DOCUMENT RESUME

ED 031 572 VT 008 787

By-Silverman, Joe
A Computer Technique for Clustering Tasks. Technical Bulletin STB 66-23.
Naval Personnel Research Activity, San Diego, Calif.
Pub Date Apr 66
Note-73p.
EDRS Price MF-$0.50 HC-$3.75
Descriptors-Bibliographies, *Cluster Grouping, Computer Oriented Programs, *Computer Programs, Grouping Procedures, Information Processing, Job Development, Job Skills, Military Training, *Models, Occupational Clusters, Program Descriptions, Questionnaires, Research Tools, Systems Approach, Tables (Data), *Task Analysis, *Techniques

The technical objective of this research was to develop a computer method for arranging a number of individual task patterns, representing job incumbents in a given occupational area, into groups or clusters. This advanced computerized technique for clustering work tasks produces homogeneous clusters of task patterns using an input of tasks performed in a sample of jobs. These clusters represent the occupational specialties that exist in a field of work. The important features of this technique are: (1) its capacity for computer analysis of task patterns of large numbers of subjects, (2) its capability for computer assistance in making research decisions at various levels of task analysis, and (3) its flexibility as a tool of pattern recognition and structuring. With only minor modification, the computer programs and concepts described in this report should be of interest to those concerned with other clustering, classifying, and taxonomic techniques. (CH)

(21)

U. S. NAVAL
PERSONNEL RESEARCH ACTIVITY

SAN DIEGO, CALIFORNIA 92152

TECHNICAL BULLETIN STB 66-23 APRIL 1966

A COMPUTER TECHNIQUE FOR CLUSTERING TASKS

Joe Silverman

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

NPRA
SAN DIEGO CALIFORNIA

AN ACTIVITY OF THE BUREAU OF NAVAL PERSONNEL

A COMPUTER TECHNIQUE FOR CLUSTERING TASKS

by

Joe Silverman

April 1966

PF016011001
Technical Bulletin STB 66-23

Submitted by

R. V. May, Jr., Director, Personnel Systems Research Department

Approved by

E. E. Dudek, Ph.D., Technical Director
G. W. Watson, Commander, USN
Officer in Charge

U. S. Naval Personnel Research Activity
San Diego, California 92152

Project Staff

Malcolm J. Carr, Project Director
Joe Silverman, Assistant Project Director
Paul A. Magnusson, Research Assistant
LTJG James W. Mosteller, USNR, Project Officer
Tandy B. Quisenberry, PNC, USN

ACKNOWLEDGEMENT

Indispensable assistance on the mathematical, statistical, and computer programming aspects of this research project has been received from the following representatives of the Statistical Department of this Activity: Dr. William J. Moonan, Director; Mrs. Margaret H. Covher, L/CPL Udo C. Pooch, and Mr. John Wolfe.

BRIEF

This report describes an advanced computerized technique for clustering work tasks which was developed in the course of research being conducted by this Activity. The objective of this research is to devise a method for determining the basic technical skills needed to man current and future weapons and support systems in order to provide a basis for the Navy enlisted personnel classification structure required in the next decade. Progress and results pertaining to this broad research objective appear in another report series issued by this Activity.

The primary purpose of this Technical Bulletin is to provide research and staff organizations involved in task analysis with a description of a new method for grouping task patterns. With an input of tasks performed in a sample of jobs, this computerized technique produces a series of relatively homogeneous clusters of task patterns. These clusters represent the occupational specialties that exist in a field of work.

The most important features of this technique are: (1) its capacity for computer analysis of task patterns of large numbers of subjects; (2) its capability for computer assistance in making research decisions at various levels of task analysis; and (3) its flexibility as a tool of pattern recognition and structuring.

In addition, this report should be of interest to those concerned with other clustering, classifying, and taxonomic techniques. The same basic problem of clustering phenomena by some criterion of similarity is encountered by physicists, mathematicians, computer designers, bio-medical engineers, information theorists, and others. With only minor modifications, the computer programs and concepts described in this report could be of value in these other fields.

CONTENTS

TABLES

FIGURES

A COMPUTER TECHNIQUE FOR CLUSTERING TASKS

I. INTRODUCTION

A. Background

The purpose of this research is to develop a method for determining the basic technical skills and their levels required for current operational weapons and support systems and for future weapons and support systems which will be introduced into operational use in the Navy during the next decade. Initial emphasis is being placed upon current skill requirements. Subsequent phases will deal with skill requirements generated by future technological developments.

The ultimate application of the method developed in this research will be the determination and description of work requirements so as to ensure their placement in the enlisted personnel classification structure in a meaningful and systematic manner. The achievement of this objective will permit the removal, replacement, and rearrangement of work requirements as this becomes necessary due to obsolescence of certain types of work, changes in others, and the addition of new work requirements associated with technological change.

As the initial step in this method development phase, a pilot study is being conducted of the engineering department in destroyers in order to determine the feasibility of the research approach and the efficacy of the associated research instruments. A report on this research was published in May 1965 (11) in which the overall concepts and research design as well as the progress to date in the pilot study are described. Readers interested in a more complete understanding of the framework of this research should consult that report.

B. Research Framework

The central concept of the research methodology in this study is that the performance of a given task or group of tasks is a function of the technical, organizational, and communicational dimensions of the work situation. Accordingly, major emphasis has been placed on the elaboration of work requirements in terms of a number of variables which are descriptive of each of these three dimensions. It is hypothesized that different occupational specialties will exhibit characteristic technical, organizational, and communicational patterns. The acronym, SAMOA, Systematic Approach to Multidimensional Occupational Analysis, has been adopted to label this approach.

The application of a multidimensional approach to occupational analysis involves, among other things, the analysis of task patterns

in terms of the technical, organizational, and communicational dimensions of the work situation. The term "task pattern" is defined as the total alignment of different tasks performed by a given individual or set of individuals in a work situation.

Before any analysis of the variables associated with these dimensions can begin, a basis for this analysis must be provided. Thus, in characterizing work requirements, it is first necessary to designate the substance and form of the work. In this research, work requirements initially take the form of a series of homogeneous and related tasks.

It is the purpose of this report to describe a technique, developed in this research, which can be used to group individual task patterns on the basis of their similarity. These groups or "clusters" of task patterns, when amplified by other variables, will help to provide the framework of work requirements necessary for a personnel classification structure.

C. Occupational Research and Task Analysis

The problem of grouping tasks and jobs for occupational classification has been approached from a variety of directions. It is the purpose of a particular method of analysis that serves as the prime criterion in choosing among alternative techniques. For instance, the manifold approaches to "job evaluation" (4,7) are all ultimately concerned with the assessment of jobs to determine their relative worth in establishing a balanced wage structure. For demographic purposes, the Bureau of the Census has approached classification in terms of broad occupational categories designed for general use (8). The Dictionary of Occupational Titles (14) provides another approach to occupational classification, employing categories which differentiate on the basis of skill level, subject matter/industry, and process/activity. This structure, like the Bureau of the Census classification, is designed for nation-wide application--particularly by guidance counselors.

These methods of grouping occupations have certain characteristics that detract from their use in this research study. Specifically, a common feature concerns the job as a basic unit of analysis, not the task. In classifying work at the job level, certain assumptions are made concerning the arrangement of work. In particular, it is assumed that "jobs" exist in the conventional sense, and not a series of task patterns that adhere to different positions depending on the specific work situation. Moreover, such approaches frequently assume that the task patterns associated with certain job titles are relatively constant and, therefore, the job can be used at the finest level of analysis.

Whatever their virtues, job level analyses are inappropriate in the context of naval occupational classification because of the unique work situation aboard ships. In order to effectively classify technical skills in naval occupations, it is necessary to approach the problem in terms of task analysis.

As with occupational analysis, the variety of techniques available for grouping tasks is considerable. Nevertheless, the purposes of this research impose a number of constraints or requirements on the kind of technique that can be used. First, tasks can be grouped by various criteria independent of their technical pattern of performance. For example, in a previous report on this research project (10), tasks were classified in terms of their technical complexity. Tasks have also been classified by their behavioral content (6), as stimulus-response events (2), in terms of learning demands (5), as man-machine elements (12), and in combinations of the above (3). In this research, tasks are associated by their pattern of technical performance and these patterns are grouped by their similarity of task content. Thus, the requirement for grouping technical task patterns as performed on board naval vessels results in constraints on the analytical techniques that can be employed.

A second constraint is concerned with the requirement to analyze large numbers of task patterns simultaneously. Conventional "job analysis" employs intensive, direct methods of obtaining occupational information. Because of the expense involved in personal contact with the job over extended periods of time, and because of the limitations in occupational coverage possible, survey methods of obtaining task pattern information are preferable for large-scale task analysis (9).

There are other considerations in this research that encouraged the development of a new approach to task pattern analysis. The feasibility of large-scale occupational analysis on a Navy-wide basis is dependent, in large part, upon the analytical speed and operational simplicity of the techniques to be employed. Thus, considerations of practicability encouraged the use of computer techniques. Also, it was advisable to minimize the amount of analytical bias introduced by contemporary occupational groupings in the Navy or other existing classification structures. It was similarly desirable to minimize the number of judgmental and inferential decisions that would have to be made in grouping combinations of tasks.

Because of the constraints imposed by the purpose of this research, it appeared feasible and desirable to develop a method which would be quantitative in approach and computerized in process. This would maximize the criteria of analytical speed and occupational scope, and provide for computer-assisted research procedures as well.

II. COMPUTER CLUSTERING TECHNIQUE

The decision to employ computerized methods of determining the "natural" task groupings in an occupational area led to a search of the research literature for possible techniques. Unfortunately, most of the existing methods are not easily adaptable to a wide range of research problems. Again, the purpose of the research dictates the limitations of the methods employed. Also, computer "soft-ware" technology has not advanced to the point that complex programs, designed for particular research objectives and written in a particular language for a specific computer, can be adapted with facility to other research purposes and other computers.

For these reasons, a new computer clustering technique was developed in response to the specific research problem involved in this study. Although many of its features are unique, there are some points which coincide with existing methods of analysis. A selected bibliography of some of these approaches to "clustering," "pattern recognition," "profile analysis," "factor analysis," and other grouping procedures, are contained in the last section of this report.

A. Technical Objective

The technical objective of the initial phase of this research was to develop a computer method for arranging a number of individual task patterns, representing job incumbents in a given occupational area, into groups or "clusters." A "cluster" is defined as a group of respondents characterized by relatively small differences in the kinds of tasks performed. In pursuing this approach, an iterative computer clustering technique was devised to group similar task patterns into homogeneous occupational segments or clusters. This technique encompasses a series of computer programs that facilitate the process of grouping task patterns and provide a variety of outputs designed to carefully regulate and control the entire procedure at any step in the process. The data collection procedures used to obtain input data, and the data processing procedures employed to obtain clusters of task patterns, are set forth in the following sections.

B. Data Collection

A number of data collection instruments were devised to obtain information on the variables associated with the three dimensions of work requirements being studied in this research. These included supervisors' questionnaires, work contact questionnaires, task lists, and others, but only the task lists are of concern for purposes of the present report.

In developing the Task List Questionnaires, a comprehensive list of tasks performed by engineering department personnel was first developed. In its final form, this list consisted of over 500 separate items. This

list was then divided into three broad work areas in engineering that appeared to be fairly discrete in terms of the work performed and the equipments involved. These are: (1) the Propulsion/Auxiliary area, encompassing work generally performed by personnel in the occupational fields of Boilerman (BT), Boilermaker (BR), Machinist's Mate (MM), and Engineman (EN); (2) the Hull/Repair area, including the work of the Damage Controlman (DC), Shipfitter (SF), and Machinery Repairman (MR); and (3) the Electrical area, covering the tasks performed by Electrician's Mates (EM) and Interior Communications Electricians (IC). Within each major area, the task list is divided into subheadings which indicate the main categories of equipment operated and maintained in that area. The instruction page and one sample page of the Task List Questionnaire, as administered to personnel in the Hull/Repair work area, are contained in Appendix A.

These task lists were administered to about 400 engineering department personnel in a sample of six destroyers in the San Diego-Long Beach area. This represents 76% of all personnel in these departments. Each man completed only that task list which pertained to his area of work.

Prior to computer processing, the task patterns of respondents, as checked on the source document (Task List Questionnaire), were key punched on cards. These cards, indicating the tasks performed by each individual, comprise the computer input.

C. Initial Computer Processing Procedures

The problem faced at this point was how to group the task patterns of respondents so that clusters of similar tasks and task patterns would emerge in a form suitable for use in determining the technical work requirements of a given occupational area.

This was accomplished in a number of steps. First, an index was developed to indicate the similarity of each individual's task pattern with that of every other individual in the sample. Second, various respondents were selected as "pivot men" on the basis of their task pattern variance, and other individuals were clustered around the pivot men by task pattern similarity. Third, the resulting clusters were analyzed by use of other computer routines in order to develop "optimum specialty clusters." Fourth, analytical procedures and computer programs were revised on the basis of the preceding analysis to refine both the technique and the data. This technique represents an interplay of mathematics, computer analysis, and human judgment. The steps employed in this procedure are described in more detail in the following sections.

Similarity Index

Prior to the actual clustering process, each individual's pattern of tasks was compared to the task pattern of every other individual who completed the same task area questionnaire. An index of similarity* was then computed for each pair of individuals based on the relative similarity of the tasks they performed.

This index is provided by:

$$S(i,j) = \frac{n\{T(i,j)\}}{n\{T(i)\} + n\{T(j)\} - n\{T(i,j)\}}$$

where

$n\{T(i,j)\}$ is the number of tasks performed by both man i and man j

$n\{T(i)\}$ is the number of tasks performed by man i

$n\{T(j)\}$ is the number of tasks performed by man j

The denominator in this expression represents the total number of different tasks performed by i and j combined.

This formula generates a continuum ranging from "0," indicating total independence (i.e., no tasks in common between man i and man j), to "1," indicating complete identity (i.e., all tasks performed by i are identical to those performed by j).

---

*This index is referred to conceptually in another form as a "Coefficient of Compositional Similarity" (CCS), in which

$$CCS = \frac{Id}{Id + Un_1 + Un_2}$$

where Id = number of tasks identical between Man 1 and Man 2

Un_1 = number of tasks unique to Man 1

Un_2 = number of tasks unique to Man 2

The CCS is an inversion of a formula originally termed the "Coefficient of Compositional Uniqueness." It was used to determine overlapping patterns of acquaintances among neighbors in a study by Carr (1), which partially replicated previous research performed by Sweetser (13).

For example, consider the following comparison of task patterns in which T(i) contains 10 elements or tasks and T(j) contains 15 elements:

T(i) = {A03,A14,A15,A19,B05,B17,C21,D04,E09,E10}

T(j) = {A03,A12,A14,A17,A19,B01,B02,B03,B17,C15,E10,E17,T01,T02,T03}

T(i,j) = {A03,A14,A19,B17,E10}

Note that man i has performed 10 tasks (indicated by the alpha-numeric codes), 5 of which are common with man j--who lists 15 tasks performed. Applying the formula,* we have:

$$S(i,j) = \frac{5}{10+15-5} = \frac{5}{20} = .25 \text{ (or 16/64ths)}$$

It was desirable to convert the quotient into 64ths because of computer processing requirements, although this is of no consequence in any subsequent stage.

This formula was applied to every possible pair of respondents in each task area and a matrix of mutual similarities was then generated by the computer.† The size of the matrix is determined by the number of personnel associated with each of the three task lists. Thus, the Propulsion/Auxiliary Task List Questionnaire, which was administered to 278 personnel, generated a semi-matrix with m(m-1)/2 or 38,503 distinct similarities, where m equals the number of personnel. The Hull/Repair list produced a semi-matrix of 741 (i.e., 39(38)/2) indices and the Electrical list resulted in 2,775 (i.e., 75(74)/2).

---

*In set notation: $S(i,j) = \frac{n\{T(i) \cap T(j)\}}{n\{T(i) \cup T(j)\}}$

where T(i) is the set of tasks performed by man i; similarly for T(j)

$n\{T(i) \cap T(j)\}$ represents the number of tasks in the "intersection" of the task lists (patterns)

$n\{T(i) \cup T(j)\}$ represents the number of tasks in the "union" of the task lists--that is, the set of tasks that belong to either or both lists

†This Index also provides the basic data for other indices used in development of the clustering technique; for instance, the Cluster Verification Score (CVS), Vector Verification Score (VVS), and Cluster Distance Score (CDS).

The similarity matrix comes in the form of a listing in which each individual is listed in serial order by identification code and all other personnel are compared with that individual by an index of similarity. For reference purposes, these data were converted to a computer-produced semi-matrix. Aside from the similarity listing and semi-matrix, the similarity indices are recorded in another form--that of a frequency distribution. For each of the three task lists, a distribution of indices was printed out in an 8 x 8 table. Examples of the initial listing, the semi-matrix, and the frequency distribution are contained in Appendix B.

Pivot Selection

In order to group the tasks performed by personnel in this sample, a starting point was necessary. In the initial computer clustering technique, this point is provided by a "pivot man"--or simply, "pivot." The pivot is the reference point for the entry of other personnel into clusters. The selection of pivots is controlled by the variance of each individual's similarity indices, where the variance is computed by:

$$s^2 = \frac{n\sum X^2-(\sum X)^2}{n(n-1)} = \frac{\sum(X-\overline{X})^2}{n-1}$$

where

X = similarity index of man i with man j, or S(i,j)

n = number of similarity indices of man i with all j

$\overline{X}$ = mean of similarity indices of man i

One of the outputs of this phase of data processing is a variance listing for each task list, as shown in Appendix B.

After the calculation of each variance, the individual with the highest variance is selected as the first pivot and becomes the reference point or core of the first cluster of task patterns. The rationale for this procedure is as follows. One of the requirements of clustering tasks is that the clusters be sizable, but also separate and distinct. A large variance indicates the presence of highly similar and highly dissimilar task patterns in a given individual's range of similarities--the maximum variance occurring where a man has one-half of his similarities = 0, and one-half = 1.

High variance is employed as the criterion for pivot selection for two reasons: first, a pivot candidate's high variance indicates that his task pattern is very similar to those of some individuals, which assures that a relatively homogeneous cluster can be formed. Second, high variance also means that the pivot candidate's pattern of tasks differs greatly from those of other personnel, thus enabling the initial cluster to be distinct from at least a portion of the body of remaining tasks. As a result, succeeding clusters can be formed around pivots that are distinct from previous clusters.

A simplified example of the relationship between an individual's range of similarities and his variance is shown in Table 1.

TABLE 1

Calculation of Variance for Two Task Pattern Samples

Man	Similarity Index X	$X-\overline{X}$	$(X-\overline{X})^2$	s^2
i	03	-17	289	
	20	0	0	
	37	17	289	
[Mean S(i,j) = 60/3 = 20]	60		578/2	= 289
j	15	- 5	25	
	20	0	0	
	25	5	25	
[Mean S(i,j) = 60/3 = 20]	60		50/2	= 25

This example shows the case of two personnel, each with a mean similarity index of 20 and a list of similarity indices with three other personnel. For man i, the similarities are both high and low (37 and 03, respectively), while for man j the similarities are grouped around the average (i.e., 15, 20, and 25). Using the deviation form,

$$s^2 = \frac{\sum(X-\overline{X})^2}{n-1} ,$$

the variance for man i is 289 while for man j, only 25. The two cases in this example are exaggerated to show the effect of variance in the selection of pivots, but the computer process is approximately the same. In terms of this computer program, man i is the better choice for pivot since highly similar task patterns (as represented by the S(i,j) of 37) can be clustered with him and still make provision for clustering other task patterns that are distinct (as represented by the S(i,j) of 03).

Cluster Grouping

After the variance is computed for each individual, and the first pivot is selected (representing the greatest variance), the initial cluster is produced by selecting those individuals with a similarity to the pivot man above a certain threshold. A "similarity threshold" (ST) was set for each computer run in order to control the process of clustering task patterns. This threshold represents the minimum similarity acceptable for inclusion in a cluster and is regulated by a "control percentage"(CP). By setting the ST at various values, the size and homogeneity of clusters can be regulated.

As noted previously, a frequency distribution of similarity indices is derived from the similarity matrix and printed out in an 8 x 8 table, with each cell representing 1/64th of the distribution. This listing was converted to a more conventional form for determining the similarity threshold to be used for each computer run. Table 2 shows the distribution of similarity indices for each of the three task areas.

Using the similarity distribution for the Hull/Repair area (m=39) in Table 2 as an example, the procedure for determining the similarity threshold can be delineated. If, for instance, the control percentage was set at 10%, a frequency count of the 741 similarities would begin at the bottom of the table and continue until 10% or 74 similarities had been counted. Note that this count ends in the frequency class of 35/64ths. The ST is thus set at 35, and the computer then generates a cluster of personnel whose similarity to the pivot is greater than the ST (i.e., ≥ 36). The resultant cluster listing contains the frequency distribution, the control percentage and ST, the identification code of the pivot, and the identification codes of all cluster members with their similarity indices to the pivot above the threshold. A partial sample cluster listing for the Hull/Repair area is shown on page 13.

Once the first pivot is selected and the members of the first cluster are chosen from those personnel with similarities to the pivot > ST, the computer initiates the selection of the second cluster. This is accomplished by setting the variances of all members of the first cluster to zero so that they will be ineligible to become pivots in succeeding clusters. The second pivot is then selected as the highest remaining variance, and a second cluster of similarities > ST is generated and printed out. As before, the variances of all personnel in the second cluster are set to zero and the third pivot is obtained by again selecting the pivot candidate with the highest variance. The procedure is reiterated and clusters are produced until a pivot candidate cannot cluster at least one other individual with a similarity to the pivot higher than the threshold.

TABLE 2

Frequency Distribution of Task Pattern Similarities in Three Task Areas

Similarity Index (64ths)	Propulsion/Auxiliary f	Hull/Repair f	Electrical f
0	2439	4	70
1	1928	5	51
2	2090	10	93
3	2084	12	85
4	2275	11	132
5	2116	15	152
6	2022	22	159
7	1976	16	137
8	2167	19	146
9	1856	20	118
10	1795	21	104
11	1700	14	89
12	1559	19	73
13	1404	21	74
14	1283	24	64
15	1028	22	53
16	1174	33	81
17	944	18	66
18	867	14	67
19	763	22	85
20	700	19	66
21	619	12	43
22	587	22	59
23	446	14	67
24	483	20	64
25	386	19	63
26	354	28	55
27	295	16	62
28	264	27	53
29	198	19	52
30	151	21	53
31	98	19	31
32	139	29	49
33	72	24	25
34	67	20	33
35	40	21	28
36	37	11	22
37	33	15	9
38	17	7	9
39	13	11	7
40	6	7	3
41	8	4	7
42	6	1	3
43	2	3	3
44	5	3	2
45	2	2	1
46	0	2	1
47	1	1	1
48	0	0	1
49	0	0	1
50	0	0	0
51	0	1	0
52	1	1	1
53	0	0	0
54	0	0	1
55	1	0	0
56	1	0	0
57	0	0	1
58	1	0	0
Total	38503	741	2775

Partial Cluster Listing
(Hull/Repair Task Area)

SIMILARITY THRESHOLD = 35 PERCENT = 10
SIMILARITY DISTRIBUTION

4	5	10	12	11	15	22	16
19	20	21	14	19	21	24	22
33	18	14	22	19	12	22	14
20	19	28	16	27	19	21	19
29	24	20	21	11	15	7	11
7	4	1	3	3	2	2	1
			1	1			

CLUSTER 1

62402	52410	62406	62419	92417	92418
	36	39	52	36	38

CLUSTER 2

92411	52410	52419	52422	52423	62431	82409	82422	92417	2404	2408
	39	39	37	41	36	39	45	36	39	40

CLUSTER 3

62403	52413	72407	82423	2409
	37	46	51	36

Computer Program Products

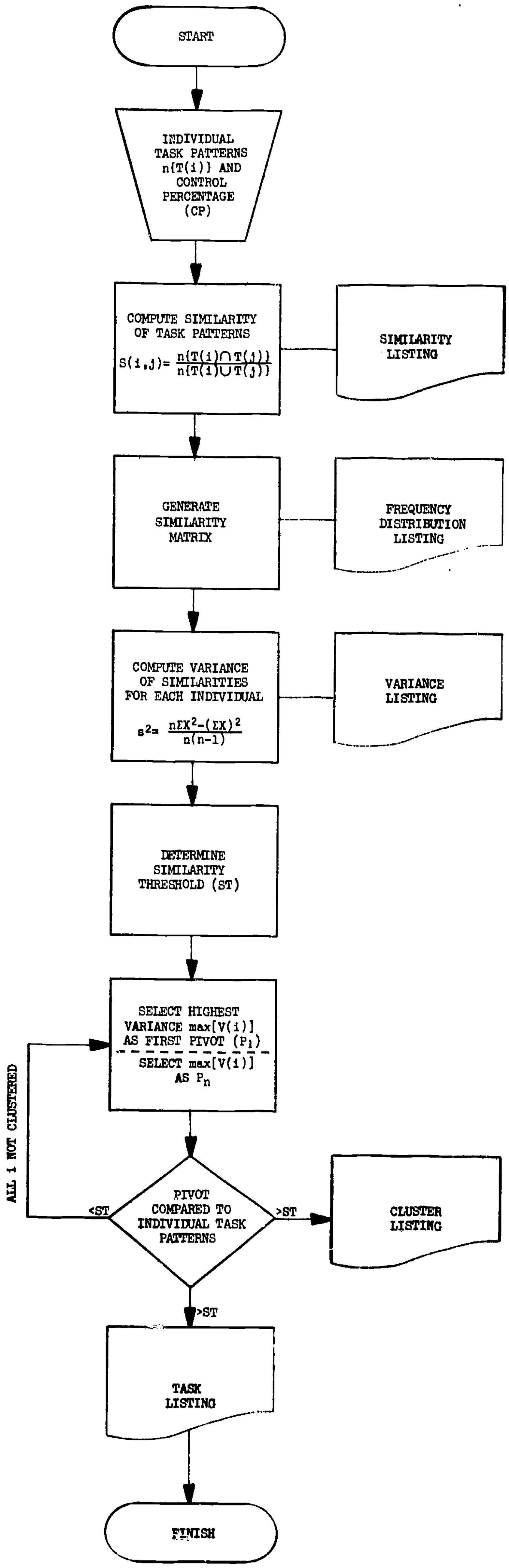
The iterative clustering program provides a number of separate but related products: (1) a similarity listing which contains an index of task pattern similarity between each man and every other man; (2) a variance listing which shows the variance of each individual's similarities; (3) a frequency distribution of similarities for each of three task lists; (4) a series of cluster listings, each showing a pivot man (the highest variance in the cluster) and all personnel with a similarity index high enough to qualify for that cluster; and (5) a task listing for each cluster, giving every task performed by personnel in that cluster and the number performing the task. The processing steps necessary to produce this output are shown in the form of a flowchart in Figure 1. There are several procedures that must be followed after production of the initial cluster runs.

The process of cluster grouping is an experimental one; that is, a series of computer runs must be made at different similarity thresholds in order to determine which ST satisfies the criteria used to evaluate the clusters. In this research, 30 different computer runs were made in the three task areas. Since each cluster run usually differs in the number of clusters, the pattern of task association, the homogeneity of the clusters, and the identity and variance of every pivot man but the first, it is necessary to examine a series of experimental clusters in order to obtain an "optimum" cluster run. The latter results in what are termed "specialty clusters."

Specialty clusters are characterized by (1) relatively low number of unclustered personnel; (2) high number of individual clusters; (3) avoidance of excessively large "initial" clusters or very small "trailing" clusters; (4) low incidence of overlapping cluster membership; (5) high variance of pivot men, especially in the last third of a cluster run; (6) low variance of low similarity cluster members, in order not to lose qualified pivots; and (7) high homogeneity of individual clusters. The analysis of computer program products is greatly facilitated by using these criteria for recognizing "optimality" in different cluster runs. However, in order to help evaluate the mass of output data produced by the computer, another program (termed "cluster identification") was required to assist in the comparison of cluster runs based on different similarity thresholds.

The examination of clusters produced by the initial program consisted of a systematic evaluation of the different sets of clusters produced by the different thresholds. Although this analysis preceded later refinements in the computer programs, it is not essential to an understanding of the clustering techniques that were ultimately adopted. As a result, the details of the "cluster identification" output and its attendant analysis are contained in Appendix C.

FIGURE 1. Computer Processing Procedures in the Initial Clustering Program



D. Program Refinement

Proceeding from a thorough analysis of the initial clustering program, a refined method of selecting pivot men and their respective clusters was developed. The revised procedure consists of two separate but related parts: the Pivot Optimization Program, which selects pivot men; and the Cluster Selection Program, which constructs the clusters around the pivots.

These techniques were developed in response to an output problem created by the initial clustering program. In the normal operation of this program, "optimum" potential pivot men could be prevented from becoming pivots by their presence or membership in preceding clusters. Figure 2 illustrates the problem manifested in the initial cluster program. A cluster with an ST of 24 is shown, in which the pivot has a variance of 100. All personnel with a similarity to the pivot over 24 are clustered, and their similarities with the pivot--as well as their variances--are also shown. Man X is included in the cluster because his similarity to the pivot is >24 (i.e., 25). Nevertheless, his variance is quite high (98)--and he could better serve as the next cluster's pivot than as a marginal member of his present cluster. Since the initial cluster program does not select pivots from among those previously clustered, man X cannot act as a pivot man. For this reason, a procedure was developed in which the selection of pivots is completed before clustering is initiated.

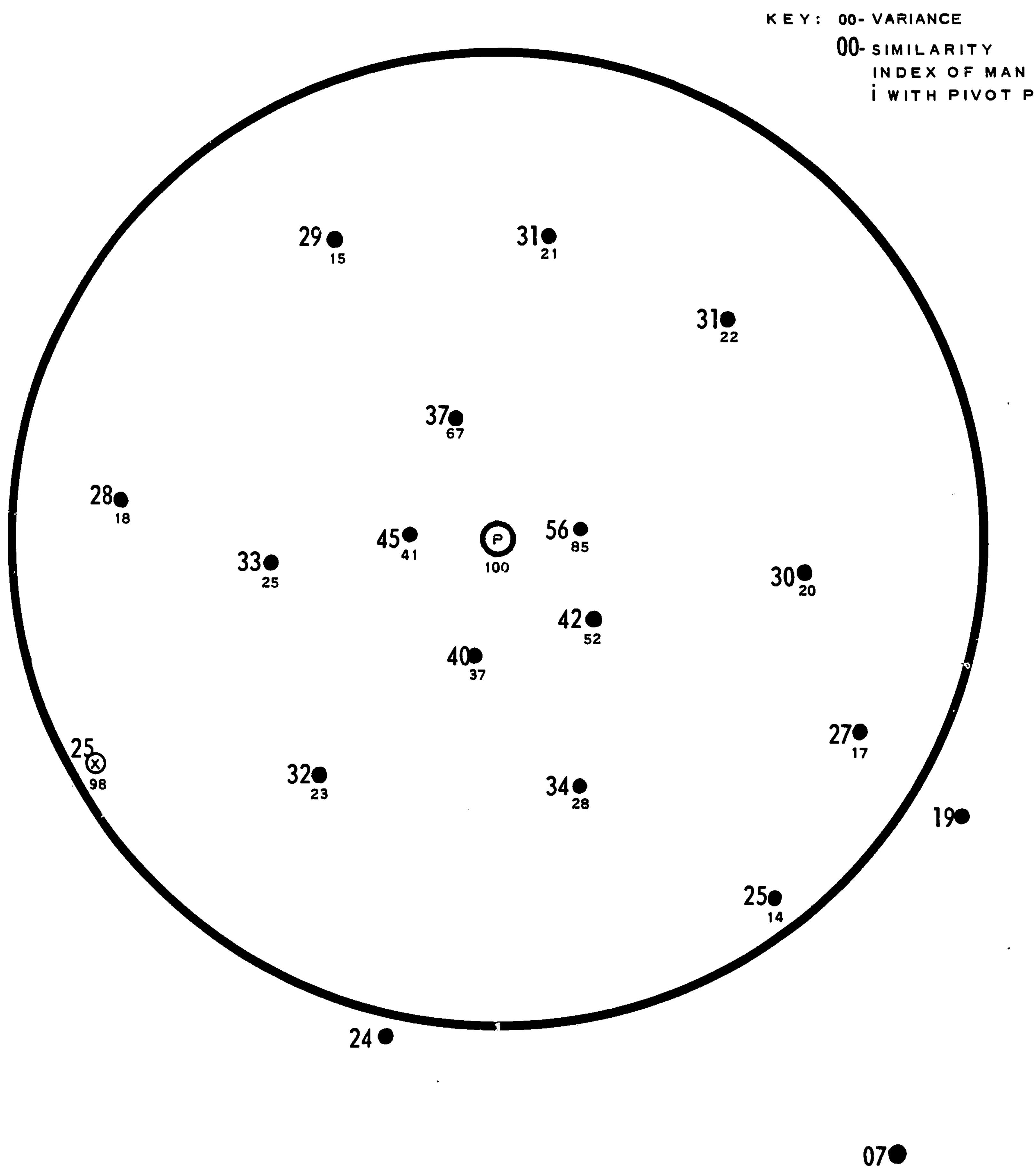
Pivot Optimization

In the refined program, the selection of pivots is optimized in two ways: first, they should have a high variance, for reasons noted previously (supra, pp. 9,10); second, they should have a relatively low similarity with each previously selected pivot man. The latter criterion enables each pivot to have a separate work area for his cluster. When two different pivots have a high similarity between them, the clusters that are formed around them are likely to be more similar than distinct. By selecting pivots who have a low similarity to preceding pivots, it is possible to avoid much of the overlapping of functional content between clusters.

The pivot selection process occurs separately for each of the three task area subsamples, and every respondent administered a task list is evaluated in terms of the two optimization criteria noted above. The function of the program is to order the men in terms of their desirability as pivots by evaluating both their individual variances as well as their similarity to previously selected pivots. The specifics of this procedure are described below.

FIGURE 2

Sample Cluster Configuration



NOTE: ST=24, THEREFORE CLUSTERING OCCURS AT S(i,P)=25

Let p, q, and r be indices of three pivot men. The individual with the highest variance, V(i), is selected as the initial pivot man p (or P_1). For all $i \neq p$, compute:

$$W(i;p) = \frac{S(i,p)}{V(i)}$$

where $S(i,p)$ = similarity of man i to pivot p

$V(i)$ = variance of man i

Select min $W(i;p)$, and designate that i = q (or P_2).

For all $i \neq p,q$, compute:

$$W(i;p,q) = \max_{p,q} \left[\frac{S(i,p)}{V(i)}, \frac{S(i,q)}{V(i)} \right]$$

Select min $W(i;p,q)$, and designate that i = r (or P_3).

A generalized procedure for selecting all pivots (other than P_1) employs the following notation. For all $i \neq \{\Pi\}$, compute:

$$W(i;\pi) = \max_{(\Pi)} \left[\frac{S(i,\pi)}{V(i)} \right]$$

where (Π) = set of all pivot men

$\pi \in \Pi$ = pivot (π) is an element of set $\{\Pi\}$

Select min $W(i;\pi)$, and designate that $i = \pi$.

Appendix D contains a computer listing representing partial output of the pivot optimization program. For P_1, it lists his identification code (ID) and variance. For all succeeding pivots, it lists his ID and variance, his similarity with π (the highest similarity with any π in the set $\{\Pi\}$), and his $W(i;\pi)$. To express the optimality of pivots, the size of $W(i;\pi)$ indices increases with each succeeding pivot man on the list. Preliminary decisions regarding the number of pivots to use for clustering (and, therefore, the number of clusters in a task area) were based on an analysis of these listings. For example, the partial Propulsion/Auxiliary listing on page 19 shows a break in the pivot optimization values after the seventh man. Thus the first seven pivots were tentatively selected to form experimental clusters. The number of clusters employed could be changed by simply altering the number of pivots chosen from the list. A more refined technique for selecting the number of clusters, without reference to pivots, was subsequently developed and is elaborated in a later section. Procedures for clustering around the pivots selected are contained in the following discussion.

Partial Pivot Optimization Listing
(Propulsion/Auxiliary Task Area)

PROGRAM-CLUSTER MODIFICATION
ALPHA .600

1ST PIVOT MAN	1072242	VARIANCE	97						
MANS ID	1062033	VARIANCE	90	SIMILARITY	11	COMPUTED VALUE	.1222	RELATED PIVOT MAN	1072242
MANS ID	1052215	VARIANCE	62	SIMILARITY	12	COMPUTED VALUE	.1935	RELATED PIVOT MAN	1072242
MANS ID	1082224	VARIANCE	69	SIMILARITY	18	COMPUTED VALUE	.2609	RELATED PIVOT MAN	1072242
MANS ID	1092021	VARIANCE	82	SIMILARITY	25	COMPUTED VALUE	.3049	RELATED PIVOT MAN	1062033
MANS ID	1102218	VARIANCE	82	SIMILARITY	30	COMPUTED VALUE	.3659	RELATED PIVOT MAN	1082224
MANS ID	1082028	VARIANCE	75	SIMILARITY	29	COMPUTED VALUE	.3867 ←	RELATED PIVOT MAN	1092021
MANS ID	1072249	VARIANCE	80	SIMILARITY	32	COMPUTED VALUE	.4000	RELATED PIVOT MAN	1062033
MANS ID	1052234	VARIANCE	80	SIMILARITY	32	COMPUTED VALUE	.4000	RELATED PIVOT MAN	1102218
MANS ID	1062015	VARIANCE	68	SIMILARITY	28	COMPUTED VALUE	.4118	RELATED PIVOT MAN	1062033
MANS ID	1072218	VARIANCE	75	SIMILARITY	31	COMPUTED VALUE	.4133	RELATED PIVOT MAN	1072242
MANS ID	1092031	VARIANCE	78	SIMILARITY	33	COMPUTED VALUE	.4231	RELATED PIVOT MAN	1082028
MANS ID	1052014	VARIANCE	87	SIMILARITY	37	COMPUTED VALUE	.4253	RELATED PIVOT MAN	1072242
MANS ID	1092010	VARIANCE	82	SIMILARITY	36	COMPUTED VALUE	.4390	RELATED PIVOT MAN	1072242
MANS ID	1092023	VARIANCE	68	SIMILARITY	30	COMPUTED VALUE	.4412	RELATED PIVOT MAN	1092031
MANS ID	1092214	VARIANCE	77	SIMILARITY	34	COMPUTED VALUE	.4416	RELATED PIVOT MAN	1082224
MANS ID	1052017	VARIANCE	76	SIMILARITY	34	COMPUTED VALUE	.4474	RELATED PIVOT MAN	1072242
MANS ID	1092025	VARIANCE	87	SIMILARITY	39	COMPUTED VALUE	.4483	RELATED PIVOT MAN	1072242
MANS ID	1072220	VARIANCE	80	SIMILARITY	36	COMPUTED VALUE	.4500	RELATED PIVOT MAN	1072208
MANS ID	1062007	VARIANCE	73	SIMILARITY	33	COMPUTED VALUE	.4521	RELATED PIVOT MAN	1072242
MANS ID	1062026	VARIANCE	97	SIMILARITY	44	COMPUTED VALUE	.4536	RELATED PIVOT MAN	1062033
MANS ID	1092024	VARIANCE	63	SIMILARITY	29	COMPUTED VALUE	.4603	RELATED PIVOT MAN	1062033
MANS ID	1082008	VARIANCE	89	SIMILARITY	41	COMPUTED VALUE	.4607	RELATED PIVOT MAN	1092025
MANS ID	1092216	VARIANCE	69	SIMILARITY	32	COMPUTED VALUE	.4638	RELATED PIVOT MAN	1102218
MANS ID	1052211	VARIANCE	83	SIMILARITY	39	COMPUTED VALUE	.4699	RELATED PIVOT MAN	1082224
MANS ID	1072255	VARIANCE	74	SIMILARITY	35	COMPUTED VALUE	.4730	RELATED PIVOT MAN	1052234
MANS ID	1062010	VARIANCE	74	SIMILARITY	35	COMPUTED VALUE	.4730	RELATED PIVOT MAN	1092031
MANS ID	1072209	VARIANCE	80	SIMILARITY	38	COMPUTED VALUE	.4750	RELATED PIVOT MAN	1072242
MANS ID	1062018	VARIANCE	63	SIMILARITY	30	COMPUTED VALUE	.4762	RELATED PIVOT MAN	1072242
MANS ID	1092029	VARIANCE	94	SIMILARITY	45	COMPUTED VALUE	.4787	RELATED PIVOT MAN	1072242
MANS ID	1082223	VARIANCE	75	SIMILARITY	36	COMPUTED VALUE	.4800	RELATED PIVOT MAN	1102218
MANS ID	1072205	VARIANCE	85	SIMILARITY	41	COMPUTED VALUE	.4824	RELATED PIVOT MAN	1072220
MANS ID	1072245	VARIANCE	64	SIMILARITY	31	COMPUTED VALUE	.4844	RELATED PIVOT MAN	1052234
MANS ID	1062009	VARIANCE	76	SIMILARITY	37	COMPUTED VALUE	.4868	RELATED PIVOT MAN	1092010
MANS ID	1062231	VARIANCE	63	SIMILARITY	31	COMPUTED VALUE	.4921	RELATED PIVOT MAN	1092214
MANS ID	1062020	VARIANCE	73	SIMILARITY	36	COMPUTED VALUE	.4932	RELATED PIVOT MAN	1092025
MANS ID	1062030	VARIANCE	75	SIMILARITY	37	COMPUTED VALUE	.4933	RELATED PIVOT MAN	1062007
MANS ID	1052016	VARIANCE	75	SIMILARITY	37	COMPUTED VALUE	.4933	RELATED PIVOT MAN	1082008
MANS ID	1092013	VARIANCE	68	SIMILARITY	34	COMPUTED VALUE	.5000	RELATED PIVOT MAN	1092021
MANS ID	1092011	VARIANCE	82	SIMILARITY	41	COMPUTED VALUE	.5000	RELATED PIVOT MAN	1062033
MANS ID	1072214	VARIANCE	70	SIMILARITY	35	COMPUTED VALUE	.5000	RELATED PIVOT MAN	1072242
MANS ID	1052010	VARIANCE	70	SIMILARITY	35	COMPUTED VALUE	.5000	RELATED PIVOT MAN	1082008
MANS ID	1052210	VARIANCE	73	SIMILARITY	37	COMPUTED VALUE	.5068	RELATED PIVOT MAN	1052211
MANS ID	1092032	VARIANCE	61	SIMILARITY	31	COMPUTED VALUE	.5082	RELATED PIVOT MAN	1092010
MANS ID	1102205	VARIANCE	80	SIMILARITY	41	COMPUTED VALUE	.5125	RELATED PIVOT MAN	1092214
MANS ID	1092012	VARIANCE	80	SIMILARITY	41	COMPUTED VALUE	.5125	RELATED PIVOT MAN	1072242
MANS ID	1092219	VARIANCE	74	SIMILARITY	38	COMPUTED VALUE	.5135	RELATED PIVOT MAN	1092214
MANS ID	1052420	VARIANCE	68	SIMILARITY	35	COMPUTED VALUE	.5147	RELATED PIVOT MAN	1082008
MANS ID	1052008	VARIANCE	81	SIMILARITY	42	COMPUTED VALUE	.5185	RELATED PIVOT MAN	1092029
MANS ID	1052235	VARIANCE	71	SIMILARITY	37	COMPUTED VALUE	.5211	RELATED PIVOT MAN	1052234
MANS ID	1092014	VARIANCE	69	SIMILARITY	36	COMPUTED VALUE	.5217	RELATED PIVOT MAN	1092021
MANS ID	1092015	VARIANCE	65	SIMILARITY	34	COMPUTED VALUE	.5231	RELATED PIVOT MAN	1082028
MANS ID	1092224	VARIANCE	59	SIMILARITY	31	COMPUTED VALUE	.5254	RELATED PIVOT MAN	1052234
MANS ID	1082218	VARIANCE	59	SIMILARITY	31	COMPUTED VALUE	.5254	RELATED PIVOT MAN	1082223
MANS ID	1092006	VARIANCE	76	SIMILARITY	40	COMPUTED VALUE	.5263	RELATED PIVOT MAN	1062026

Cluster Selection

Once the "optimum" pivots have been determined, the selection of their clusters is a relatively simple process. All personnel to be clustered (i.e., those other than pivots) are considered separately. Each individual is selected for membership in that cluster with whose pivot man he has the greatest similarity. An individual thus appears in only one cluster, excepting those instances in which his highest similarity is with two or more pivot men. In the latter case, such individuals appear in all clusters with whose pivots the tie occurs and also appear in a separate listing of ties.

The output of the cluster selection program is in the form of a listing of clusters, a sample of which is shown on page 21. This listing gives the ID of the pivot man around which the cluster is formed, and then lists other cluster members by order of descending similarity to the pivot. The cluster member's ID, his index of similarity with the pivot man, and his variance constitute each line entry in the cluster listing. Appendix E contains some examples of cluster listings for each of the three engineering task areas.

Partial Cluster Selection Listing
(Propulsion/Auxiliary Task Area)

PIVOT MAN CLUSTERED ON 1092021

IDNUMBER	1072205	SIMILARITY WITH PIVOT MAN	37	VARIANCE	85
IDNUMBER	1092014	SIMILARITY WITH PIVOT MAN	36	VARIANCE	69
IDNUMBER	1072220	SIMILARITY WITH PIVOT MAN	35	VARIANCE	80
IDNUMBER	1092017	SIMILARITY WITH PIVOT MAN	34	VARIANCE	62
IDNUMBER	1092013	SIMILARITY WITH PIVOT MAN	34	VARIANCE	68
IDNUMBER	1082020	SIMILARITY WITH PIVOT MAN	34	VARIANCE	66
IDNUMBER	1082026	SIMILARITY WITH PIVOT MAN	32	VARIANCE	54
IDNUMBER	1052026	SIMILARITY WITH PIVOT MAN	32	VARIANCE	69
IDNUMBER	1052023	SIMILARITY WITH PIVOT MAN	32	VARIANCE	57
IDNUMBER	1052233	SIMILARITY WITH PIVOT MAN	31	VARIANCE	46
IDNUMBER	1092023	SIMILARITY WITH PIVOT MAN	30	VARIANCE	68
IDNUMBER	1082017	SIMILARITY WITH PIVOT MAN	30	VARIANCE	58
IDNUMBER	1082009	SIMILARITY WITH PIVOT MAN	30	VARIANCE	53
IDNUMBER	1072218	SIMILARITY WITH PIVOT MAN	30	VARIANCE	66
IDNUMBER	1052218	SIMILARITY WITH PIVOT MAN	30	VARIANCE	41
IDNUMBER	1082028	SIMILARITY WITH PIVOT MAN	29	VARIANCE	75
IDNUMBER	1102032	SIMILARITY WITH PIVOT MAN	28	VARIANCE	49
IDNUMBER	1102009	SIMILARITY WITH PIVOT MAN	28	VARIANCE	55
IDNUMBER	1082016	SIMILARITY WITH PIVOT MAN	28	VARIANCE	49
IDNUMBER	1072241	SIMILARITY WITH PIVOT MAN	28	VARIANCE	56
IDNUMBER	1072228	SIMILARITY WITH PIVOT MAN	28	VARIANCE	57
IDNUMBER	1102033	SIMILARITY WITH PIVOT MAN	27	VARIANCE	45
IDNUMBER	1102030	SIMILARITY WITH PIVOT MAN	27	VARIANCE	54
IDNUMBER	1092022	SIMILARITY WITH PIVOT MAN	27	VARIANCE	53
IDNUMBER	1082029	SIMILARITY WITH PIVOT MAN	27	VARIANCE	56
IDNUMBER	1072204	SIMILARITY WITH PIVOT MAN	26	VARIANCE	52
IDNUMBER	1052208	SIMILARITY WITH PIVOT MAN	26	VARIANCE	55
IDNUMBER	1052020	SIMILARITY WITH PIVOT MAN	26	VARIANCE	53
IDNUMBER	1052205	SIMILARITY WITH PIVOT MAN	25	VARIANCE	39
IDNUMBER	1092007	SIMILARITY WITH PIVOT MAN	24	VARIANCE	44
IDNUMBER	1072257	SIMILARITY WITH PIVOT MAN	17	VARIANCE	26
IDNUMBER	1062013	SIMILARITY WITH PIVOT MAN	11	VARIANCE	23
IDNUMBER	1092409	SIMILARITY WITH PIVOT MAN	8	VARIANCE	8

III. ANALYSIS OF CLUSTERING TECHNIQUES

There are several considerations which emerge from the preceding discussion of computer techniques for clustering tasks. First, in using methods which employ a "pivotal" task pattern as the reference point for grouping similar task patterns, the selection of those pivots is critically important. Second, the particular technique used to cluster task patterns around a "core" can vary, depending on the criteria used to evaluate clusters and the particular research objectives involved. Third, the development of "optimum specialty clusters" necessitates some procedure for regulating the size and homogeneity of clusters.

Each of these three problem areas was examined in detail in the process of developing techniques for the analysis of task patterns. The procedures employed in this analysis and the determinations resulting from it are contained in the following discussion.

A. Effects of Differential Pivot Selection

The primary criterion for the selection of pivots in this research, regardless of the specific technique employed, has been the magnitude of an individual's variance. Thus, for a given task area, those individuals who possessed a high variance of task pattern similarities were more likely to become pivots than those with lower variances. By comparing the initial pivot selection technique with that employed in the pivot optimization program, the more effective method for optimizing the selection of pivots can be determined. Table 3 shows the results of this comparison in an abbreviated list of pivots produced by the two programs.

Of the two methods of pivot selection in Table 3, note that the initial pivot selection technique is shown under four different conditions; that is, pivots were selected with similarity thresholds set at 25, 21, 19, and 16. Although the ST is used primarily to regulate entry into clusters, it also affects the number and kinds of pivots selected--see Appendix C.

Under the four conditions, the range of variance among pivots P_1 to P_{15} runs from 97-44, 97-36, 97-31, and 97-21, respectively. It can be seen from Table 3 that as the ST is lowered, so is the variance of P_n. But even with ST = 25 (a relatively high threshold), P_{15} has a variance of only 44.

In contrast to the initial method of selecting pivots, the pivot optimization technique results in a single list of pivots in which the variance is maximized for each--while still producing pivots that are mutually distinctive in their task patterns. Although the threshold is set at ST = 1 (which, in effect, sets <u>no</u> restriction on cluster membership), Table 3 shows the range of pivot variances to be 97-62.

TABLE 3

Comparison of Two Pivot Selection Techniques
as Applied to the Propulsion/Auxiliary Task Area

Order of Selection	Initial Pivot Selection Technique								Pivot Optimization Technique	
	ST=25		ST=21		ST=19		ST=16		ST=1	
	ID Code	s^2	ID Code	s^2	ID Code	s^2	ID Code	s^2	ID Code	s^2
P_1	72242	97	72242	97	72242	97	72242	97	72242	97
P_2	62026	97	62026	97	62026	97	62026	97	62033	90
P_3	52211	83	52211	83	52234	80	72245	64	52215	62
P_4	82028	75	82201	56	82016	49	52217	53	82224	69
P_5	72245	64	92022	53	52409	44	52409	44	92021	82
P_6	82201	56	82228	49	82222	43	82222	43	02218	82
P_7	02030	54	62206	45	52218	41	52006	38	82028	75
P_8	82009	53	52409	44	52006	38	92205	31	72249	80
P_9	82013	50	92007	44	62017	36	52227	31	52234	80
P_{10}	72251	49	02016	41	72237	35	92020	30	62015	68
P_{11}	02224	47	52218	41	72227	33	02026	29	72208	75
P_{12}	92033	45	92229	39	92034	33	72257	26	92031	78
P_{13}	02033	45	82203	38	02210	32	62013	23	52014	87
P_{14}	52409	44	62021	38	92205	31	82024	22	92010	82
P_{15}	92226	44	62017	36	52227	31	92225	21	92023	68

In fact, few of the pivots produced by the initial method (regardless of the ST) are even listed in the first 15 pivots selected by the optimization method. The critical feature of the latter technique is the avoidance of the problem illustrated in Figure 2 (<u>supra</u>, p. 17), whereby potential pivots are lost through inclusion in preceding clusters.

B. Effects of Differential Cluster Formation

Aside from the particular method used to select pivots, the application of the two clustering techniques results in different cluster effects. In the initial cluster program, individuals are grouped into clusters when their similarity to a pivot exceeds a stated minimum. In contrast, the cluster selection technique associated with the pivot optimization program produces clusters by grouping individuals together by their highest similarity to a given pivot.

The resulting clusters produced by these two techniques differ in one important respect. In the initial cluster program, a sizable number of personnel appear in more than one cluster because their similarity to a number of pivots exceeds the threshold. For example, in the Propulsion/Auxiliary task area these multiple memberships comprise between 54% and 72% of the sample--depending on the particular similarity threshold set for the cluster run. Multiple memberships constitute a factor which frequently has a negative effect on cluster homogeneity. This is due to the introduction of heterogeneous segments of task patterns into more than one cluster.

Conversely, the cluster selection program clusters by reference to an individual's highest similarity and, as a result, individuals generally appear in a single cluster. The only exception occurs when an individual's <u>highest</u> similarity relates to more than one pivot. In order to understand the differential effect of these two methods of cluster formation, an analysis of cluster homogeneity was undertaken.

The evaluation of a set of clusters is accomplished by reference to the criterion of homogeneity. "Optimum specialty clusters" are those which maximize task pattern homogeneity within a cluster. Since clusters are formed by the relationship of an individual's <u>similarity to a pivot</u>, there is no assurance that this relationship will automatically result in high <u>similarity among all personnel</u> in a given cluster. In order to maximize the criterion of homogeneity, a computer program called the Cluster Verification routine was developed. This program employs an input of individual task patterns in a given cluster, generates an intra-cluster similarity matrix, and produces an output which shows the mean task pattern similarity of the entire cluster (cluster verification score or CVS) and the standard deviation.

It also identifies each individual in the cluster by code, the mean similarity of each individual's similarities with all other cluster

members (vector verification score or VVS), and the standard deviation. Appendix F contains an example of the computer output for one cluster in the Electrical task area.

Using the verification scores (CVS) to measure cluster homogeneity, different cluster arrangements produced by the two computer programs can be compared and evaluated. Table 4 indicates the various CVS values for five clusters in the Electrical area. These clusters, produced by the two clustering techniques, employ identical pairs of pivots and identical thresholds. By holding the pivot factor and threshold factor constant, the effect of multiple memberships can be examined in isolation.

Table 4 shows the differences in homogeneity of clusters produced by the two clustering techniques under varying conditions. In most cases, the effect of multiple memberships has been the dilution of cluster homogeneity. For instance, with the initial cluster program run at ST=25, the five clusters show mean similarities (CVS) of 29.90, 28.82, 27.55, 27.93, and 28.67. On the other hand, the five clusters produced by the cluster selection program at the same threshold (i.e., 25) show consistently higher CVS values of 34.98, 30.79, 28.11, 28.31, and 28.73. For some clusters (e.g., C_5) the increase in homogeneity is minimal, but for others (e.g., C_1) it is fairly large. Aside from the multiple memberships in the "initial" clusters, these task groupings are identical.

It is interesting to note that with the cluster selection program set at ST=1 (where 100% of the respondents are clustered), and the initial program set at ST=25 (where only 75% are clustered), the homogeneity of one "optimization" cluster (i.e., C_4) is still greater than its counterpart, and another (i.e., C_1) is quite similar. Thus, even with no effective threshold, the cluster selection technique sometimes produces greater homogeneity than the initial technique <u>with</u> a threshold.

When higher threshold runs are compared, there appears to be little difference between the two clustering techniques. However, at those thresholds (i.e., ST=27, 30, or 33) the number of personnel that are clustered is small. At ST=33, for instance, only 44% are clustered--compared with 75% at ST=25.

C. <u>Effects of Differential Threshold Regulation</u>

Regardless of the method employed in either pivot selection or cluster formation, the extent of homogeneity in a cluster will ultimately depend on the "entry level" established for the particular cluster. The entry level is a designated value which defines the minimum level of similarity required for inclusion in a cluster.

In the initial cluster program, the entry level is stated in terms of a similarity threshold (ST) which regulates the entry of personnel

TABLE 4

Comparison of Two Cluster Formation Techniques as Applied to the Electrical Task Area

Cluster Number	ST=	Initial Cluster Technique 25	27	30	33	Cluster Selection Technique 1	25	27	30	33
C_1	Mean (CVS)	29.90	30.21	33.04	37.08	29.37	34.98	34.98	35.18	38.78
Pivot: 72427	SD	6.12	6.20	6.57	5.73	10.01	7.50	7.50	8.09	6.81
	m	25	24	17	12	14	12	12	11	9
C_2	Mean (CVS)	28.82	29.79	34.86	35.93	19.68	30.79	32.24	39.00	39.00
Pivot: 02812	SD	5.70	5.64	4.89	5.48	12.54	5.84	6.02	6.72	6.72
	m	17	15	8	6	11	8	7	4	4
C_3	Mean (CVS)	27.55	28.14	29.27	31.93	17.04	28.11	28.11	28.62	32.20
Pivot: 52405	SD	5.26	5.23	5.79	5.24	8.54	5.49	5.49	6.11	4.69
	m	21	15	10	6	19	8	8	7	5
C_4	Mean (CVS)	27.93	29.29	35.76	35.76	28.13	28.31	29.96	36.53	36.53
Pivot: 72410	SD	6.23	6.36	3.06	3.06	7.19	6.87	6.79	3.00	3.00
	m	18	13	7	7	13	13	10	6	6
C_5	Mean (CVS)	28.67	29.43	30.00	33.13	22.86	28.73	29.30	29.76	32.00
Pivot: 52406	SD	5.70	5.37	5.26	5.39	8.51	5.13	5.14	5.20	4.93
	m	29	25	23	15	21	16	15	14	10
Personnel Clustered	No.	56 (Minus 44 Mults)	51 (Minus 41 Mults)	41 (Minus 24 Mults)	33 (Minus 13 Mults)	75 (Minus 3 Ties)	56 (Minus 1 Tie)	51 (Minus 1 Tie)	41 (Minus 1 Tie)	33 (Minus 1 Tie)
	%	74.7	68.0	54.7	44.0	100.0	74.7	68.0	54.7	44.0

into a cluster by their similarity to the pivot. By increasing the ST, and thus making entrance to a cluster more restrictive, the homogeneity of a cluster is also raised. However, because the more restrictive cluster entrance requirement necessarily excludes more personnel, every increase in the ST results in an increased number of unclustered personnel. Thus, a trade-off in improved cluster homogeneity requires the exclusion of a sizable portion of the sample of task patterns.

Aside from the similarity threshold as a method of cluster regulation, there is a different kind of entry level that might be used in maximizing the homogeneity function of clusters. The latter is obtained from a cluster verification listing (see Appendix F) which shows the mean similarity of the cluster as a whole (CVS), but also shows the mean similarity of each cluster member's relationship with all other members (VVS). The relative effectiveness of these two types of threshold is shown in Table 5.

Table 5 contains the components of a single cluster: listed thereon are the identification codes, the similarity of each individual with the pivot of cluster [S(i;p)], and the mean similarity (VVS) of each cluster member's task pattern relationships. Employing the ST method of regulating the size and homogeneity of clusters, the thresholds were set at ST = 34, ST = 28, and ST = 14--yielding clusters with m = 10, m = 15, and m = 20, respectively. The cluster verification scores (CVS) for these potential clusters, as well as the total, are listed at the bottom of the table.

With the same size clusters, thresholds were set by the mean similarity of each individual's vector of similarities (VVS). The CVS scores for clusters set at those thresholds (i.e., VVS = 25, 23, and 13) are also listed at the bottom of the table.

For the complete cluster (m = 21), the CVS is necessarily identical--because the cluster membership is identical. Similarly, the same CVS is obtained for both cluster regulation methods at m = 15; again, because of identical memberships. However, for the most restrictive threshold (m = 10), the homogeneity of the cluster is greater when using the mean vector similarity (VVS) as a threshold than by using the similarity to the pivot (ST). Similarly, the same result emerges when the two clusters of m = 20 are compared.

Based on this analysis, the results indicate that the size of a cluster and its homogeneity can, in some cases, be optimized by employing mean similarity, rather than similarity to the pivot, as the method of regulating clusters. However, the difference in results produced by the two techniques is not so great as those shown between the two pivot selection techniques and the two methods of cluster formation.

TABLE 5

Comparison of Two Threshold Regulation Techniques
As Applied to a Cluster in the Electrical Task Area

Identification Code (Arrayed by Similarity to the Pivot)		Similarity With Pivot [S(i,p)]	Identification Code (Arrayed by Mean Similarity to All Cluster Members)		Mean Similarity of Each Cluster Member [VVS]
52406 (pivot)			52406 (pivot)		30
02810		44	72420		30
77420		41	02810		29
62428		38	62428		27
72418		36	72418		26
52424		36	52424		26
82415		35	62422		26
62430		35	62430		25
62422		35	02804		25
82418	m=10	34	82427	m=10	25
92416		33	62433		24
02804		31	82415		23
82427		31	82418		23
52425		31	92416		23
62433	m=15	28	52425	m=15	23
92208		26	92208		23
62413		22	62413		19
92406		20	92406		17
82416		18	82416		14
02811	m=20	14	72434	m=20	13
72434		13	02811		11

Cluster Designation	Mean Similarity of Cluster (CVS)	Cluster Designation	Mean Similarity of Cluster (CVS)
m=10	32.00	m=10	32.60
m=15	29.30	m=15	29.30
m=20	23.99	m=20	24.23
m=21	22.86	m=21	22.86

D. Summary

The preceding sections have emphasized the three stages in developing clusters; namely, (1) the selection of optimum pivots, (2) the formation of clusters around pivots, and (3) the regulation of size and homogeneity of clusters by a threshold. For each of these processes, two techniques have been compared.

In the case of pivot selection, the initial pivot program and the pivot optimization program were analyzed in terms of their respective output. The latter technique was found to be the more effective in maximizing the variance of pivots, while still maintaining the task pattern distinctions among pivots.

The two methods of forming clusters were compared in terms of the membership of clusters and their homogeneity. Of the two programs, the cluster selection technique was found to contribute more to cluster homogeneity, through the avoidance of multiple memberships, than the initial program.

In considering techniques for regulating clusters, the similarity threshold (ST) contributes somewhat less to cluster homogeneity than the threshold derived from mean vector similarities (VVS). Because the regulation of clusters through manipulation of thresholds is so important in developing homogeneous clusters of work requirements, a more detailed analysis of this area was conducted in terms of the Unified Cluster System (UCS)--elaborated in the following section.

IV. UNIFIED CLUSTER SYSTEM

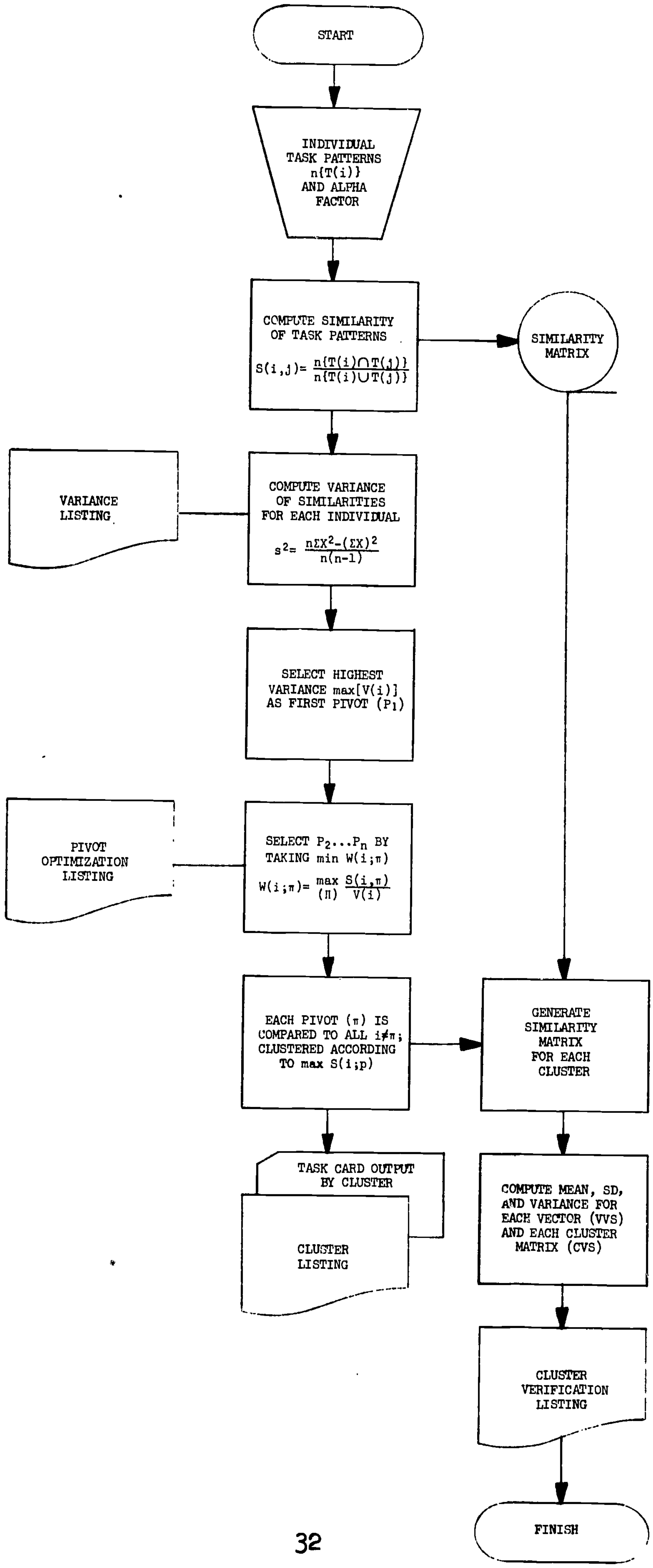
In analyzing the effects of differential pivot selection and cluster formation, it was possible to develop a "unified" computer program which could--in a single run--produce most of the desired outputs necessary to formulate decisions regarding the size and number of optimum specialty clusters in a given occupational field. To accomplish this end, a series of computer programs was integrated in a single "package" designed to provide the data necessary for a comprehensive analysis; this integrated group of programs was designated the "Unified Cluster System" or UCS.

Input for the UCS consists of a magnetic tape containing the original matrices of task pattern similarities derived from the deck of cards produced by responses on the task list questionnaires. Output is comprised of printouts that were previously the result of separate computer runs. These outputs include a variance listing, pivot optimization listing, printout of "ties," cluster listing, cluster verification listing, and an output of punch cards containing the task patterns of respondents arranged by cluster in the same form as the cluster listing (see Figure 3 for processing procedures). The unique feature of UCS is its capability of producing multiple runs with card output.

Instead of clustering all respondents' task patterns around a predetermined number of pivots judged to be appropriate for a given task area, the UCS contains an iterative procedure for fixing the number of clusters. This technique groups all personnel into two clusters, then produces a complete UCS output package. The program then recycles and groups the task patterns by their respective similarity to three pivots--again, with the appropriate output. Each time the process is iterated, it adds a pivot from the pivot optimization listing in preferential order. Thus, the program results in a series of cluster sets; the first set containing two clusters, the second set three clusters, the third set four clusters, and so on until the pivot optimization list is exhausted. With this output, an occupational area can be evaluated in terms of two or more optimum specialty clusters, without initial decisions as to the optimum number and size of clusters.

Although each set includes the pivots of the preceding set, the addition of each new pivot causes the successive iterations to form different task patterns. This is because the personnel are redistributed in terms of their highest similarity to a pivot. As a result, when the members of two clusters are presented with a third pivot for comparison of task pattern similarity, that third pivot will usually attract some marginal members of the initial clusters. Each time the program recycles, the same pivot will frequently attract a somewhat different constellation of cluster members. Thus, the first pivot selected (P_1) will provide the basis for a maximum of k different clusters (k = number of iterations or sets); and the terminal pivot (P_t) necessarily attracts a single cluster.

FIGURE 3. Computer Processing Procedures in the UCS



A. Designation of Specialty Clusters

In order to isolate the specialty clusters in an occupational area, a few limitations must be imposed on the process of cluster analysis. First, the size of the sample in a task area dictates the upper and lower limits for a cluster in that area. For example, in this research it did not appear feasible to employ clusters of less than ten respondents. The description of clusters in terms of the technical, organizational, and communicational variables would not be statistically meaningful with very small clusters because of the paucity of data. Similarly, excessively large clusters would exhaust most of the sample in a particular task area, leaving few respondents as a source of data to describe other clusters in the area. As a result, the particular constraints of size in this occupational sample were set within the flexible limits of between 10 and 50 personnel. The clusters which emerged from UCS did not indicate that these constraints posed a significant limitation on the process of cluster analysis.

A second constraint in designating specialty clusters involves threshold regulation. The UCS, unlike the initial cluster program, clusters <u>all</u> personnel in the sample according to their highest similarity to a pivot. Because of this, there are a number of respondents' task patterns that do not adhere <u>closely</u> to any pivot, but are nevertheless included in those clusters to whose pivot they are most similar. These personnel have marginal or deviate task patterns because: (1) they were new arrivals on board ship at the time of sampling (and thus performed an erratic and incomplete list of tasks); (2) they did not complete the task list questionnaire; (3) the questionnaire was improperly filled out; (4) the survey instructions were misunderstood; or (5) simply because their task patterns were relatively unique on the particular ship(s) sampled. Whatever the reason, the task patterns associated with these personnel detract from cluster homogeneity to a significant degree. It is the precise purpose of the similarity threshold (ST) to eliminate such deviant cases, providing that cluster similarity is not promoted at the expense of a sizable portion of the sample.

In light of the constraints discussed above, the initial step in designating specialty clusters involves setting thresholds on <u>all</u> clusters produced in the three task areas by the UCS program. This process depends in part on the judgment of the research staff in analyzing the UCS output cluster by cluster. The procedure employed is identical for every cluster, so that reference to one example will suffice to describe the process used for all clusters.

The following page contains a partial UCS printout of a cluster listing. Identification codes for the pivot and all cluster members are shown, along with each individual's similarity index ordered from high to low. To eliminate marginal cluster members, one proceeds from the top of the list and skips the first ten indices (which represent the minimum limit on cluster size). Continuing down the list, note that the similarity indices are sequentially continuous until one

Partial UCS Cluster Listing
(Electrical Task Area)

PRINT OUT OF TIES

IDENTIFICATION	1062411	SIMILARITY	34	CLUSTER	1	PMAN	1072427	VARIANCE	117
IDENTIFICATION	1062411	SIMILARITY	34	CLUSTER	5	PMAN	1052405	VARIANCE	117
IDENTIFICATION	1082428	SIMILARITY	6	CLUSTER	4	PMAN	1082407	VARIANCE	1
IDENTIFICATION	1082428	SIMILARITY	6	CLUSTER	5	PMAN	1052405	VARIANCE	1
IDENTIFICATION	1092416	SIMILARITY	29	CLUSTER	1	PMAN	1072427	VARIANCE	79
IDENTIFICATION	1092416	SIMILARITY	29	CLUSTER	5	PMAN	1052405	VARIANCE	79

THE SELECTION OF THE ACTUAL CLUSTERS FOLLOW

PIVOT MAN CLUSTERED ON 1072427

IDNUMBER	1072429	SIMILARITY WITH PIVOT MAN	57	VARIANCE	172
IDNUMBER	1072423	SIMILARITY WITH PIVOT MAN	52	VARIANCE	169
IDNUMBER	1082421	SIMILARITY WITH PIVOT MAN	48	VARIANCE	127
IDNUMBER	1052412	SIMILARITY WITH PIVOT MAN	46	VARIANCE	129
IDNUMBER	1052418	SIMILARITY WITH PIVOT MAN	43	VARIANCE	119
IDNUMBER	1092408	SIMILARITY WITH PIVOT MAN	41	VARIANCE	119
IDNUMBER	1102810	SIMILARITY WITH PIVOT MAN	36	VARIANCE	137
IDNUMBER	1072420	SIMILARITY WITH PIVOT MAN	36	VARIANCE	143
IDNUMBER	1052406	SIMILARITY WITH PIVOT MAN	36	VARIANCE	149
IDNUMBER	1102808	SIMILARITY WITH PIVOT MAN	35	VARIANCE	107
IDNUMBER	1062428	SIMILARITY WITH PIVOT MAN	34	VARIANCE	129
IDNUMBER	1062411	SIMILARITY WITH PIVOT MAN	34	VARIANCE	117
IDNUMBER	1092420	SIMILARITY WITH PIVOT MAN	33	VARIANCE	110
IDNUMBER	1062415	SIMILARITY WITH PIVOT MAN	32	VARIANCE	97
IDNUMBER	1072418	SIMILARITY WITH PIVOT MAN	32	VARIANCE	113
IDNUMBER	1062430	SIMILARITY WITH PIVOT MAN	32	VARIANCE	108
IDNUMBER	1062416	SIMILARITY WITH PIVOT MAN	31	VARIANCE	79
IDNUMBER	1082427	SIMILARITY WITH PIVOT MAN	30	VARIANCE	92
IDNUMBER	1072417	SIMILARITY WITH PIVOT MAN	30	VARIANCE	52
IDNUMBER	1092416	SIMILARITY WITH PIVOT MAN	29	VARIANCE	79
IDNUMBER	1062422	SIMILARITY WITH PIVOT MAN	29	VARIANCE	108
IDNUMBER	1052425	SIMILARITY WITH PIVOT MAN	29	VARIANCE	60
IDNUMBER	1092410	SIMILARITY WITH PIVOT MAN	22	VARIANCE	29
IDNUMBER	1062407	SIMILARITY WITH PIVOT MAN	19	VARIANCE	30
IDNUMBER	1082416	SIMILARITY WITH PIVOT MAN	16	VARIANCE	15

reaches those individuals with an index of 29. Thereafter, begins a series of gaps starting with a space of seven between the continuous similarities of 29 and the index of 22--as indicated by the arrow. If the last three individuals with low similarities were included in this cluster, it would dilute the homogeneity of the cluster disproportionately.

It is the identification of the significant interstice in a series of similarity indices that depends on the judgment of the researcher--although it is not so arbitrary as it may appear. If cluster verification scores (CVS) were computed for this cluster, starting with the initial ten indices and adding one additional individual each time, the first large drop in cluster homogeneity would appear at the same point (i.e., between 29 and 22) identified in this example.

In an identical manner, thresholds were set for each cluster to eliminate marginal contributors to cluster homogeneity. Verification scores (CVS) were then computed for all "refined" UCS clusters using the Cluster Verification routine discussed in a previous section (supra, pp. 25,26).

The next step in designating optimum specialty clusters is involved with the decision as to which set of clusters produced by UCS (and refined by setting thresholds) are to represent the homogeneous segments of work that are characteristic of an occupational area. Each iteration of the UCS produced a set of clusters utilizing the entire sample in a task area; thus some choice must be made among the k sets of clusters. Table 6 shows a partial array of cluster sets from the Propulsion/Auxiliary task area, beginning with three and terminating with fourteen clusters. The first two columns contain the identification code of the pivots (e.g., "72242") and their cluster number (e.g., "C_1"); the next and all succeeding columns, each contain a set of clusters showing the size of each cluster (m) in that set as well as the degree of internal homogeneity (as determined by the measure of mean similarity provided by CVS computations).

In selecting an optimum set of clusters to represent a given task area, there are a series of criteria which can be used to delimit the scope of the problem. Thus, the object in making a choice among alternative sets produced by UCS is to (1) maximize cluster homogeneity, (2) maximize the number of clusters representing the task area, (3) maximize the number of personnel (i.e., task patterns) accounted for within the bounds of the similarity thresholds, and (4) minimize the number of clusters that exceed the size constraints of 10 to 50.

Initially, half of the sets listed in Table 6 can be eliminated from consideration because they clearly exceed the criteria noted above. That is, of the 12 sets shown, six sets (S_1, S_2, S_9, S_{10}, S_{11}, and S_{12}) can be excluded because of the relatively small number of personnel accounted for in clustering and/or because of the relatively large number of clusters invalidated by exceeding the size constraints (i.e.,

TABLE 6

Summary Array of Partial UCS Output
for the Propulsion/Auxiliary Task Area

Cluster Number	Pivot Identif.		Set Number S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}	S_{11}	S_{12}
C_1	72242	m	65	59	47	47	43	43	43	43	32	30	28	24
		CVS	22.4	23.9	25.0	25.0	25.4	25.4	25.4	25.4	29.4	29.5	29.6	30.9
C_2	62033	m	45	43	23	23	22	21	21	19	19	19	19	19
		CVS	21.5	22.2	25.6	25.6	25.9	25.9	25.9	26.2	26.2	26.2	26.2	26.2
C_3	52215	m	65	50	50	36	42	42	32	32	23	23	23	23
		CVS	21.4	21.2	21.2	21.2	20.3	20.3	20.6	20.6	22.5	22.5	22.5	22.5
C_4	82224	m		39	51	21	21	21	16	16	16	16	16	16
		CVS		21.7	17.7	20.6	20.6	20.6	21.3	21.3	21.3	21.3	21.3	21.3
C_5	92021	m			31	31	25	23	23	22	17	17	15	15
		CVS			23.8	23.8	24.3	24.5	24.5	24.6	23.8	23.8	23.2	23.2
C_6	02218	m				36	36	36	21	20	17	17	17	17
		CVS				22.4	22.4	22.4	22.4	23.2	25.7	25.7	25.7	25.7
C_7	82025	m					20	19	19	19	15	11	11	9
		CVS					20.6	20.8	20.8	20.8	23.8	25.0	25.0	25.1
C_8	72249	m						9	9	9	7	6	6	6
		CVS						19.1	19.1	19.1	20.6	24.2	24.2	24.2
C_9	52234	m							35	35	35	35	26	26
		CVS							21.7	21.7	21.7	21.7	23.9	23.9
C_{10}	62015	m								6	5	4	4	4
		CVS								24.7	26.2	32.3	32.3	32.3
C_{11}	72208	m									9	9	8	8
		CVS									29.3	29.3	28.8	28.8
C_{12}	92031	m										8	8	7
		CVS										25.6	25.6	26.2
C_{13}	52014	m											7	6
		CVS											33.2	34.1
C_{14}	92010	m												9
		CVS												28.0
Number of Clusters			3	4	5	6	7	8	9	10	11	12	13	14
No. Personnel Clustered			175	191	202	194	209	214	219	221	195	195	188	189
No. Clusters <10 and >50			2	1	1	0	0	1	1	2	3	4	5	7

clusters which are <10 and >50). Of the remaining six sets, S_3 can be excluded because of the small number of clusters (i.e., 5), and S_4, because of the relatively small number of personnel clustered (i.e., 194). S_8, although containing the highest number of clustered individuals (221), has two clusters of less than 10 respondents each. In the three sets left, there is little to choose in the way of cluster homogeneity among those clusters that can be commonly compared (i.e., C_1 to C_7). Therefore, the final selection must be made on the basis of number of clusters in a set and the number of personnel clustered. On both criteria, S_7 "optimizes" the choice--even though one cluster in that set is slightly undersized (C_8, where m=9). As a final check on this process, a computer program was developed to analyze internal cluster homogeneity in terms of the task pattern similarity between clusters.

B. Evaluation of Cluster Similarity Distance

In order to evaluate the task pattern differences between clusters, a computer program was developed to build a matrix of similarities parallel to the similarity matrix used for Cluster Verification (CVS); the output of which provides measures of inter-cluster distance. This is done by computing the mean value of all cells in the task pattern similarity matrix of two clusters. These values (termed Cluster Distance Scores or CDS) indicate the extent to which the clusters in a set, taken two at a time, are discrete or similar. Ideally, the difference in task patterns between clusters should be significantly greater than the difference in task patterns within clusters. Since the CVS and CDS are identical in terms of computational procedures, a direct comparison is possible. Table 7 contains a matrix of inter-cluster similarities for eight Propulsion/Auxiliary area clusters in set seven (although there were nine UCS clusters listed for S_7, the smallest cluster [C_8] was eliminated because of its low homogeneity and inadequate size).

From an analysis of this matrix, it is possible to evaluate the cluster set to determine which clusters are most similar and which are most discrete. It is not the absolute cluster distance score (CDS) that is important in this evaluation; instead, it is the size of the CDS relative to the internal homogeneity (CVS) of the two clusters being compared. In all comparisons, the CDS should be smaller than the mean similarity of either of the two clusters which make up the similarity distance matrix. If this were not the case (i.e., if the mean similarity between clusters were greater than that within clusters), the rationale for maintaining separate clusters would collapse. Table 7 shows there are no exceptions to this research expectation. Thus, there are more differences in task patterns between clusters than within clusters.

In some cluster pairings, the CDS indicates wide disparities in the work performed [e.g., between pairs (C_2,C_3);(C_2,C_4);(C_2,C_6);(C_2,C_8); (C_4,C_5); and (C_4,C_7)]. Of the 28 CDS cell entries for the eight Propulsion/Auxiliary task area clusters, the six lowest values are related to C_2 pairings and C_4 pairings. Conversely, of the nine most

TABLE 7

Cluster Distance Matrix for Eight Clusters
In the Propulsion/Auxiliary Task Area

Cluster Number	m	Intra-Cluster Similarity (CVS)	Inter-Cluster Similarity (CDS)						
			C_2	C_3	C_4	C_5	C_6	C_7	C_8
C_1	43	25.4	11.4	9.6	11.7	18.0	13.2	19.5	12.0
C_2	21	25.9		8.8	3.4	21.0	6.0	15.4	7.1
C_3	32	20.6			11.7	11.7	17.3	9.7	18.2
C_4	16	21.3				7.1	19.6	7.1	17.0
C_5	23	24.5					10.6	19.0	10.9
C_6	21	22.4						9.7	20.0
C_7	19	20.8							10.0
C_8	35	21.7							--

similar cluster pairings, three clusters (C_5, C_6, and C_8) account for half of the high CDS values. It is interesting to note that with one exception--i.e., (C_6,C_8)--the similarity of these three clusters among themselves is not particularly great.

With the designation of optimum specialty clusters noted previously, and the aid of output from the cluster distance program, it then becomes possible to describe an occupational field or task area in terms of its task pattern interaction. The relationships between relatively homogeneous segments of work requirements can be best illustrated in an n-dimensional space--which, unfortunately, is impossible in the planar surface of this report. Nevertheless, Table 7 does indicate the constituents of some of these relationships. For instance, a macro-cluster can be developed from C_4, C_6, and C_8--all of which have a considerable amount of mutual task pattern similarities. On the other hand, C_2 appears to be relatively independent of all other clusters except C_5.

The task pattern relationships described above are influenced to a very large degree by the source of the data. Inasmuch as the task patterns were derived from engineering personnel on destroyers, the similarities in tasks performed between individuals and between clusters of nominally different occupational areas are much greater than would be the case for other ship types or other work situations (e.g., industrial occupations), where the division of labor and specialization of functions are more prominent. Destroyers are generally characterized by jobs which evidence a large amount of overlapping in task patterns. Because of this, the specialty clusters produced from a matrix of task pattern similarities reflect this relative lack of specialization and are much more difficult to separate clearly. However, this does not invalidate the clustering process; the clusters produced by these techniques simply reflect the way in which tasks are performed in a specific work situation.

V. RESEARCH APPLICATIONS OF COMPUTER CLUSTERING TECHNIQUES

The primary application for computer clustering techniques in this research is in the area of task analysis. All of the data processing decisions and program designs have been directed toward the development of optimum specialty clusters. These clusters, which constitute groups of homogeneous task patterns, will be characterized by a series of technical, organizational, and communicational variables. By this process, clusters of work requirements will be developed--each cluster reflecting a particular profile of skills and knowledges.

Computer clustering techniques are not limited to task analysis alone. For instance, in the same research, the series of programs associated with UCS is being employed to determine existing patterns of communications networks in destroyers. With only minor modifications, these same clustering programs will employ an input of "contact lists" to produce clusters of communications patterns. A considerable amount of the work in this area has been heretofore limited to experimental networks of three to seven persons in a laboratory setting. With the advent of more advanced and sophisticated techniques, such as UCS, it becomes possible to test hypotheses about occupational and organizational behavior in actual shipboard situations. These homogeneous patterns of work contacts will be contrasted with "official" designations of organizational structure and formal work group arrangements, to determine cases of deviation and the circumstances under which such deviation occurs.

Methods of computer clustering can be adapted to a wide range of research problems, in addition to the above. Problems of unidimensional pattern recognition are especially suitable for UCS solution. In particular, this assortment of clustering programs provides quantitative criteria for research decisions that are frequently arbitrary, or based on "estimates," in other research techniques.

REFERENCES

1. Carr, M. J., "Differential Acquaintance and Association Among Neighbors in an Area of Low Socioeconomic Status," (Unpublished M.A. Thesis), Indiana University, Bloomington, Indiana, 1948.

2. Cotterman, T. E., Task Classification: An Approach to Partially Ordering Information on Human Learning, Dayton, Ohio: Wright Air Development Center, Jan. 1959 (WADC Technical Note 58-374).

3. Haggard, D. F., The Feasibility of Developing a Task Classification Structure for Ordering Training Principles and Training Content, Fort Knox, Kentucky: Human Resources Research Office, Research Memorandum, Jan. 1963.

4. International Labour Office, Job Evaluation, Geneva, Switzerland: ILO, 1960.

5. MacCaslin, E. F., A Tentative Taxonomy of Task Demands, Paper read at APA Convention, Sept. 1963.

6. Miller, R. B., Task Taxonomy: Inventive Approach, Paper read at APA Convention, Sept. 1965.

7. Otis, J. L. and Leukart, R. H., Job Evaluation, Englewood Cliffs, N. J.: Prentice-Hall, 1954.

8. Shartle, C. L., Occupational Information, Englewood Cliffs, N. J.: Prentice-Hall, 1959, pp. 137-144.

9. Shartle, op.cit., pp. 94-134.

10. Silverman, J., A Method for Structuring Technical Tasks, San Diego: U. S. Naval Personnel Research Activity, Aug. 1965 (STB 66-4).

11. Silverman, J. and Carr, M. J., Method Development for Basic Technical Skills Research, San Diego: U. S. Naval Personnel Research Activity, May 1965 (SRR 65-4).

12. Stolurow, L. M., The Classification of Learning Tasks: A Systems Approach, University of Illinois: Training Research Laboratory, 1960 (Memo. Rep. 12).

13. Sweetser, F. L., "Neighborhood Acquaintance and Association," (Unpublished Ph.D. Dissertation), Columbia University, New York, 1941.

14. U. S. Employment Service, Dictionary of Occupational Titles, U. S. Department of Labor, Vols. I and II, 1965.

/

SELECTED BIBLIOGRAPHY OF CLUSTER ANALYSIS AND ASSOCIATED TECHNIQUES

This bibliography contains a selection of books, professional articles, and other publications which focus on the problem of defining, describing, measuring, and recognizing groupings of entities. In the behavioral sciences, this interest would focus upon one or more common features of human groups or patterns of human behavior. But the techniques employed to classify, group, or cluster humans on the basis of some criterion of similarity are not necessarily different in kind from those techniques used on the same type of problem by physicists, mathematicians, computer designers, information theorists, and electronic engineers. Unfortunately, there appears to be relatively little interaction on the part of scientists from diverse disciplines who, nevertheless, are concerned with similar technical problems.

The selection of publications which follows represents an attempt to bring together some of the wide variety of literature concerned with cluster analysis, pattern recognition, hierarchical grouping, factor analysis, profile grouping, and other clustering, classifying, and taxonomic techniques. Chronologically, only 25 percent of the items listed were published prior to 1960, and there are no items dated before 1949-50. Thus, the emphasis in this bibliography has been on the currency of research. Further, the stress is on statistical techniques--particularly those employing computerized procedures--rather than non-quantitative methods of analysis. Most of the entries in this bibliography have been reviewed in the course of developing the UCS technique. However, there are a number of items which are still unevaluated in terms of the research problem concerned in this report--their inclusion is based on the possibility of stimulating greater inter-disciplinary exchange than now exists.

Abraham, C. T., "Evaluation of Clusters on the Basis of Random Graph Theory," (Unpublished Report), Yorktown Heights, New York: IBM Corp.

Abramson, N. and Braverman, D., "Learning to Recognize Patterns in a Random Environment," IRE Trans. Information Theory, September 1962, IT-8(5), pp. 558-563.

Albert, A., "A Mathematical Theory of Pattern Recognition," Annals of Mathematical Statistics, 1963, 34(1), pp. 284-299.

Allais, D. C., The Selection of Measurements for Prediction, Stanford, Calif.: Stanford Electronics Laboratory, Nov. 1964 (TR6103-9).

Ball, G. H. and Hall, D. J., Isodata, A Novel Method of Data Analysis and Pattern Classification, Menlo Park, Calif.: Stanford Research Institute, 1965.

Bledsoe, W. W., "A Corridor-Projection Method for Determining Orthogonal Hyperplanes for Pattern Recognition," (Unpublished Report), Palo Alto, Calif.: Panoramic Research Corp., 1963.

Bledsoe, W. W. and Browning, I., "Pattern Recognition and Reading by Machine," Proc. Eastern Joint Computer Conference, 1959, pp. 225-233.

Block, H. D., "The Perceptron: A Model for Brain Functioning, I," Reviews of Modern Physics, 1962, 34(1), pp. 123-135.

Block, H. D., Knight, B. W., and Rosenblatt, F., "Analysis of a Four-Layer, Series-Coupled Perceptron, II," Reviews of Modern Physics, 1962, 34(1), pp. 135-142.

Block, H. D., Nilsson, N. J., and Duda, R. O., "Determination and Detection of Features in Patterns," in Tou, J. T. and Wilcox, R. H., (eds.), Computer and Information Sciences, Washington, D. C.: Spartan Books, 1964, pp. 75-110.

Bonner, R. E., "On Some Clustering Techniques," IBM Journal of Research and Development, Jan. 1964, pp. 22-32.

Bottenberg, R. A. and Christal, R. E., An Iterative Technique for Clustering Criteria Which Retains Optimum Predictive Efficiency, Lackland Air Force Base, Texas: Personnel Research Laboratory, Mar. 1961 (WADD-TN-61-30).

Braverman, D. J., Machine Learning and Automatic Pattern Recognition, Stanford, Calif.: Stanford Electronics Laboratory, Feb. 1961 (TR2003-1).

Brennan, E. J., An Analysis of the Adaptive Filter, Syracuse, New York: General Electric Corp., (Electronics Laboratory Technical Information Series Report R61 ELS-20), 1961.

Brick, D., A Mathematical Approach to Pattern Recognition and Self Organization, Waltham, Mass.: Sylvania Applied Research Laboratory, June 1962 (Research Memorandum No. 296).

Cardinet, J., "The Use of Profiles for Differential Classification," Educational and Psychological Measurement, 1959, 19(2), pp. 192-205.

Cattell, R. B., "A Note on Correlation Clusters and Cluster Search Methods," Psychometrika, 1944, 9(3), pp. 169-184.

Cattell, R. B., "r_p and Other Coefficients of Pattern Similarity," Psychometrika, 1949, 14, pp. 279-298.

Chow, C. K., "An Optimum Character Recognition System Using Decision Functions," IRE Trans. Electronic Computers, December 1957, EC-6(4), pp. 247-254.

Coombs, C. H. and Satter, G. A., "A Factorial Approach to Job Families," Psychometrika, 1949, 14, pp. 33-42.

Cooper, D. B. and Cooper, P. W., "Nonsupervised Adaptive Signal Detection and Pattern Recognition," Journal of Information and Control, 1964, 7(3).

Cooper, P. W., "The Hyperplane in Pattern Recognition," Cybernetica, 1962, 5(4).

Cooper, P. W., "The Hypersphere in Pattern Recognition," Journal of Information and Control, 1962, 5(4), pp. 324-346.

Cooper, P. W., "Statistical Classification with Quadratic Forms," Biometrika, December 1963, 50, pp. 439-448.

Cronbach, L. J. and Gleser, G. C., "Assessing Similarity Between Profiles," Psychological Bulletin, 1953, 50, pp. 456-473.

Daly, R. F., The Adaptive Binary Detection, Stanford, Calif.: Stanford Electronics Laboratory, June 1961 (TR2003-2).

duMas, F. M., "The Coefficient of Profile Similarity," Journal of Clinical Psychology, 1949, 5, pp. 123-131.

Dunnette, M. D. and England, G. W., "A Checklist for Differentiating Engineering Jobs," Personnel Psychology, 1957, 10, pp. 191-198.

Dunnette, M. D. and Kirchner, W. K., "A Checklist for Differentiating Different Kinds of Sales Jobs," Personnel Psychology, 1959, 12, pp. 191-198.

Dwyer, P. S., "Solution of the Personnel Classification Problem with the Method of Optimal Regions," Psychometrika, 1954, 19, pp. 11-26.

Farley, B. C. and Clark, W. A., "Generalization of Pattern Recognition in a Self-Organizing System," Proc. Western Joint Computer Conference, 1955.

Farley, B. C. and Clark, W. A., "Simulation of Self-Organizing Systems by Digital Computer," IRE Trans. Information Theory, September 1954, IT-4, pp. 76-84.

Fierschein, G. and Fischler, M., "Automatic Subclass Determination for Pattern Recognition Applications," IRE Trans. Electronic Computers, April 1963, EC-12(2), pp. 137-141.

Fortier, J. J. and Solomon, H., Clustering Procedures, Stanford, Calif.: Stanford University, Department of Statistics, March 1964 (Tech. Report No. 7).

Fralick, S. C., The Synthesis of Machines Which Learn Without a Teacher, Stanford, Calif.: Stanford University, April 1964 (Tech. Report No. 6103-8).

Fruchter, B., Introduction to Factor Analysis, Princeton, N. J.: Van Nostrand, 1954.

Fu, K. S., "A Sequential Decision Model for Optimum Recognition," in Bernard, E. E. and Kare, M. R., (eds.), Biological Prototypes and Synthetic Systems, New York: Plenum, 1962, p. 270.

Fu, K. S., "A Statistical Approach to the Design of Intelligent Machines - Pattern Recognition and Learning," Cybernetica, 1962, 5(2), pp. 88-102.

Gaier, E. L. and Lee, M. C., "Pattern Analysis: The Configural Approach to Predictive Measurement," Psychological Bulletin, 1953, 50, pp. 140-148.

Gamba, A., "New Developments in Artificial Intelligence and Pattern Recognition," in Tou, J. T. and Wilcox, R. H., (eds.), Computer and Information Sciences, Washington, D. C.: Spartan Books, 1964, pp. 219-229.

Ghiselli, E. E., The Measurement of Occupational Aptitude, Berkeley, Calif.: University of California Press, 1955.

Glaser, E. M., "Signal Detection by Adaptive Filters," IRE Trans. Information Theory, April 1961, IT-7(2), pp. 87-98.

Griffin, J. S., Jr., King, J. H., Jr., and Tunis, C. J., "A Pattern-Identification Device Using Linear Decision Functions," in Tou, J. T. and Wilcox, R. H., (eds.), Computer and Information Sciences, Washington, D. C.: Spartan Books, 1964, pp. 169-193.

Guilford, J. P., Psychometric Methods, New York: McGraw-Hill, 1954.

Harman, H. H., Modern Factor Analysis, Chicago: University of Chicago Press, 1960.

Harrison, R., "Cumulative Communality Cluster Analysis of Workers' Job Attitudes," Journal of Applied Psychology, 1961, 45, pp. 123-125.

Helmstadter, G. C., "An Empirical Comparison of Methods for Estimating Profile Similarity," Educational and Psychological Measurement, 1957, 17, pp. 71-82.

Highleyman, W. H., "Linear Decision Functions, with Application to Pattern Recognition," Proc. Institute of Radio Engineers, June 1962, 50(6), pp. 1501-1514.

Hyvarinen, L., "Classification of Qualitative Data," British Information Theory Journal, 1962, pp. 83-89.

Jacobs, I. M., Classification of Noisy Patterns Following a Finite Learning Period, Waltham, Mass.: Sylvania Applied Research Laboratory, March 1961 (Research Memorandum No. 239).

Jakowotz, C. V., Shuey, R. L., and White, G. M., "Adaptive Waveform Recognition," in Cherry, E. C., (ed.), Information Theory, London: Butterworth, 1961.

Kamentsky, L. A. and Liu, C. N., "A Theoretical and Experimental Study of a Model for Pattern Recognition," in Tou, J. T. and Wilcox, R. H., (eds.), Computer and Information Sciences, Washington, D. C.: Spartan Books, 1964, pp. 194-218.

Kaskey, G., Krishnaiah, P. R., and Azzari, A., "Cluster Formation and Diagnostic Significance in Psychiatric Symptom Evaluation," Proc. Fall Joint Computer Conference, 1962, 22, p. 285.

Kazmierczak, H. and Steinbuch, K., "Adaptive Systems in Pattern Recognition," IEEE Trans. Electronic Computers, December 1963, EC-12(6).

Lewis, P. M., II, "The Characteristic Selection Problem in Recognition Systems," IRE Trans. Information Theory, February 1962, IT-8(2), pp. 171-178.

Lorr, M., "A Comparison of Two Methods of Cluster Analysis," Outpatient Studies in Psychiatry, Pre-publication Report No. 25, Sept. 1965.

Marill, T. and Green, D. M., "On the Effectiveness of Receptors in Recognition Systems," IRE Trans. Information Theory, January 1963, IT-9(1), pp. 11-17.

Marill, T. and Green, D. M., "Statistical Recognition Functions and the Design of Pattern Recognizers," IRE Trans. Electronic Computers, December 1960, EC-9(4), pp. 472-477.

Mattson, R. L. and Dammann, J. E., "A Technique for Determining and Coding Subclasses in Pattern Recognition Problems," March 1965, (Submitted for Publication to IBM Journal of Research and Development).

McCormick, E. J., Finn, R., and Scheips, C. D., "Patterns of Job Requirements," Journal of Applied Psychology, 1957, 41, pp. 358-364.

McCracken, R. R., "Job Cluster Analysis in Terms of Aptitude Activity Elements," Dissertation Abstracts, 1959, 20, pp. 1861-1862.

McQuitty, L. L., "Comprehensive Hierarchical Analysis," Educational and Psychological Measurement, 1960, 20, pp. 805-816.

Michener, C. D. and Sokal, R. R., "A Quantitative Approach to a Problem in Classification," Evolution, June 1957, 11, pp. 130-162.

Needham, R. M., The Theory of Clumps, II, Cambridge, England: The Cambridge Language Research Unit, March 1961 (Report M.L. 139).

Nunnally, J., "The Analysis of Profile Data," Psychological Bulletin, 1962, 59(4), pp. 311-319.

Okajima, M., Stark, L., Whipple, G. H., and Yasui, S., "Computer Pattern Recognition Techniques: Some Results with Real Electrocardiographic Data," IEEE Trans. Bio-Medical Electronics, July 1963, BME-10 (3), pp. 106-114.

Orr, D. B., "The Distance Measure as a Statistic for Clustering Jobs," Dissertation Abstracts, 1956, 16, pp. 1290-1291.

Orr, D. B., "A New Method for Clustering Jobs," Journal of Applied Psychology, 1960, 44, pp. 44-49.

Osgood, C. E. and Suci, G. J., "A Measure of Relation Determined by Both Mean Difference and Profile Information," Psychological Bulletin, 1952, 49, pp. 251-262.

Palmer, G. J. and McCormick, E. J., "A Factor Analysis of Job Activities," Journal of Applied Psychology, 1961, 45, pp. 289-294.

Parker-Rhodes, A. F., Contributions to the Theory of Clumps, I, Cambridge, England: The Cambridge Language Research Unit, March 1961 (Report M.L. 138).

Patrick, E. A. and Hancock, J. C., The Non-Supervised Learning of Probability Spaces and Recognition of Patterns, Lafayette, Indiana: Purdue University, 1965 (Technical Report).

Rimoldi, H. J. A. and Grib, T. F., "Pattern Analysis," British Journal of Statistical Psychology, 1960, 13, pp. 137-149.

Rogers, D. J. and Tanimoto, T. T., "A Computer Program for Classifying Plants," Science, Oct. 1960, 132, pp. 1115-1118.

Sawrey, W. L., Keller, L., and Conger, J. J., "An Objective Method of Grouping Profiles by Distance Functions and its Relation to Factor Analysis," Educational and Psychological Measurement, 1960, 20, pp. 651-674.

Sebestyen, G. S., Classification Decisions in Pattern Recognition, Cambridge, Mass.: MIT-Research Laboratory of Electronics, April 1960 (Tech. Report No. 301).

Sebestyen, G. S., Decision-making Processes in Pattern Recognition, New York: Macmillan, 1962.

Sebestyen, G. S., "Pattern Recognition by an Adaptive Process of Sample Set Construction," IRE Trans. Information Theory, September 1962, IT-8(5), pp. S-82-S-91.

Sebestyen, G. S., "Recognition of Membership in Classes," IRE Trans. Information Theory, January 1961, IT-7(1), pp. 48-50.

Sebestyen, G. S. and Edie, J., Pattern Recognition Research, Bedford, Mass.: Air Force Cambridge Research Laboratory, June 1964 (Report 64-821, AD 608 692).

Selfridge, O. G., "Pattern Recognition and Learning," in Cherry, E. C., (ed.), Information Theory, London: Butterworth, 1956.

Selfridge, O. G. and Neisser, U., "Pattern Recognition by Machine," Scientific American, August 1960, 203(2), pp. 60-68.

Smith, J. W., "The Analysis of Multiple Signal Data," IEEE Trans. Information Theory, July 1964, IT-10(3), pp. 208-214.

Sokal, R. R. and Michener, C. D., "A Statistical Method for Evaluating Systematic Relationships," The University of Kansas Science Bulletin, Mar. 20, 1958.

Sokal, R. R. and Sneath, P. N. A., Principles of Numerical Taxonomy, San Francisco: W. H. Freeman & Co., 1963.

Spilker, J. J., Jr., Luby, D. D., and Lawhorn, R. D., Progress Report--Adaptive Binary Waveform Detection, Palo Alto, Calif.: Philco Western Development Laboratory (Communication Sciences Department Tech. Report No. 75), December 1963.

Stark, L., Okajima, M., and Whipple, G. H., "Computer Pattern Recognition Techniques: Electrocardiographic Diagnosis," Communications of the Assn. for Computing Machinery, Oct. 1962, 5, pp. 527-531.

Steinbuch, K. and Piske, U. A. W., "Learning Matrices and Their Applications," IEEE Trans. Electronic Computers, December 1963, EC-12(6), pp. 846-862.

Thomas, L. L., "A Cluster Analysis of Office Operations," Journal of Applied Psychology, 1952, 36, pp. 238-242.

Thorndike, R. L., "Who Belongs in the Family?," Psychometrika, 1953, 18, pp. 267-276.

Thorndike, R. L., Hagen, E. P., Orr, D. B., and Rosner, B., An Empirical Approach to the Determination of Air Force Job Families, Lackland Air Force Base, Texas: AF Personnel & Training Research Center, Aug. 1957 (AFPTRC-TR-57-5).

Tomlinson, H., Classification of Information Topics by Clustering Interest Profiles, Lackland Air Force Base, Texas: Personnel Research Laboratory, Nov. 1965 (PRL-TR-65-19).

Tou, J. T. and Wilcox, R. H., (eds.), Computer and Information Sciences: Collected Papers on Learning, Adaptation and Control in Information Systems, Washington, D. C.: Spartan Books, 1964.

Tryon, R. C., "Communality of a Variable: Reformulation by Cluster Analysis," Psychometrika, 1957, 22, pp. 241-260.

Tryon, R. C., "Cumulative Communality Cluster Analysis," Educational and Psychological Measurement, 1958, 18, pp. 3-35.

Tryon, R. C., "General Dimensions of Individual Differences: Cluster Analysis versus Multiple Factor Analysis," Educational and Psychological Measurement, 1958, 18, pp. 477-495.

Turner, R. D., "First-Order Experiential Concept Formation," in Bernard, E. E. and Kare, M. R., (eds.), Biological Prototypes and Synthetic Systems, New York: Plenum Co., 1962.

Uhr, L., Pattern Recognition: Theory, Experiment, Computer Simulations, and Dynamic Models of Form Perception and Discovery, New York: Wiley, 1965.

Uhr, L. and Vossler, C., "A Pattern Recognition Program that Generates, Evaluates, and Adjusts Its Own Operators," Proc. Western Joint Computer Conference, 1961, pp. 555-569.

Ward, J. H., Jr., Hierarchical Grouping to Maximize Payoff, Lackland Air Force Base, Texas: Personnel Research Laboratory, Mar. 1961 (WADD-TN-61-29).

Ward, J. H., Jr. and Hook, M. E., A Hierarchical Grouping Procedure Applied to a Problem of Grouping Profiles, Lackland Air Force Base, Texas: Personnel Research Laboratory, Oct. 1961 (ASD-TN-61-55).

Widrow, B., "Pattern Recognition and Adaptive Control," Proc. Joint Automatic Control Conference, June 1962.

Widrow, B. and Smith, F. W., "Pattern-Recognizing Control Systems," in Tou, J. T. and Wilcox, R. H., (eds.), Computer and Information Sciences, Washington, D. C.: Spartan Books, 1964, pp. 288-317.

APPENDICES

APPENDIX A

Task List Questionnaire

TASK LIST INSTRUCTIONS for HULL AND REPAIR AREA

1. The Task List on the following pages should be filled out only by enlisted personnel working in the Hull and Repair area of the Engineering Department. This includes DC's, SF's, MR's, and strikers for these ratings. It also includes personnel of other ratings assigned to this area.

2. The Task List is divided into 11 subject headings. Read the subject heading first to determine if the heading applies to your present work area.

   -If it applies to your work, then read each task below the heading and make an "X" on the line following each task if you have actually performed the task in your present assignment on this ship within the past 3 months.

   -If the heading does not apply to your work, go on to the next subject heading.

3. Many of the tasks contain several different parts. Be sure to check the task if you perform any of the parts, even though you do not perform all the parts.

4. Remember-

   -Do NOT check any tasks just because you "know how" to do them, or because you did them in school or in past duty assignments.

   -Do NOT check tasks which, during the past 3 months, you have supervised only.

   -Do NOT check tasks when you give only minor assistance, such as handing parts or tools to another man who is actually performing the task.

5. Do not hesitate to ask questions if you need assistance.

13. Perform angular, compound, and differential indexing; cut spur gears, T-slots and dovetails using milling machine. 13.______

14. Perform spline cutting and broaching; cut spur, bevel, helical and worm gears using milling machine. 14.______

15. Perform balancing machine operations. 15.______

16. Stow, lubricate, adjust, and clean shop equipment, machines and tools. 16.______

17. Lubricate machine tool bearings, guide-rollers, fittings and designated parts; fill oil holes and oil cups; and change oil. 17.______

18. Clean exposed surfaces of all machines and tools. 18.______

19. Check and adjust leveling of machine foundations. 19.______

20. Perform machining operations using lathe grinding attachments and milling attachments. 20.______

B. PIPEFITTING (Plumbing, Steamfitting, Pipe Covering, Piping and Valve Work)

1. Make temporary repairs to pipe with plugs, clamps, plastic, or patches. 1.______

2. Make permanent repairs to pipes with plugs (rivet or screws), welded or brazed patches, or by straightening and aligning. 2.______

3. Replace piping sections and fittings. 3.______

4. Layout and assemble sections of piping using templates and targets, pipe bending machines, and cutting-burring-threading machines. 4.______

5. Hydrostatically test pipes, tubes, valves and fittings. 5.______

6. Clean and flush piping and plumbing lines. 6.______

7. Determine cause of troubles in flushing and firemain systems. 7.______

8. Install, patch and repair pipe lagging and insulation, and molded pipe covering on steam, water and refrigeration lines. 8.______

APPENDIX B

Initial Cluster Program Output

Part 1. Partial Similarity Listing and Variance Listing (Hull/Repair Task Area)

ID NUMBER 1	ID NUMBER 2	SIMILARITY
52402	52408	
52402	52410	22
52402	52413	29
52402	52419	16
52402	52421	
52402	52422	20
52402	52423	21
52402	62402	13
52402	62403	29
52402	62406	16
52402	62419	16
52402	62425	16
52402	62431	16
52402	72402	29
52402	72407	33
52402	72408	9
52402	72416	10
52402	72419	19
52402	72424	16
52402	72426	20
52402	82404	24
52402	82409	16
52402	82419	6
52402	82420	13
52402	82422	14
52402	82423	27
52402	82425	13
52402	92402	25
52402	92405	11
52402	92411	16
52402	92417	12
52402	92418	7
52402	92421	17
52402	2404	15
52402	2405	17
52402	2408	19
52402	2409	31
52402	2411	18
52402	2413	2
52402	2414	21
52408	52410	
52408	52413	
52408	52419	
52408	52421	
52408	52422	
52408	52423	
52408	62402	
52408	62403	
52408	62406	
52408	62419	
52408	62425	
52408	62431	
52408	72402	

ID NUMBER	VARIANCE
52402	47
52410	130
52413	46
52419	120
52422	96
52423	115
62402	148
62403	135
62406	105
62419	147
62425	19
62431	86
72402	29
72407	127
72408	66
72416	15
72419	53
72424	98
72426	51
82404	90
82409	88
82419	118
82420	81
82422	128
82423	111
82425	64
92402	19
92405	93
92411	138
92417	111
92418	131
92421	97
2404	113
2405	35
2408	119
2409	71
2411	22
2413	44
2414	59

Part 2. Semi-Matrix (Hull/Repair Task Area)

	52410	52413	52419	52422	52423	62402	62403	62406	62419	62425	62431	72402	72407	72408	72416	72419	72424	72426	82404	82409	82419	82420	82422	82423	82425	92402	92405	92411	92417	92418	92421	02404	02405	02408	02409	02411	02413	02414
52402	22	29	16	20	21	13	29	16	16	16	16	29	33	9	10	19	16	20	24	16	6	13	14	27	13	25	11	16	12	7	17	15	17	19	31	18	2	21
52410		19	45	42	46	36	11	34	40	15	33	28	13	22	5	26	49	28	47	32	32	34	36	9	26	19	29	39	32	32	35	37	7	43	14	14	14	33
52413			16	18	20	12	37	17	15	20	13	29	33	13	10	16	18	19	22	15	10	14	15	30	13	25	14	16	12	8	16	15	17	16	28	22	4	20
52419				34	43	35	10	33	38	16	35	19	11	22	5	24	34	24	35	30	30	32	35	9	26	18	32	39	29	33	35	34	7	44	12	10	16	33
52422					40	32	11	37	37	19	30	23	13	22	5	30	32	28	37	34	29	30	39	9	26	19	32	37	31	26	37	34	7	37	16	17	15	31
52423						34	12	35	39	20	31	26	14	21	5	26	33	27	43	34	28	30	41	9	25	20	32	41	30	35	38	36	8	39	15	15	14	33
62402							2	39	52	17	31	19	4	30	3	25	28	22	34	32	26	26	33	2	25	15	27	34	36	38	29	32	2	35	6	8	18	25
62403								8	6	14	8	18	46	3	12	10	9	10	13	9	2	8	9	51	6	19	6	7	4	1	9	9	24	11	36	24	1	13
62406									44	26	34	23	10	23	4	27	31	23	32	36	28	27	40	6	28	19	32	33	30	30	32	35	7	35	12	14	16	32
62419										22	34	22	8	27	2	25	34	29	35	34	28	29	36	4	25	19	28	35	38	37	34	33	5	40	10	12	18	32
62425											16	15	13	23	9	23	15	26	20	23	14	17	23	12	20	27	23	16	18	13	23	19	15	17	16	14	9	23
62431												21	11	19	3	27	30	18	25	38	28	27	31	6	24	16	25	36	23	27	21	31	6	39	15	16	17	28
72402													21	14	7	19	21	17	32	21	14	17	19	15	18	19	19	21	17	14	16	18	11	21	24	16	5	19
72407														3	14	10	10	10	16	10	2	9	10	37	9	22	7	8	5	2	10	10	24	13	44	24	0	14
72408															4	29	20	24	22	24	24	25	30	3	26	15	20	27	26	22	27	26	4	22	8	6	16	22
72416																7	1	6	6	5	0	4	5	12	6	8	5	3	3	3	3	4	19	6	14	9	0	6
72419																	28	28	25	29	26	30	37	7	28	16	26	28	22	24	26	33	10	30	15	13	16	26
72424																		28	33	24	35	22	28	6	19	13	31	34	29	24	31	38	6	37	14	12	16	29
72426																			27	17	20	25	26	9	28	17	31	24	30	22	31	26	13	22	12	11	7	32
82404																				29	29	28	35	11	26	20	30	35	31	26	33	37	8	38	17	12	12	31
82409																					28	29	41	7	25	21	27	39	28	28	24	33	7	35	13	19	16	30
82419																						26	32	2	25	10	31	35	27	25	26	33	1	32	5	9	15	30
82420																							39	5	28	18	24	33	24	36	26	33	9	35	13	10	22	26
82422																								6	31	20	35	45	32	36	37	41	8	40	13	15	20	32
82423																									5	20	6	6	4	1	5	6	15	8	29	25	2	11
82425																										15	32	26	25	27	30	31	10	22	10	11	5	22
92402																											12	16	12	13	18	13	12	20	24	16	7	20
92405																												33	32	32	30	31	7	28	9	14	8	32
92411																													36	34	33	39	3	40	12	14	20	34
92417																														31	29	32	3	33	8	11	15	35
92418																															26	28	3	33	4	8	21	25
92421																																37	8	33	12	11	11	28
02404																																	8	39	14	14	18	32
02405																																		6	23	8	0	9
02408																																			17	14	22	34
02409																																				23	6	17
02411																																					7	17
02413																																						16

Part 3. Frequency Distribution of Similarity Indices

Format

0/64	1/64	2/64	3/64	4/64	5/64	6/64	7/64
8/64	9/64	10/64	11/64	12/64	13/64	14/64	15/64
16/64	17/64	18/64	19/64	20/64	21/64	22/64	23/64
24/64	25/64	26/64	27/64	28/64	29/64	30/64	31/64
32/64	33/64	34/64	35/64	36/64	37/64	38/64	39/64
40/64	41/64	42/64	43/64	44/64	45/64	46/64	47/64
48/64	49/64	50/64	51/64	52/64	53/64	54/64	55/64
56/64	57/64	58/64	59/64	60/64	61/64	62/64	63/64

Propulsion/Auxiliary Task Area (m=278)

2439	1928	2090	2084	2275	2116	2022	1976
2167	1856	1795	1700	1559	1404	1283	1028
1174	944	867	763	700	619	587	446
483	386	354	295	264	198	151	98
139	72	67	40	37	33	17	13
6	8	6	2	5	2		1
				1			1
1		1					

Hull/Repair Task Area (m=39)

4	5	10	12	11	15	22	16
19	20	21	14	19	21	24	22
33	18	14	22	19	12	22	14
20	19	28	16	27	19	21	19
29	24	20	21	11	15	7	11
7	4	1	3	3	2	2	1
			1	1			

Electrical Task Area (m=75)

70	51	93	85	132	152	159	137
146	118	104	69	73	74	64	53
81	66	67	85	66	43	59	67
64	63	55	62	53	52	53	31
49	25	33	28	22	9	9	7
3	7	3	3	2	1	1	1
1	1			1		1	
	1						

APPENDIX C

Cluster Identification Analysis

The examination of potential specialty clusters in the initial program was accomplished with the aid of a special program labelled "cluster identification." This program arrays the cluster data in a table which facilitates visual examination of the structural characteristics of different clusters. This table shows the identification code of cluster members, their respective variances, their presence in one or more clusters and their similarity to the pivot in those clusters, and their status (whether clustered or unclustered, pivot or non-pivot). By comparing a series of these cluster identification tables and calculating a few summary statistics, the analysis of specialty clusters can proceed more effectively.

Cluster identification tables were computed for different similarity thresholds in the three task areas. Those computer runs in which the control percentage was set at 10% for each of the three task areas are shown on pages 65-67. Table 8 contains a summary of cluster identification data for 12 experimental cluster runs.

There are a number of observations that can be made through analysis of Table 8. First, it is clear that as the control percentage (CP) increases, the similarity threshold (ST) decreases. The reason for this is based on the method of obtaining thresholds by using the similarity distribution, as noted previously. Note also, that the range in variances between the first cluster's pivot (P_1) and the last cluster's pivot (P_t) in a given run increases as the ST decreases. Thus, in the Electrical area, the control percentage set at 5% yields an ST=33 and a pivot range of 173-97, while the percentage set at 20% yields an ST=25 and a pivot range of 173-41. In terms of the criteria for selection of specialty clusters, the higher similarity thresholds result in greater homogeneity in each cluster because of the more restrictive cluster entry requirement, and also improve the quality of the pivots associated with the "trailing" clusters because of their higher variance.

Second, the number of clusters differs for each run. The evaluation of this factor is based on the criterion concerned with "optimizing" the number and size of clusters. In order to develop specialty clusters, the initial clusters (C_1) should not be surfeited with personnel (as in the Propulsion/Auxiliary area run at CP=20 corresponding to ST=16), nor should the "trailing" clusters (C_t) be too small (as in the Propulsion/Auxiliary area run at CP=5 corresponding to ST=25). One can obtain a good idea of the kinds of "trade-offs" required by the different cluster structures in simply fulfilling the criteria of cluster number and size.

TABLE 8

Summary of Cluster Identification Tables

Program Run			Cluster Size			Number of Clusters	Percent of Personnel Unclustered	Percent of Personnel with Multiple Membership
CP	ST	Pivot Range	Initial Cluster	Terminal Cluster	Difference			
%	64ths	$(P_1 \rightarrow P_t)$	(C_1)	(C_t)	(C_1-C_t)			
Propulsion/ Auxiliary (m = 278)								
5	25	97-38	42	2	40	31	30	54
10	21	97-33	63	11	52	19	23	63
15	19	97-20	76	4	72	27	14	65
20	16	97-15	109	10	99	21	10	72
Hull/Repair (m = 39)								
5	38	148-135	3	3	0	3	64	0
10	35	148-135	6	5	1	3	49	10
15	33	148-86	11	7	4	5	31	37
20	32	148-81	12	7	5	7	33	46
Electrical (m = 75)								
5	33	173-97	13	6	7	12	51	35
10	30	173-90	18	4	14	6	41	49
15	27	173-60	25	2	23	9	22	47
20	25	173-41	26	5	21	10	15	47

Third, the number and percent of unclustered personnel (those with similarities to pivots $\leq$ST) also differs for each run. In the Hull/Repair area, for instance, the percent of personnel unclustered runs from 64% at ST=38 to 33% at ST=32. Thus, with a threshold difference of only 6/64ths, the percent of unclustered personnel almost doubles in size. In selecting specialty clusters it is desirable to minimize the percent of unclustered personnel so that a major portion of the different task patterns sampled is included in the cluster analysis. Table 8 shows the effect of decreasing the percent of unclustered personnel: namely, reducing the ST and, therefore, the degree of homogeneity in each cluster. As with the other criteria of cluster "optimality" some trade-off must be made.

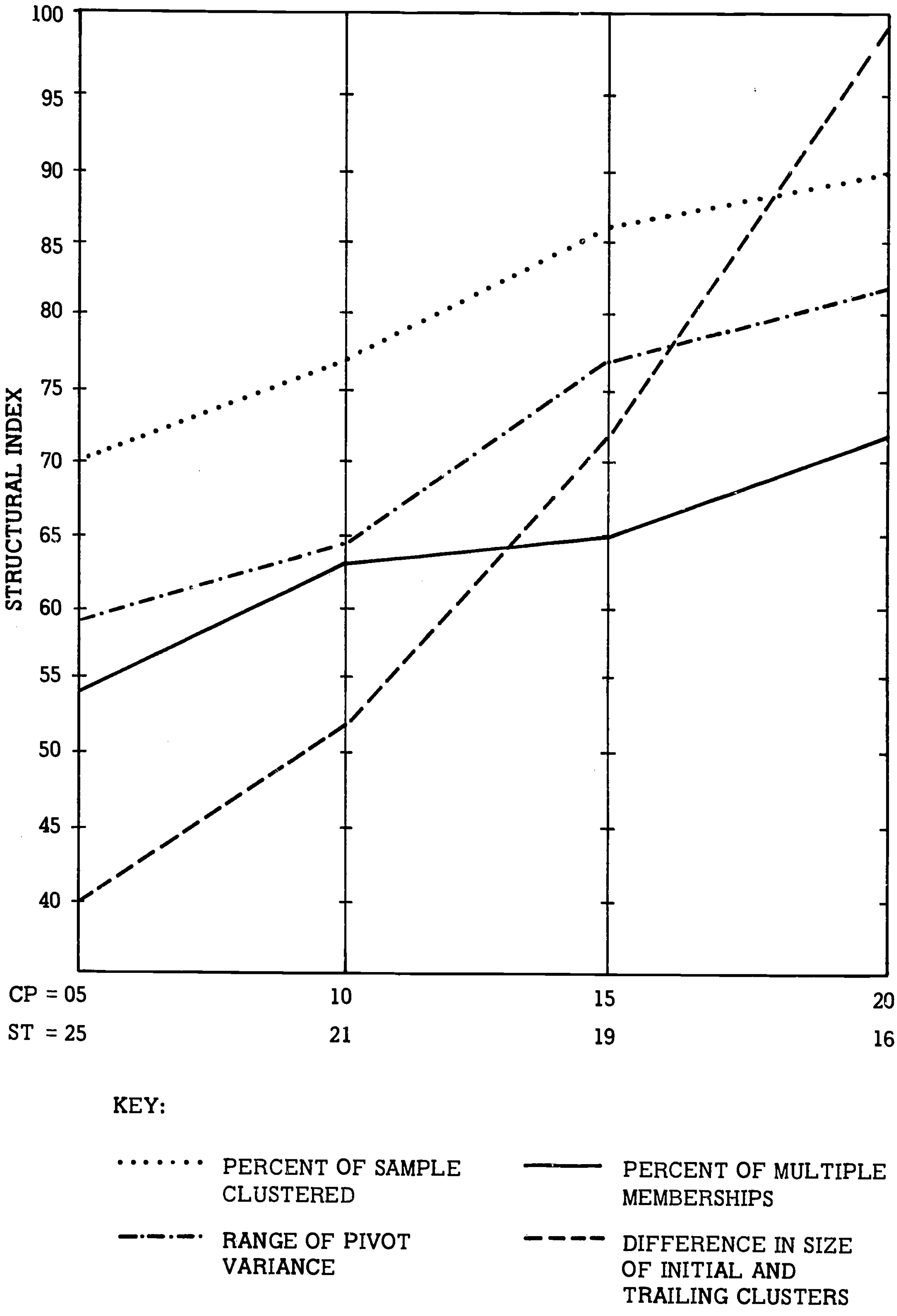
Fourth, the number of multiple memberships must also be considered in selecting specialty clusters. Multiple memberships occur when an individual has a similarity >ST to more than one pivot, and thereby becomes a member of more than one cluster. In some cluster runs, the number of multiple cluster memberships can be quite high. This is undesirable because it results in overlapping task patterns among clusters and tends to dilute the homogeneity of the clusters in which the individual appears. Table 8 shows the percent of multiple memberships for each run increasing as the ST decreases. For example, in the Propulsion/Auxiliary area the percent of multiple membership runs from 54% at ST=25 to 72% at ST=16.

On the basis of most criteria of cluster selection, the higher thresholds seem to provide the "optimal" clusters. There is, however, the problem of high numbers of unclustered personnel at those thresholds. Figure 4 shows the interrelationships of some structural features for different Propulsion/Auxiliary computer runs. The four sets of cluster run characteristics charted in Figure 4 are linearly related to CP (positive) and ST (negative).

As a result of the considerations noted above, an intensive examination of the program logic behind the pivot and cluster selection techniques led to some refinements in the pivot theory as well as methods for obtaining optimum specialty clusters.

FIGURE 4

Relationship of Selected Structural
Features of the Initial Clustering Technique



CLUSTER IDENTIFICATION TABLE

TASK LIST 4 PERCENT = 10 ST = 21

CLUSTER NUMBERS

ID	VARIANCE	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	UNCL
52006	38										23				23				33															
52008	81	42														22																		
52010	70	33				25				23																								
52011	39	22																																
52013	65			26	27			30						23																				
52014	87	37	24			22																												
52016	75	35														25																		
52017	76	34				23																												
52019	62	28														24																		
52020	53	24				22					23																							
52023	57	23				25				26		22																						
52026	69	27	25			23				25		29																						
52027	83	37	24			22																												
52028	67	38																																
52029	37										23								30															
52201	13																																	*
52202	15																																	*
52203	25																																	*
52205	39											24																						
52206	81	25		38				27	22					23																				
52207	13																																	*
52208	55		32																															
52209	16																																	*
52210	73	22		37				22						27																				
52211	83			XX																														
52212	29						27																											
52215	62			22	34			25																										
52216	56			27	25			22																										
52217	53			32																														
52218	41											XX																						
52219	48				24			24					25																					
52222	26																																	*
52223	42				22								29							23														
52225	5																																	*
52226	43				29		23						27																					
52227	31			22																														
52228	8																																	*
52229	46				25			22																										
52232	28																																	*
52233	46		26			22				24																								
52234	80			35	24		26	23						24																				
52235	71			35				22																										
52404	16																																	*
52409	44								XX																									
52414	33												22																					
52416	28								24																									
52420	68	32																																
62007	73	33																																
62008	26	22																																

CLUSTER IDENTIFICATION TABLE

TASK LIST 5

PERCENT = 10

ST = 55

CLUSTER NUMBERS

ID	VARIANCE	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	UNCL
52402	47																																	*
52410	130	36	39																															
52413	46			37																														
52419	120		39																															
52422	96		37																															
52423	115		41																															
62402	148	XX																																
62403	135			XX																														
62406	105	39																																
62419	147	52																																
62425	19																																	*
62431	86		36																															
72402	29																																	*
72407	117			46																														
72408	66																																	*
72416	15																																	*
72419	53																																	*
72424	98																																	*
72426	51																																	*
82404	90																																	*
82409	88		39																															
82419	118																																	*
82420	81																																	*
82422	128		45																															
82423	111			51																														
82425	64																																	*
92402	19																																	*
92405	93																																	*
92411	138		XX																															
92417	111	36	36																															
92418	131	38																																
92421	97																																	*
2404	113		39																															
2405	35																																	*
2408	119		40																															
2409	71			36																														
2411	22																																	*
2413	44																																	*
2414	59																																	*

CLUSTER IDENTIFICATION TABLE

TASK LIST 6 PERCENT = 10 ST = 30

CLUSTER NUMBERS

ID	VARIANCE	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	UNCL
52204	4																																	*
52403	12																																	*
52405	97				32																													
52406	149	36			36	35																												
52407	97				36																													
52411	116		37	35																														
52412	129	46																																
52415	138		35	40																														
52417	80																																	*
52418	119	43			31																													
52424	109				XX																													
52425	60																																	*
52426	78																																	*
62405	55																																	*
62407	30																																	*
62409	15																																	*
62411	117	34																																
62412	36																																	*
62413	60																																	*
62414	90			31																														
62416	79	31																																
62417	102		31	32																														
62421	79		36																															
62422	108					XX																												
62428	129	34			34	33	35																											
62430	108	32				31																												
62433	93					32																												
62434	16																																	*
72403	61																																	*
72404	41																																	*
72410	135		34	41																														
72414	73																																	*
72417	52																																	*
72418	113	32				32																												
72420	143	36			36																													
72423	169	52			34																													
72427	173	XX																																
72428	82																																	*
72429	172	57				31																												
72430	47																																	*
72434	42																																	*
72437	60																																	*
82403	72																																	*
82406	103		41	32																														
82407	63																																	*
82412	70		33																															
82415	97	32																																
82416	18																																	*
82418	77				31																													

APPENDIX D

Pivot Optimization Listing (Electrical Task Area)

PROGRAM-CLUSTER MODIFICATION
ALPHA .500

1ST PIVOT MAN	1072427	VARIANCE	173													
MANS ID	1102812	VARIANCE	145	SIMILARITY	4	COMPUTED VALUE	.0276	RELATED PIVOT MAN	1072427							
MANS ID	1052405	VARIANCE	97	SIMILARITY	22	COMPUTED VALUE	.2268	RELATED PIVOT MAN	1072427							
MANS ID	1072410	VARIANCE	135	SIMILARITY	32	COMPUTED VALUE	.2370	RELATED PIVOT MAN	1102812							
MANS ID	1052406	VARIANCE	149	SIMILARITY	36	COMPUTED VALUE	.2416	RELATED PIVOT MAN	1072427							
MANS ID	1052415	VARIANCE	138	SIMILARITY	39	COMPUTED VALUE	.2826	RELATED PIVOT MAN	1072410							
MANS ID	1072420	VARIANCE	143	SIMILARITY	41	COMPUTED VALUE	.2867	RELATED PIVOT MAN	1052406							
MANS ID	1062411	VARIANCE	117	SIMILARITY	34	COMPUTED VALUE	.2906	RELATED PIVOT MAN	1072427							
MANS ID	1062428	VARIANCE	129	SIMILARITY	38	COMPUTED VALUE	.2946	RELATED PIVOT MAN	1052406							
MANS ID	1072423	VARIANCE	169	SIMILARITY	52	COMPUTED VALUE	.3077	RELATED PIVOT MAN	1072427							
MANS ID	1102810	VARIANCE	137	SIMILARITY	44	COMPUTED VALUE	.3212	RELATED PIVOT MAN	1052406							
MANS ID	1062433	VARIANCE	93	SIMILARITY	30	COMPUTED VALUE	.3226	RELATED PIVOT MAN	1062428							
MANS ID	1062430	VARIANCE	108	SIMILARITY	35	COMPUTED VALUE	.3241	RELATED PIVOT MAN	1052406							
MANS ID	1062422	VARIANCE	108	SIMILARITY	35	COMPUTED VALUE	.3241	RELATED PIVOT MAN	1052406							
MANS ID	1102808	VARIANCE	107	SIMILARITY	35	COMPUTED VALUE	.3271	RELATED PIVOT MAN	1072427							
MANS ID	1092420	VARIANCE	110	SIMILARITY	36	COMPUTED VALUE	.3273	RELATED PIVOT MAN	1062428							
MANS ID	1052424	VARIANCE	109	SIMILARITY	36	COMPUTED VALUE	.3303	RELATED PIVOT MAN	1072420							
MANS ID	1072429	VARIANCE	172	SIMILARITY	57	COMPUTED VALUE	.3314	RELATED PIVOT MAN	1072427							
MANS ID	1102412	VARIANCE	147	SIMILARITY	49	COMPUTED VALUE	.3333	RELATED PIVOT MAN	1102812							
MANS ID	1062417	VARIANCE	102	SIMILARITY	35	COMPUTED VALUE	.3431	RELATED PIVOT MAN	1052415							
MANS ID	1092408	VARIANCE	119	SIMILARITY	41	COMPUTED VALUE	.3445	RELATED PIVOT MAN	1072427							
MANS ID	1052411	VARIANCE	116	SIMILARITY	41	COMPUTED VALUE	.3534	RELATED PIVOT MAN	1052415							
MANS ID	1102801	VARIANCE	115	SIMILARITY	41	COMPUTED VALUE	.3565	RELATED PIVOT MAN	1072410							
MANS ID	1052412	VARIANCE	129	SIMILARITY	46	COMPUTED VALUE	.3566	RELATED PIVOT MAN	1072427							
MANS ID	1102804	VARIANCE	98	SIMILARITY	35	COMPUTED VALUE	.3571	RELATED PIVOT MAN	1072420							
MANS ID	1052418	VARIANCE	119	SIMILARITY	43	COMPUTED VALUE	.3613	RELATED PIVOT MAN	1072427							
MANS ID	1072418	VARIANCE	113	SIMILARITY	41	COMPUTED VALUE	.3628	RELATED PIVOT MAN	1072420							
MANS ID	1092427	VARIANCE	107	SIMILARITY	39	COMPUTED VALUE	.3645	RELATED PIVOT MAN	1062417							
MANS ID	1092424	VARIANCE	108	SIMILARITY	40	COMPUTED VALUE	.3704	RELATED PIVOT MAN	1052415							
MANS ID	1052407	VARIANCE	97	SIMILARITY	36	COMPUTED VALUE	.3711	RELATED PIVOT MAN	1052424							
MANS ID	1082421	VARIANCE	127	SIMILARITY	48	COMPUTED VALUE	.3780	RELATED PIVOT MAN	1072427							
MANS ID	1092415	VARIANCE	90	SIMILARITY	35	COMPUTED VALUE	.3889	RELATED PIVOT MAN	1062428							
MANS ID	1082406	VARIANCE	103	SIMILARITY	41	COMPUTED VALUE	.3981	RELATED PIVOT MAN	1102412							
MANS ID	1082415	VARIANCE	97	SIMILARITY	39	COMPUTED VALUE	.4021	RELATED PIVOT MAN	1072423							
MANS ID	1082427	VARIANCE	92	SIMILARITY	37	COMPUTED VALUE	.4022	RELATED PIVOT MAN	1102810							
MANS ID	1082426	VARIANCE	98	SIMILARITY	42	COMPUTED VALUE	.4286	RELATED PIVOT MAN	1102412							
MANS ID	1062414	VARIANCE	90	SIMILARITY	44	COMPUTED VALUE	.4889	RELATED PIVOT MAN	1062417							

APPENDIX E

Cluster Selection Listings

Propulsion/Auxiliary Task Area

PIVOT MAN CLUSTERED ON 1102218

IDNUMBER	1052206	SIMILARITY WITH PIVOT MAN	44	VARIANCE	81
IDNUMBER	1102225	SIMILARITY WITH PIVOT MAN	39	VARIANCE	68
IDNUMBER	1102205	SIMILARITY WITH PIVOT MAN	39	VARIANCE	80
IDNUMBER	1072238	SIMILARITY WITH PIVOT MAN	38	VARIANCE	71
IDNUMBER	1082223	SIMILARITY WITH PIVOT MAN	36	VARIANCE	75
IDNUMBER	1052235	SIMILARITY WITH PIVOT MAN	34	VARIANCE	71
IDNUMBER	1052210	SIMILARITY WITH PIVOT MAN	34	VARIANCE	73
IDNUMBER	1092216	SIMILARITY WITH PIVOT MAN	32	VARIANCE	69
IDNUMBER	1082205	SIMILARITY WITH PIVOT MAN	32	VARIANCE	51
IDNUMBER	1072255	SIMILARITY WITH PIVOT MAN	32	VARIANCE	74
IDNUMBER	1052234	SIMILARITY WITH PIVOT MAN	32	VARIANCE	80
IDNUMBER	1102209	SIMILARITY WITH PIVOT MAN	31	VARIANCE	61
IDNUMBER	1082218	SIMILARITY WITH PIVOT MAN	30	VARIANCE	59
IDNUMBER	1072213	SIMILARITY WITH PIVOT MAN	30	VARIANCE	54
IDNUMBER	1062226	SIMILARITY WITH PIVOT MAN	29	VARIANCE	62
IDNUMBER	1062207	SIMILARITY WITH PIVOT MAN	29	VARIANCE	55
IDNUMBER	1072231	SIMILARITY WITH PIVOT MAN	28	VARIANCE	50
IDNUMBER	1072215	SIMILARITY WITH PIVOT MAN	28	VARIANCE	64
IDNUMBER	1082211	SIMILARITY WITH PIVOT MAN	27	VARIANCE	49
IDNUMBER	1062206	SIMILARITY WITH PIVOT MAN	27	VARIANCE	45
IDNUMBER	1052216	SIMILARITY WITH PIVOT MAN	27	VARIANCE	56
IDNUMBER	1092224	SIMILARITY WITH PIVOT MAN	26	VARIANCE	59
IDNUMBER	1072251	SIMILARITY WITH PIVOT MAN	26	VARIANCE	49
IDNUMBER	1082220	SIMILARITY WITH PIVOT MAN	25	VARIANCE	46
IDNUMBER	1072229	SIMILARITY WITH PIVOT MAN	25	VARIANCE	42
IDNUMBER	1062204	SIMILARITY WITH PIVOT MAN	24	VARIANCE	41
IDNUMBER	1082203	SIMILARITY WITH PIVOT MAN	23	VARIANCE	38
IDNUMBER	1102221	SIMILARITY WITH PIVOT MAN	22	VARIANCE	35
IDNUMBER	1092217	SIMILARITY WITH PIVOT MAN	22	VARIANCE	53
IDNUMBER	1062213	SIMILARITY WITH PIVOT MAN	20	VARIANCE	33
IDNUMBER	1052227	SIMILARITY WITH PIVOT MAN	20	VARIANCE	31
IDNUMBER	1062230	SIMILARITY WITH PIVOT MAN	19	VARIANCE	27
IDNUMBER	1102213	SIMILARITY WITH PIVOT MAN	18	VARIANCE	20
IDNUMBER	1102223	SIMILARITY WITH PIVOT MAN	17	VARIANCE	26
IDNUMBER	1082210	SIMILARITY WITH PIVOT MAN	17	VARIANCE	47
IDNUMBER	1082417	SIMILARITY WITH PIVOT MAN	12	VARIANCE	26
IDNUMBER	1062408	SIMILARITY WITH PIVOT MAN	11	VARIANCE	28
IDNUMBER	1102402	SIMILARITY WITH PIVOT MAN	9	VARIANCE	16
IDNUMBER	1092028	SIMILARITY WITH PIVOT MAN	9	VARIANCE	19
IDNUMBER	1052201	SIMILARITY WITH PIVOT MAN	8	VARIANCE	13
IDNUMBER	1102417	SIMILARITY WITH PIVOT MAN	7	VARIANCE	7
IDNUMBER	1052228	SIMILARITY WITH PIVOT MAN	7	VARIANCE	8
IDNUMBER	1062410	SIMILARITY WITH PIVOT MAN	3	VARIANCE	7
IDNUMBER	1092227	SIMILARITY WITH PIVOT MAN	2	VARIANCE	4

Hull/Repair Task Area

PRINT OUT OF TIES

IDENTIFICATION	1082420	SIMILARITY	26	CLUSTER	1	PMAN	1062402	VARIANCE	81
IDENTIFICATION	1082420	SIMILARITY	26	CLUSTER	3	PMAN	1082419	VARIANCE	81
IDENTIFICATION	1082425	SIMILARITY	25	CLUSTER	1	PMAN	1062402	VARIANCE	64
IDENTIFICATION	1082425	SIMILARITY	25	CLUSTER	3	PMAN	1082419	VARIANCE	64

THE SELECTION OF THE ACTUAL CLUSTERS FOLLOW

PIVOT MAN CLUSTERED ON 1062402

IDNUMBER	1062419	SIMILARITY WITH PIVOT MAN	52	VARIANCE	147
IDNUMBER	1062406	SIMILARITY WITH PIVOT MAN	39	VARIANCE	105
IDNUMBER	1092418	SIMILARITY WITH PIVOT MAN	38	VARIANCE	131
IDNUMBER	1092417	SIMILARITY WITH PIVOT MAN	36	VARIANCE	111
IDNUMBER	1052410	SIMILARITY WITH PIVOT MAN	36	VARIANCE	130
IDNUMBER	1102408	SIMILARITY WITH PIVOT MAN	35	VARIANCE	119
IDNUMBER	1052419	SIMILARITY WITH PIVOT MAN	35	VARIANCE	120
IDNUMBER	1082404	SIMILARITY WITH PIVOT MAN	34	VARIANCE	90
IDNUMBER	1052423	SIMILARITY WITH PIVOT MAN	34	VARIANCE	115
IDNUMBER	1082422	SIMILARITY WITH PIVOT MAN	33	VARIANCE	128
IDNUMBER	1082409	SIMILARITY WITH PIVOT MAN	32	VARIANCE	88
IDNUMBER	1052422	SIMILARITY WITH PIVOT MAN	32	VARIANCE	96
IDNUMBER	1062431	SIMILARITY WITH PIVOT MAN	31	VARIANCE	86
IDNUMBER	1072408	SIMILARITY WITH PIVOT MAN	30	VARIANCE	66
IDNUMBER	1092421	SIMILARITY WITH PIVOT MAN	29	VARIANCE	97
IDNUMBER	1082420	SIMILARITY WITH PIVOT MAN	26	VARIANCE	81
IDNUMBER	1082425	SIMILARITY WITH PIVOT MAN	25	VARIANCE	64
IDNUMBER	1072426	SIMILARITY WITH PIVOT MAN	22	VARIANCE	51
IDNUMBER	1072402	SIMILARITY WITH PIVOT MAN	19	VARIANCE	29
IDNUMBER	1102413	SIMILARITY WITH PIVOT MAN	18	VARIANCE	44
IDNUMBER	1062425	SIMILARITY WITH PIVOT MAN	17	VARIANCE	19

Electrical Task Area

PRINT OUT OF TIES

THE SELECTION OF THE ACTUAL CLUSTERS FOLLOW

PIVOT MAN CLUSTERED ON 1072427

IDNUMBER	1072429	SIMILARITY WITH PIVOT MAN	57	VARIANCE	172
IDNUMBER	1072423	SIMILARITY WITH PIVOT MAN	52	VARIANCE	169
IDNUMBER	1082421	SIMILARITY WITH PIVOT MAN	48	VARIANCE	127
IDNUMBER	1052412	SIMILARITY WITH PIVOT MAN	46	VARIANCE	129
IDNUMBER	1052418	SIMILARITY WITH PIVOT MAN	43	VARIANCE	119
IDNUMBER	1092408	SIMILARITY WITH PIVOT MAN	41	VARIANCE	119
IDNUMBER	1102810	SIMILARITY WITH PIVOT MAN	36	VARIANCE	137
IDNUMBER	1072420	SIMILARITY WITH PIVOT MAN	36	VARIANCE	143
IDNUMBER	1052406	SIMILARITY WITH PIVOT MAN	36	VARIANCE	149
IDNUMBER	1102808	SIMILARITY WITH PIVOT MAN	35	VARIANCE	107
IDNUMBER	1062428	SIMILARITY WITH PIVOT MAN	34	VARIANCE	129
IDNUMBER	1062411	SIMILARITY WITH PIVOT MAN	34	VARIANCE	117
IDNUMBER	1092420	SIMILARITY WITH PIVOT MAN	33	VARIANCE	110
IDNUMBER	1082415	SIMILARITY WITH PIVOT MAN	32	VARIANCE	97
IDNUMBER	1072418	SIMILARITY WITH PIVOT MAN	32	VARIANCE	113
IDNUMBER	1062430	SIMILARITY WITH PIVOT MAN	32	VARIANCE	108
IDNUMBER	1062416	SIMILARITY WITH PIVOT MAN	31	VARIANCE	79
IDNUMBER	1082427	SIMILARITY WITH PIVOT MAN	30	VARIANCE	92
IDNUMBER	1072417	SIMILARITY WITH PIVOT MAN	30	VARIANCE	52
IDNUMBER	1052424	SIMILARITY WITH PIVOT MAN	30	VARIANCE	109
IDNUMBER	1092416	SIMILARITY WITH PIVOT MAN	29	VARIANCE	79
IDNUMBER	1062422	SIMILARITY WITH PIVOT MAN	29	VARIANCE	108
IDNUMBER	1052425	SIMILARITY WITH PIVOT MAN	29	VARIANCE	60
IDNUMBER	1102804	SIMILARITY WITH PIVOT MAN	28	VARIANCE	98
IDNUMBER	1052407	SIMILARITY WITH PIVOT MAN	26	VARIANCE	97
IDNUMBER	1082418	SIMILARITY WITH PIVOT MAN	25	VARIANCE	77
IDNUMBER	1062433	SIMILARITY WITH PIVOT MAN	24	VARIANCE	93
IDNUMBER	1092415	SIMILARITY WITH PIVOT MAN	23	VARIANCE	90
IDNUMBER	1092410	SIMILARITY WITH PIVOT MAN	22	VARIANCE	29
IDNUMBER	1052405	SIMILARITY WITH PIVOT MAN	22	VARIANCE	97
IDNUMBER	1102802	SIMILARITY WITH PIVOT MAN	21	VARIANCE	59
IDNUMBER	1092208	SIMILARITY WITH PIVOT MAN	21	VARIANCE	86
IDNUMBER	1052417	SIMILARITY WITH PIVOT MAN	21	VARIANCE	80
IDNUMBER	1092426	SIMILARITY WITH PIVOT MAN	19	VARIANCE	80
IDNUMBER	1092406	SIMILARITY WITH PIVOT MAN	19	VARIANCE	40
IDNUMBER	1062407	SIMILARITY WITH PIVOT MAN	19	VARIANCE	30
IDNUMBER	1072414	SIMILARITY WITH PIVOT MAN	18	VARIANCE	73
IDNUMBER	1082424	SIMILARITY WITH PIVOT MAN	17	VARIANCE	74
IDNUMBER	1082416	SIMILARITY WITH PIVOT MAN	16	VARIANCE	18
IDNUMBER	1072403	SIMILARITY WITH PIVOT MAN	16	VARIANCE	61
IDNUMBER	1072430	SIMILARITY WITH PIVOT MAN	15	VARIANCE	47
IDNUMBER	1062405	SIMILARITY WITH PIVOT MAN	15	VARIANCE	55
IDNUMBER	1062413	SIMILARITY WITH PIVOT MAN	14	VARIANCE	60
IDNUMBER	1082407	SIMILARITY WITH PIVOT MAN	13	VARIANCE	63
IDNUMBER	1102807	SIMILARITY WITH PIVOT MAN	11	VARIANCE	43
IDNUMBER	1072434	SIMILARITY WITH PIVOT MAN	10	VARIANCE	42
IDNUMBER	1102811	SIMILARITY WITH PIVOT MAN	9	VARIANCE	11
IDNUMBER	1062412	SIMILARITY WITH PIVOT MAN	9	VARIANCE	36
IDNUMBER	1052403	SIMILARITY WITH PIVOT MAN	8	VARIANCE	12
IDNUMBER	1082428	SIMILARITY WITH PIVOT MAN	2	VARIANCE	1

APPENDIX F

Cluster Verification Listing
(Propulsion/Auxiliary Task Area)

CLUSTER 4

82224	52211	72221	72225	2205	92219	92214	52206	52235	52210	2218	62219	52234	92211	72238	62231
19	20	19	18	18	18	19	18	18	17	18	17	18	16	18	17
97	127	121	97	121	126	114	124	103	112	117	86	116	70	99	91
52217	92226	92216	62207	82211	72215	62226	92224	72229	82220	2212	72251	72227	72240	52409	52203
16	14	17	16	15	16	16	16	15	15	14	15	13	12	14	11
74	71	101	73	68	76	100	74	59	61	80	73	53	32	76	35
62228	52232	82228	62213	52227	2221	2220	62230	82210	72217	62420	82413	72248	62426	92209	52416
11	11	14	12	13	13	11	12	13	13	13	12	10	11	8	11
33	31	76	37	39	49	29	44	68	75	93	78	34	56	16	49
52212	92414	82417	52202	2222	82225	62408	72425	2219	92425	82429	82227	62424	2417	52404	52228
9	10	11	9	8	8	11	8	7	8	7	7	9	7	7	6
39	44	55	15	20	22	71	33	25	36	18	19	71	12	43	14
52201	2402	2031	2416	2406	92227	62410									
7	8	3	2	3	2	4									
19	33	9	4	17	6	10									

CLUSTER NUMBER = 4 TASK LIST = 5
NUMBER OF SUBJECTS = 71
MEAN = 12.24
STANDARD DEVIATION = 8.98

DISTRIBUTION LIST

Chief of Naval Operations (OP-01)
(OP-07)
Bureau of Naval Material (Code A3m)
Bureau of Naval Personnel (Pers-A3) (25)
Office of Naval Research (Code 458)
Center for Naval Analysis
Naval Post Graduate School
Naval Personnel Research Laboratory (3)
U. S. Marine Corps (G1)
Office of Research and Development, U. S. Army
HumRRO, U. S. Army
Personnel Management Division, U. S. Army
Personnel Research Laboratory, Lackland AFB
Human Factors Operations Research Laboratory, Bolling AFB
U. S. Employment Service, Department of Labor
Personnel Research and Development Center, Civil Service Commission
Science and Technology Division, Library of Congress
Defense Documentation Center (20)