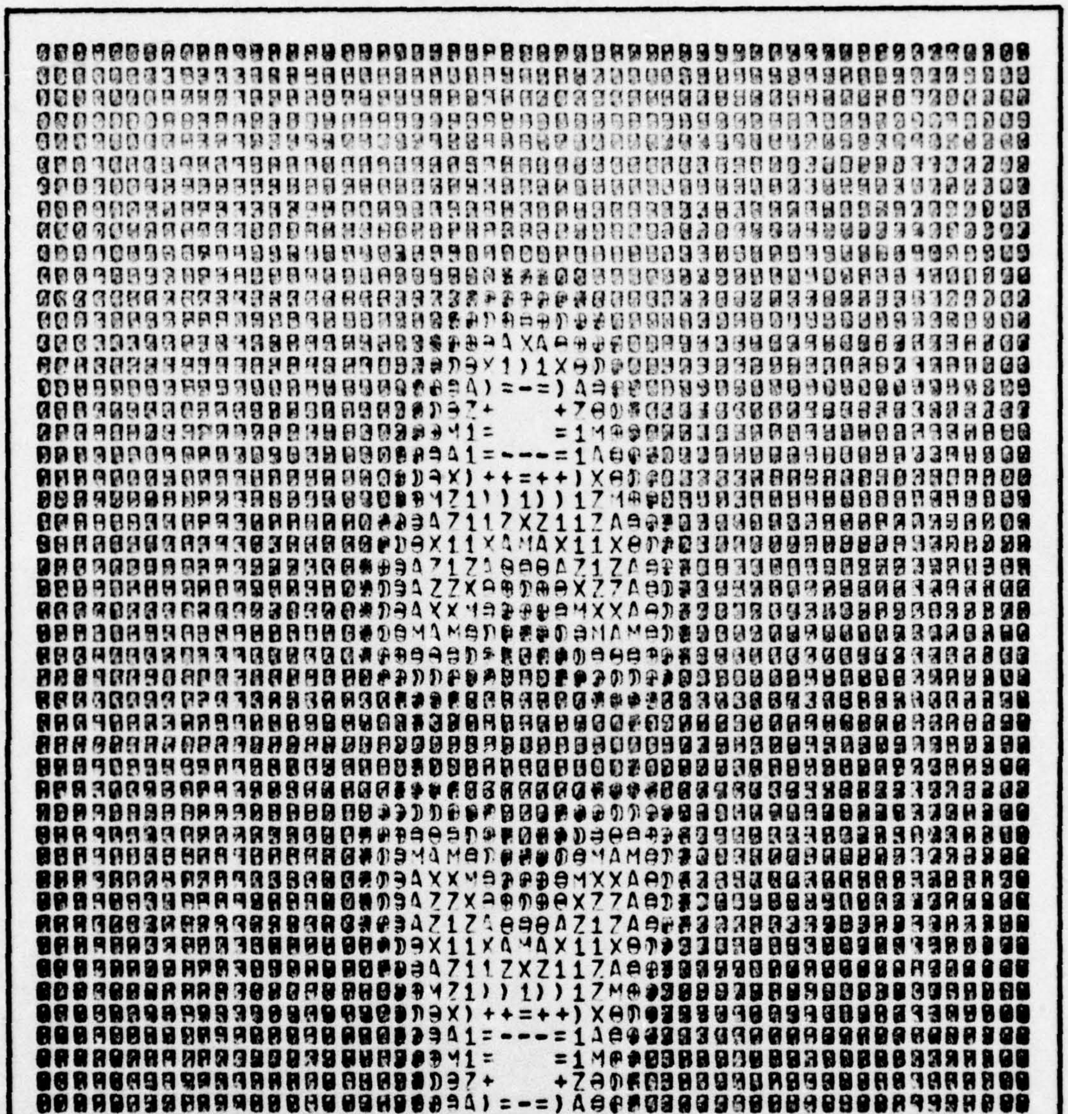


Figure 8. "New" Filtered NAV Symbol

Table VI
Comparison of Distances from NAV Symbol

Changed Most Confusable Symbol		Digitally Reproduced Symbol	
Symbol	Distance	Symbol	Distance
RAD	.5110	RAD	.3418
WXS	.5360	WXS	.4523
APT	.6160	WXR	.4854
WXR	.7152	APT	.6128
VOR	.7418	TAC	.6679
TAC	.7540	VOR	.7017
ORP	.7660	ORP	.7324
AAM	.8397	AAM	.7529
HDG	.8506	HDG	.7935
EAC	.8901	VTC	.8027
VTC	.8956	EAC	.8221
TGT	.8984	TGT	.8454
SAM	.9096	ITX	.8554
ITX	.9142	SAM	.8705
AAA	.9178	AAA	.9142
FAC	.9844	MRP	.9634
MRP	1.0516	FAC	1.0121

basic form of the symbol shown in Figure 6. To determine how accurately the basic form of this "new" symbol was interpreted, the filtered version of the reproduced symbol, depicted in Figure 8, should be compared to the basic form in Figure 6. Any errors in interpretation are easily seen. Also, Table VI indicates how accurately the changed symbol was reproduced if one compares the distances and the order of the symbols between the column corresponding to the changed symbol and the column corresponding to the reproduced symbol. Ideally, the order of symbols and the distances to those symbols from the reproduced symbol would be identical to those of the changed symbol. However, the resulting differences between the symbols are caused by the loss of the higher frequency detail of the changed symbol due to filtering.

Complete Symbol Set Change. The results of repeated applications of the symbol change algorithm on symbol set one are shown in Figure 9 and in Table VIII. Only eight symbols of the original 18 have been altered in order to achieve the minimum Euclidian distance of 0.5 between all possible pairs of symbols. Figure 9 shows these eight symbols with each new symbol depicted to the right of the original symbol. The original distance matrix of symbol set one is shown in Table VII. The data in this table should be compared with that of Table VIII, which contains the distance matrix of set one after the eight symbols have been changed. Only those distances in the columns and rows corresponding to the eight changed symbols are affected. Those distances that were below the threshold in Table VII have been increased to at least 0.5 in Table VIII.

These results indicate that the symbol change algorithm made significant changes in the input symbol set with a corresponding increase in the distances between

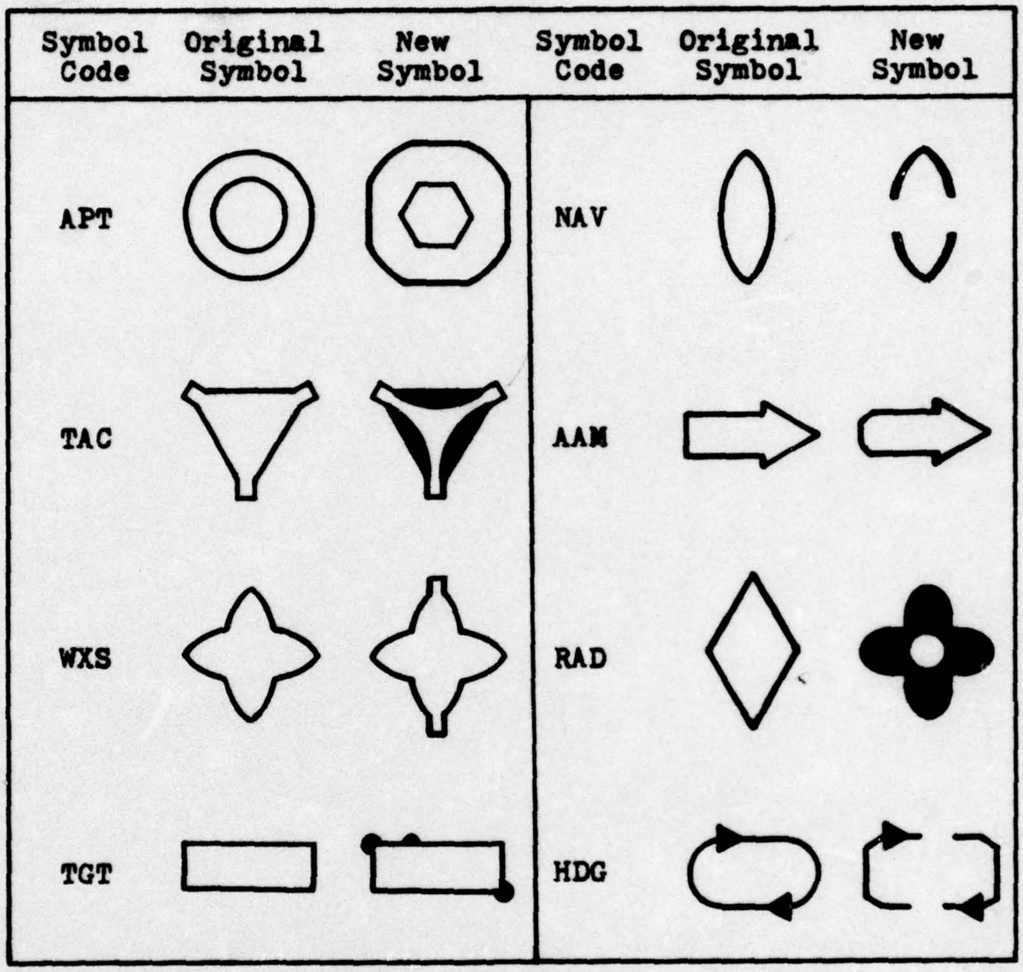


Figure 9. Changed Symbols of Symbol Set One

the symbols. The algorithm also lowered the symbol set confusability number from 0.910207 to 0.805384. Therefore, the number of discriminability errors associated with the symbol set should be fewer after the changes. Thus, the application of the symbol change algorithm shows a definite improvement in the original symbol set.

Table VII
Distance Matrix for Symbol Set One

	APT	VOR	VTC	TAC	WXS	ORP	WXR	MRP	EAC	AAA	SAM	TGT	NAV	AAM	RAD	ITX	HDG	FAC
APT	.00	.48	.68	.53	.48	.56	.75	.99	.63	.74	.63	.53	.59	.57	.59	.87	.50	.73
VOR	.48	.00	.69	.57	.56	.66	.83	1.11	.69	.88	.79	.67	.74	.67	.75	.96	.53	.88
VTC	.68	.69	.00	.25	.77	.94	.89	1.19	.89	.99	.93	.84	.81	.81	.78	.81	.78	1.03
TAC	.53	.57	.25	.00	.60	.78	.78	1.08	.78	.84	.78	.73	.65	.68	.62	.79	.67	.87
WXS	.48	.56	.77	.60	.00	.69	.49	1.01	.80	.79	.60	.67	.44	.56	.46	.95	.61	.76
ORP	.56	.66	.94	.78	.69	.00	.88	.87	.73	.79	.81	.81	.69	.77	.67	1.02	.77	.84
WXR	.75	.83	.89	.78	.49	.88	.00	.93	.89	.96	.81	.85	.52	.71	.50	.93	.83	1.05
MRP	.99	1.11	1.19	1.08	1.01	.87	.93	.00	.87	.90	1.00	1.06	.88	.96	.83	1.11	1.08	1.06
EAC	.63	.69	.89	.78	.80	.73	.89	.87	.00	.89	.86	.67	.83	.67	.80	.97	.66	.95
AAA	.74	.88	.99	.84	.79	.79	.96	.90	.89	.00	.63	.72	.83	.66	.78	.96	.79	.58
SAM	.63	.79	.93	.78	.60	.81	.81	1.00	.86	.63	.00	.54	.79	.50	.73	.99	.59	.59
TGT	.53	.67	.84	.73	.67	.81	.85	1.06	.67	.72	.54	.00	.84	.38	.80	.97	.32	.80
NAV	.59	.74	.81	.65	.44	.69	.52	.88	.83	.83	.79	.84	.00	.74	.19	.85	.80	.89
AAM	.57	.67	.81	.68	.56	.77	.71	.96	.67	.66	.50	.38	.74	.00	.69	.89	.41	.79
RAD	.59	.75	.78	.62	.46	.67	.50	.83	.80	.78	.73	.80	.19	.69	.00	.84	.78	.83
ITX	.87	.96	.81	.79	.95	1.02	.93	1.11	.97	.96	.99	.97	.85	.89	.84	.00	.97	1.13
HDG	.50	.53	.78	.67	.61	.77	.83	1.08	.66	.79	.59	.32	.80	.41	.78	.97	.00	.81
FAC	.73	.88	1.03	.87	.76	.84	1.05	1.06	.95	.58	.59	.80	.89	.79	.83	1.13	.81	.00

Table VIII
Distance Matrix for New Symbol Set One

	APT	VOR	VTC	TAC	WXS	ORP	WXR	MRP	EAC	AAA	SAM	TGT	NAV	AAM	RAD	ITX	HDG	FAC
APT	.00	.54	.66	.60	.59	.54	.84	.98	.59	.77	.71	.61	.75	.64	.85	.86	.60	.80
VOR	.54	.00	.69	.67	.55	.66	.83	1.11	.69	.88	.79	.72	.72	.68	.81	.96	.55	.88
VTC	.66	.69	.00	.54	.79	.94	.89	1.19	.89	.99	.93	.92	.82	.80	.94	.81	.82	1.03
TAC	.60	.67	.54	.00	.59	.78	.73	1.04	.87	.77	.71	.84	.70	.65	.65	.90	.76	.79
WXS	.59	.55	.79	.59	.00	.67	.58	1.04	.79	.74	.53	.67	.59	.53	.52	.95	.58	.66
ORP	.54	.66	.94	.78	.67	.00	.88	.87	.73	.79	.81	.82	.77	.77	.87	1.02	.78	.84
WXR	.84	.83	.89	.73	.58	.88	.00	.93	.89	.96	.81	.90	.51	.63	.51	.93	.89	1.05
MRP	.98	1.11	1.19	1.04	1.04	.87	.93	.00	.87	.90	1.00	1.09	.96	.93	1.00	1.11	1.11	1.06
EAC	.59	.69	.89	.87	.79	.73	.89	.87	.00	.89	.86	.67	.86	.70	1.00	.97	.71	.95
AAA	.77	.88	.99	.77	.74	.79	.96	.90	.89	.00	.63	.76	.95	.67	.92	.96	.81	.58
SAM	.71	.79	.93	.71	.53	.81	.81	1.00	.86	.63	.00	.58	.91	.51	.71	.99	.61	.59
TGT	.61	.72	.92	.84	.67	.82	.90	1.09	.67	.76	.58	.00	.95	.55	.90	1.01	.53	.80
NAV	.75	.72	.82	.70	.59	.77	.51	.96	.86	.95	.91	.95	.00	.72	.73	.89	.84	1.06
AAM	.64	.68	.80	.65	.53	.77	.63	.93	.70	.67	.51	.55	.72	.00	.69	.87	.51	.81
RAD	.85	.81	.94	.65	.52	.87	.51	1.00	1.00	.92	.71	.90	.73	.69	.00	1.12	.87	.86
ITX	.86	.96	.81	.90	.95	1.02	.93	1.11	.97	.96	.99	1.01	.89	.87	1.12	.00	.99	1.13
HDG	.60	.55	.82	.76	.58	.78	.89	1.11	.71	.81	.61	.53	.84	.51	.87	.99	.00	.78
FAC	.80	.88	1.03	.79	.66	.84	1.05	1.06	.95	.58	.59	.80	1.06	.81	.86	1.13	.78	.00

V. Conclusions and Recommendations

The results in Chapter IV are based on the symbol change algorithm discussed in Chapter III. This chapter reviews this symbol change algorithm and recommends some areas for further research. Table IX summarizes the symbol change algorithm as a part of the predictor of human visual performance (PREVIP) algorithm (Ref 7).

Symbol Change Algorithm

The symbol change algorithm is designed to produce an optimally human-separable symbol set by changing one or more symbols within a given set of symbols in order to reduce the number of discriminability errors caused by similarities in the basic form or shape of certain symbols within the set. The symbols that cause these confusion errors are changed one at a time until the Euclidian distance between each possible pair of symbols is greater than or equal to an arbitrary threshold distance. A threshold distance of 0.5 is used for the results of Chapter IV. This value is chosen to approximate an error free distance and is based on a review of a number of different alphanumeric symbol sets. If all of the distances are increased to 0.5 or greater, the number of discriminability errors associated with the symbol set should decrease.

Table IX
Symbol Change Algorithm

1. Simulate each symbol in digital form by sampling the intensity pattern of the symbol at a finite number of evenly spaced points or by a comparable method.

2. Transform each digitally represented symbol into the Fourier domain.

3. Energy normalize each transformed symbol.

4. Low-pass filter each transformed symbol.

5. Compute the Euclidian distance between each possible pair of filtered, transformed symbols.

6. Compute the symbol set confusability number (SSC) from the ordered distance matrix according to Eq (4).

7. Determine the most confusable symbol in the set from the method using the ordered distance matrix.

8. Change the most confusable symbol by controlled manipulation of the Fourier components according to Eq (7), the weighted component approach.

9. Obtain a spatial image of the changed symbol by taking the inverse Fourier transform using the changed, filtered Fourier components.

10. Digitally reproduce the changed symbol from the spatial image.

11. Repeat steps two through ten until the minimum Euclidian distance between each possible pair of symbols is greater than or equal to the threshold Euclidian distance.

The symbol that is changed is the most confusable symbol in the symbol set. For the results in Chapter IV, the most confusable symbol is determined from the method using the ordered distance matrix rather than the symbol set confusability matrix. This method is used since the most confusion is between those symbols that are separated by the smallest Euclidian distance.

The Fourier components of the most confusable symbol are changed by the weighted component approach using Eq (7). This approach provides the most controllable change in the shape of the symbol so that the resulting spatial image derived from the changed Fourier components is easily interpreted to enable digital simulation of the changed symbol. This new symbol in digital form then replaces the original symbol and the algorithm is repeated until the minimum Euclidian distance threshold is achieved for all distances in the symbol set.

The symbol change algorithm effectively improves a given symbol set based on the increase in the Euclidian distances. The symbol set confusability number (SSC), which is computed from the ordered distance matrix according to Eq (4), also provides a measure of the relative separability of a symbol set. A symbol set with a lower SSC has fewer discriminability errors associated with it. Based on the Euclidian distances of the four symbol sets used in this study and the results of Table IV, an SSC of less than one implies that the given symbol set is a comparatively good symbol set that would produce relatively few errors.

Recommendations for Further Research

Although a symbol set with a low SSC implies relatively few confusion errors, the actual error rate in any arbitrary application cannot be predicted with our present knowledge. Also, the hypothesized reduction in errors after employing the symbol change algorithm is unknown. Further research is needed to verify these results with actual data obtained by testing human subjects as Goble did for discriminability errors (Ref 10) but in this case using symbol sets both before and after application of the symbol change algorithm. Human performance data would provide a more definite relationship between the number of errors and the SSC so as to enable the number of errors to be determined from each SSC.

The number of errors is also related to the Euclidian distance between symbols. This is an inverse relationship, however, the minimum distance for an error free symbol set requires further study. Although a threshold distance of 0.5 was used for this particular project, this threshold value is still an arbitrary number that is intended to approximate an error free distance. Is there, in fact, one such universal error free distance, or is this distance dependent upon the number of symbols, the type of symbols (alphanumeric, geometric, or size variant, for example), or other factors? It is this question that is addressed in the appendix but still requires more investigation in order to reach an answer.

Another question for consideration involves the definition of the most confusable symbol. Would more errors be eliminated by choosing a different most confusable symbol? A second definition was discussed in Chapter III. Another definition for this symbol would be to select the nearest neighbor to the currently defined most confusable symbol. It may be possible to reduce the errors even further by selecting a different order for the symbols to be changed. Human performance data may verify that another definition of the most confusable symbol may better optimize a symbol set.

A different direction of movement of the most confusable symbol in effecting the symbol change may also reduce the number of errors associated with the completely changed symbol set. One method would weight the Fourier components of all or part of the neighboring symbols in the set. This method would slightly alter symbol change Eq (7) by redefining FC_{nbr}. The value of FC_{nbr} could be the average of the corresponding Fourier components for all or part of the neighboring symbols. FC_{nbr} could also be the average of these components which are weighted by a factor equal to the reciprocal of the rank of each neighboring symbol where the rank is one for the nearest neighbor, two for the next closest symbol and so on. A weighting factor corresponding to one minus the distance between the symbols might also be used.

The symbol change approach which separates like Fourier components according to Eq (9) also merits further investigation. If the energy that is contributed to the spatial background by the changed Fourier components can be controlled or eliminated, this approach may provide a larger reduction in the number of discriminability errors.

Another related area for possible future study deals with the reproduction of the changed symbol. Can the changed symbol be interpreted from the spatial image more accurately by some other means? It may be possible to redigitize the changed symbol from the spatial image created from the inverse Fourier transform of the filtered components. One suggested technique would restrict the changes in the symbol to only neighboring points or regions of the original intensity pattern or the previously changed intensity pattern. Another technique would not allow a change if the correlation of the changed spatial image with the original or previously changed spatial image did not exceed some minimum threshold.

Although the approach used for this symbol change algorithm is thought to be the best method, additional data from other symbol sets and human performance tests may offer new insight such that other alternatives might further reduce the number of discriminability errors in a symbol set. These recommended areas for further research may later provide a more efficient algorithm to generate an optimally human-separable symbol set.

Bibliography

1. Andrews, Harry C. Computer Techniques in Image Processing. New York: Academic Press, 1970.

2. Attneave, Fred. Applications of Information Theory to Psychology: A Summary of Basic Concepts, Methods, and Results. New York: Henry Holt and Company, Inc., 1959.

3. Carl, J.W. Generalized Harmonic Analysis for Pattern Recognition: A Biologically Derived Model. Unpublished thesis. Wright-Patterson Air Force Base, Ohio: Air Force Institute of Technology, September 1969. AD 862441.

4. Carl, J.W. and Hall, C.F. "The Application of Filtered Transforms to the General Classification Problem." IEEE Transactions on Computers, C-21: 785-790 (July 1972).

5. Cooley, J.W. and Tukey, J.W. "An Algorithm for Machine Calculation of Complex Fourier Series." Mathematics of Computation, 19: 297-301 (April 1965).

6. Defense Mapping Agency Aerospace Center. Type and Symbol Catalog for Flight Information Publications and Associated Products (Third Edition). St. Louis Air Force Station, Missouri: Defense Mapping Agency Aerospace Center, March 1974.

7. Gagnon, Roger A. A Predictor of Human Visual Performance at Form Discrimination Tasks (PREVIP). Unpublished technical report. Wright-Patterson Air Force Base, Ohio: Aerospace Medical Research Laboratory, 1976.

8. Garner, W.R. and Hake, H.W. "The Amount of Information in Absolute Judgements." Psychological Review, 58: 446-459 (1951).

9. Ginsburg, A.P., et al. "Psychological Aspects of a Model for the Classification of Visual Images." Proceedings 1972 International Congress of Cybernetics: 92-103 (1973).

10. Goble, Larry G. Filtered 2-Dimensional Discrete Fourier and Walsh Transform Correlation with Recognition Errors and Similarity Judgements. Dissertation. Ann Arbor, Michigan: University of Michigan, 1975.

11. Granlund, G.H. "Fourier Preprocessing for Hand Print Character Recognition." IEEE Transactions on Computers, C-21: 195-201 (February 1972).

12. Honigfeld, A.R. Radar Symbology: A Literature Review. Technical Memorandum 14-64. Aberdeen Proving Grounds, Maryland: Human Engineering Laboratories, 1964. AD 461180.

13. Kabrisky, Matthew. A Proposed Model for Visual Information Processing in the Human Brain. Urbana, Illinois: University of Illinois Press, 1966.

14. Kama, W.N., et al. Display Design for Electronic Countermeasures Application -- Symbol Coding and Range Presentation. AMRL-TR-73-42. Wright-Patterson Air Force Base, Ohio: Aerospace Medical Research Laboratory, November 1973. AD 781032.

15. Muller, P.F., et al. The Symbolic Coding of Information on Cathode Ray Tubes and Similar Displays. WADC TR 55-375. Wright-Patterson Air Force Base, Ohio: Wright Air Development Center, October 1955.

16. Radoy, Charles H. Pattern Recognition by Fourier Series Transformations. Unpublished thesis. Wright-Patterson Air Force Base, Ohio: Air Force Institute of Technology, March 1967.

17. Tallman, Oliver H.,II. The Classification of Visual Images by Spatial Filtering. Dissertation. Wright-Patterson Air Force Base, Ohio: Air Force Institute of Technology, June 1969.

18. Tallman, Oliver H.,II. Processing of Visual Imagery by an Adaptive Model of the Visual System: Its Performance and Its Significance. AMRL-TR-70-45. Wright-Patterson Air Force Base, Ohio: Aerospace Medical Research Laboratory, November 1970.

19. Thomas, John K. A Method for Predicting Human Performance of Certain Symbol Recognition Tasks. Unpublished thesis. Wright-Patterson Air Force Base, Ohio: Air Force Institute of Technology, December 1973.

20. Thomson CSF. EADI Maintenance Book 1. Paris, France: AVS Division, Thomson CSF, 1973.

21. Van Cott, H.P., et al. Human Engineering Guide to Equipment Design (Revised Edition). Washington, D.C.: American Institute for Research, 1972.

Appendix

Minimum Error Free Distance

The problem of selecting a suitable threshold Euclidian distance for the symbol change algorithm led to the possibility of discovering some minimum distance with which no confusion errors would occur. It would be desirable to change the minimum number of symbols in a given symbol set as is required to increase the Euclidian distance between all possible pairs of symbols so that the symbol set would cause no discriminability errors. Some additional research and results on this aspect of symbology standardization are discussed and presented in this appendix.

The Euclidian distance calculations for symbol set two (see Figure 2) yielded the interesting result that the pair of symbols with the largest distance between them (1.0444) were the two squares NAV and RAD. The only difference in these two symbols was the size. The NAV symbol was a 36x36 square on the 64x64 point grid while the RAD symbol was an 18x18 square.

This size variation has previously been used in visual coding schemes and a maximum and recommended number of symbols or coding steps which a human can differentiate has been linked with various error rates (Ref 21:69-77). Various coding techniques were investigated by a research group headed by Muller in 1955 (Ref 15). It was believed that the data gathered by Muller would provide some measure between the number of errors of the size variant coding techniques and the Euclidian distances. Although the appropriate information was originally recorded by Muller, the number of errors between pairs of symbols was not included in his report. However, he did report a maximum number and a recommended number of symbols for each coding technique and an associated error rate using the Garner-Hake method of measuring the accuracy of stimulus identification in informational terms (Ref 2:42-67 and Ref 8). Two of these coding techniques were selected for study in an effort to arrive at a minimum error free Euclidian distance.

Magnitude coding and area coding (Ref 21:73-75) were investigated using Gagnon's predictor of human visual performance (PREVIP) algorithm to generate the distance matrices (Ref 7). A field of view of 0.63 degrees by 0.63 degrees and a size of interest of 0.25 cycles per degree were the input parameters used for each of six symbol sets. The symbols were digitized on a 64x64 point grid and were confined in this grid to a 40x40 point area which corresponded to 30 minutes of visual angle by applying the field of view and size of interest input parameters. The largest symbols were a 40 point diameter circle which represented Muller's area coded circular blip of 30 minutes of visual angle (Ref 15:11-12 and Ref 21:75) as well as his magnitude coded ellipse of unit axis ratio (Ref 15:9-10 and Ref 21:74) and a 40x40 square which represented the largest symbol in four additional magnitude and area coded symbol sets. The latter four symbol sets were chosen because the resolution on the 64x64 point grid could not accurately represent circles and ellipses in digital form, especially for smaller diameters and axis ratios. The square provided a similar starting point in size of visual angle for the largest symbol in each set but, as one or both dimensions were varied as a percentage of the 40 point maximum dimension, the smaller symbols were simulated more accurately than similarly smaller symbols in the circular or elliptical sets.

These six symbol sets are depicted in Figures 10 through 15 and the corresponding distance matrices are tabulated in Tables X through XV. The distances were larger between the middle symbols of each set than the distances between the symbols at the end points. The average distance between symbols A and B, B and C, C and D, and D and E in the three five-symbol sets was 0.6835 while that, following a similar progression, in the three eight-symbol sets was 0.4452 which yielded an average distance of 0.5968 (four distances in the five-symbol sets and seven in the eight-symbol sets). The average of all the distances in the three five-symbol sets was 0.9300 while that for the three eight-symbol sets was 0.6895 yielding an overall average distance of 0.8667 (ten distances in each five-symbol set and 28 in each eight-symbol set).

Thus, in conclusion, both average distances of 0.5968 and 0.8667 are greater than the 0.5 threshold distance used in the symbol change algorithm. Since errors are assumed to exist in these six symbol sets based on Muller's results, does the threshold distance of 0.5 change the symbol enough? Should a higher threshold distance be used and, if so, will it significantly reduce the number of errors even further? Can each Euclidian distance be associated with a definite number of errors or is the number of errors dependent upon the number of symbols in the set and/or the type of coding used? And, finally, is there a universal error free distance? It would be useful if some or all of these questions could be answered, and this appendix is written to provide some data as a base for approaching these answers.

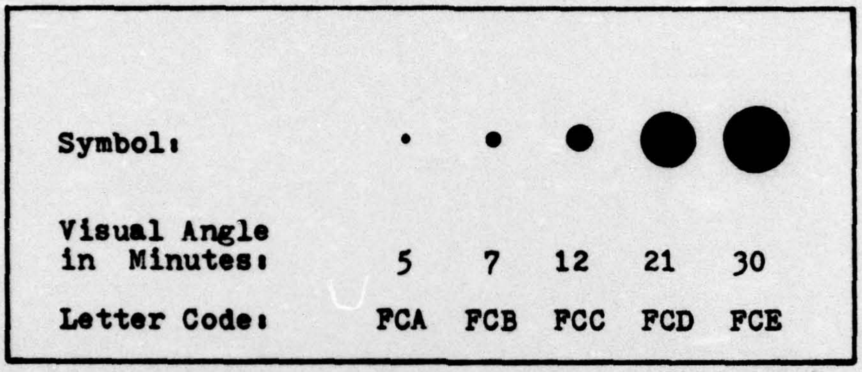


Figure 10. Five-Symbol Set of Circular Blips

Table X
Distance Matrix for
Five-Symbol Set of Circular Blips

	FCA	FCB	FCC	FCD	FCE
FCA	.0000	.2139	.8226	1.2474	1.2763
FCB	.2139	.0000	.6990	1.2677	1.3353
FCC	.8226	.6990	.0000	1.0793	1.3171
FCD	1.2474	1.2677	1.0793	.0000	.8759
FCE	1.2763	1.3353	1.3171	.8759	.0000

Symbol:

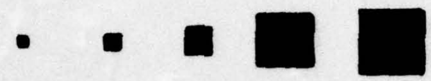
Axis Ratio:	0.00	0.05	0.11	0.21	0.37	0.58	0.82	1.00
Letter Code:	ELA	ELB	ELC	ELD	ELE	ELF	ELG	ELH

Figure 11. Eight-Symbol Set of Ellipses

Table XI
Distance Matrix for Eight-Symbol Set of Ellipses

	ELA	ELB	ELC	ELD	ELE	ELF	ELG	ELH
ELA	.0000	.2445	.2042	.4923	.7508	.7528	.6886	.8220
ELB	.2445	.0000	.1871	.6276	.9178	.9230	.8593	.9781
ELC	.2042	.1871	.0000	.4729	.8245	.8545	.7886	.9102
ELD	.4923	.6276	.4729	.0000	.5663	.7214	.6620	.7987
ELE	.7508	.9178	.8245	.5663	.0000	.5682	.6434	.7797
ELF	.7528	.9230	.8545	.7214	.5682	.0000	.5245	.7140
ELG	.6886	.8593	.7886	.6620	.6434	.5245	.0000	.5144
ELH	.8220	.9781	.9102	.7987	.7797	.7140	.5144	.0000

Symbol:



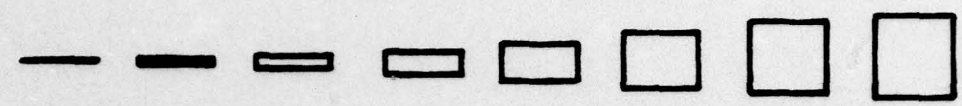
Visual Angle in Minutes :	5	7	12	21	30
Letter Code:	FSA	FSB	FSC	FSD	FSE

Figure 12. Five-Symbol Set of Square Blips

Table XII
Distance Matrix for
Five-Symbol Set of Square Blips

	FSA	FSB	FSC	FSD	FSE
FSA	.0000	.2459	.9176	1.2789	1.3027
FSB	.2459	.0000	.7851	1.2844	1.3461
FSC	.9176	.7851	.0000	1.0291	1.2787
FSD	1.2789	1.2844	1.0291	.0000	.8648
FSE	1.3027	1.3461	1.2787	.8648	.0000

Symbol:

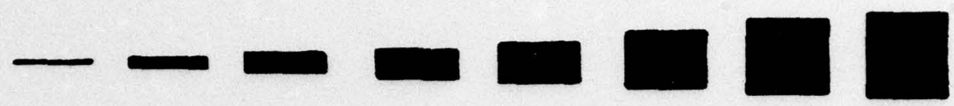


Height Ratio:	0.00	0.05	0.11	0.21	0.37	0.58	0.82	1.00
Letter Code:	RTA	RTB	RTC	RTD	RTE	RTF	RTG	RTH

Figure 13. Eight-Symbol Set of Rectangles

Table XIII
Distance Matrix for Eight-Symbol Set of Rectangles

	RTA	RTB	RTC	RTD	RTE	RTF	RTG	RTH
RTA	.0000	.2445	.2142	.5221	.8003	.8046	.7753	.7638
RTB	.2445	.0000	.1856	.6489	.9636	.9713	.9373	.9243
RTC	.2142	.1856	.0000	.4797	.8579	.8881	.8537	.8398
RTD	.5221	.6489	.4797	.0000	.5953	.7553	.7334	.7165
RTE	.8003	.9636	.8579	.5953	.0000	.5875	.7070	.6920
RTF	.8046	.9713	.8881	.7553	.5875	.0000	.6109	.6688
RTG	.7753	.9373	.8537	.7334	.7070	.6109	.0000	.4747
RTH	.7638	.9243	.8398	.7165	.6920	.6688	.4747	.0000

Symbol: 

Height Ratio:	0.00	0.05	0.11	0.21	0.37	0.58	0.82	1.00
Letter Code:	FRA	FRB	FRC	FRD	FRE	FRF	FRG	FRH

Figure 14. Eight-Symbol Set of Filled Rectangles

Table XIV
Distance Matrix for Eight-Symbol Set of Filled Rectangles

	FRA	FRB	FRC	FRD	FRE	FRF	FRG	FRH
FRA	.0000	.2445	.4679	.6659	.8300	.9085	.9218	.9118
FRB	.2445	.0000	.2459	.5121	.8113	.9602	1.0090	1.0113
FRC	.4679	.2459	.0000	.3420	.8007	1.0378	1.1298	1.1464
FRD	.6659	.5121	.3420	.0000	.6207	.9963	1.1612	1.1994
FRE	.8300	.8113	.8007	.6207	.0000	.6156	.9625	1.0515
FRF	.9085	.9602	1.0378	.9963	.6156	.0000	.6171	.8174
FRG	.9218	1.0090	1.1298	1.1612	.9625	.6171	.0000	.4074
FRH	.9118	1.0113	1.1464	1.1994	1.0515	.8174	.4074	.0000

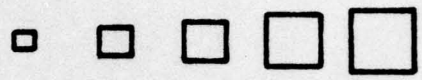
Symbol:					
Visual Angle in Minutes:	12	14	19	25	30
Side Ratio:	0.40	0.48	0.63	0.82	1.00
Letter Code:	SQA	SQB	SQC	SQD	SQE

Figure 15. Five-Symbol Set of Squares

Table XV
Distance Matrix for
Five-Symbol Set of Squares

	SQA	SQB	SQC	SQD	SQE
SQA	.0000	.3728	.7839	.8607	.8234
SQB	.3728	.0000	.6279	.8609	.8372
SQC	.7839	.6279	.0000	.7347	.8566
SQD	.8607	.8609	.7347	.0000	.6734
SQE	.8234	.8372	.8566	.6734	.0000

VITA

James A. Johnson was born in Fairmont, West Virginia, on 1 Jul 1946. He graduated from Fairmont Senior High School in 1964 and attended the United States Air Force Academy from which he received a Bachelor of Science degree in Basic Sciences in 1968. Upon graduation, he was commissioned in the Regular United States Air Force. After completing Undergraduate Pilot Training at Laughlin Air Force Base, Texas, in August 1969, he was assigned to the Aerospace Cartographic and Geodetic Service (ACGS) at Forbes Air Force Base, Kansas, as an RC-130A pilot. He piloted the AC-130A gunship in Southeast Asia from October 1971 to October 1972 and returned to the ACGS in Kansas and subsequently at Keesler Air Force Base, Mississippi, with temporary duty to Southeast Asia, the Pacific, South America, and Central America. Captain Johnson served as a flight commander in the 53rd Weather Reconnaissance Squadron of the Air Weather Service at Keesler Air Force Base until entering the School of Engineering, Air Force Institute of Technology, in June 1975.

Permanent address: Route 7, Box 338
Fairmont, West Virginia 26554

REPORT DOCUMENTATION PAGE — READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: GE/EE/76D-26
2. GOVT ACCESSION NO.:
3. RECIPIENT'S CATALOG NUMBER:
4. TITLE (and Subtitle): AN ALGORITHM TO GENERATE AN OPTIMALLY HUMAN-SEPARABLE SYMBOL SET
5. TYPE OF REPORT & PERIOD COVERED: MS Thesis
6. PERFORMING ORG. REPORT NUMBER:
7. AUTHOR(s): James A. Johnson, Captain
8. CONTRACT OR GRANT NUMBER(s):
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Air Force Institute of Technology(AFIT-EN), Wright-Patterson AFB, Ohio 45433
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS:
11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Flight Dynamics Laboratory (FGR), Air Force Systems Command, Wright-Patterson AFB, Ohio 45433
12. REPORT DATE: December 1976
13. NUMBER OF PAGES: 70
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office):
15. SECURITY CLASS. (of this report): Unclassified
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:
16. DISTRIBUTION STATEMENT (of this Report): Approved for public release; distribution unlimited
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report):
18. SUPPLEMENTARY NOTES: Approved for public release; IAW AFR 190-17

JERRAL F. GUESS, Captain, USAF
Director of Information

19. KEY WORDS (Continue on reverse side if necessary and identify by block number):
Pattern Recognition
Fourier Analysis
Fourier Transformation
Spatial Frequencies
Spatial Filters

20. ABSTRACT (Continue on reverse side if necessary and identify by block number):

The purpose of this study was to develop an algorithm which could produce from a given symbol set a new symbol set that would be optimally separable by human subjects. A symbol change algorithm was developed based on the Fourier transform model of the human visual system and the inverse relationship between the number of human confusion errors that result from the basic shapes of symbols and the Euclidian distances between these symbols in

the Fourier domain. The algorithm effectively changes the shape of certain symbols of a given symbol set by controlled manipulation of the symbol in the transform domain until a minimum, Euclidian distance threshold is achieved between the changed symbol and all other symbols in the set. A symbol set confusability number was also developed which can be used to evaluate different symbol sets with the lower number identifying the least confusable symbol set. The author believes that by applying this algorithm for a given symbol set the number of human confusion errors that occur due to the shapes of these given symbols will be effectively reduced.