Cluster Analysis for the Computer-Assisted Statistical Analysis of Melodies

Luigi Logrippo and Bernard Stepien

...notoriamente no hay clasificacion del universo que no sea arbitraria y conjetural. La razon es muy simple: no sabemos que cosa es el universo.
—*J.L. Borges, "El idioma analitico de John Wilkins"*

Computer usage in musicology seems to be developing somewhat more slowly than in other research areas in the humanities. Many papers in the field limit themselves to discussing musical databases, and elementary types of statistical analyses, involving note or interval counts. With some exceptions (for example, Morando, 1979 and Steinbeck, 1976) there is little mention of the powerful statistical techniques that have become everyday tools in other areas. In this paper we will show that some of these techniques can be adapted with relative ease to the needs of research in musicology. Part of our research used standard software packages which, with some adaptation, proved quite adequate to the task.

Our main statistical method is cluster analysis. Cluster analysis attempts to classify a set of entities according to their similarity from the point of view of some predefined set of characters, or "indicators." For example, if the set of entities being considered is a set of melodies, and the set of indicators is the set of notes used in each melody, cluster analysis will group together melodies using similar sets of notes.

This method has been used to classify melodies from several points of view. We will limit ourselves to the presentation of examples in which songs are classified by scales (i.e., notes used), intervals, and similarity of melodic patterns. The results are immediately useful to a musicologist attempting to

Luigi Logrippo and Bernard Stepien are in the Department of Computer Science at the University of Ottawa.

classify melodies by modes, for which scales, intervals, and melodic patterns are among the most important characters of identification. Other classification criteria investigated include melodic contours, rhythmic contours, and rhythmic patterns. The musicological results of this research were presented in Logrippo, Pelinski and Stepien (1981) and Pelinski (1981a).

Our project consisted of two phases. In the first, over one hundred monophonic songs were placed in a database, and analyzed using a variety of packages, one of which was a cluster analysis package. These melodies were Inuit personal ("a-ja-jai") songs, some collected by Pelinski (1981) in the settlements of Rankin Inlet and Eskimo Point (two corpora), others collected by Hauser (1977) in Thule. These are the three "corpora" discussed in Classifications of Songs by Tonal Density, and Combined Tonal/Interval Density Analysis. After completing this phase (in 1980), the second author developed some ideas on how to use cluster analysis for classifying melodies by similarity of melodic patterns. He wrote a software package to do this, and tested it on some other small corpora of melodic fragments. This more recent work is also discussed below in Similarity of Melodies.

Related work was previously reported by Wenker (1977) and Morando (1979). The first author presented a computer-assisted statistical analysis of scales and intervals in North-American folk music, without however using cluster analysis or factor analysis. The second author used factor analysis in order to uncover statistical relationships between the musical styles of several classical composers.

Cluster Analysis
Cluster analysis was developed in the sixties by various researchers, and today is extensively used in such diverse fields as the social and health sciences, biology,

and physical sciences. For example, it is used to classify insects in entomology, and to detect patterns of behavior in psychology. There is an abundant literature on this topic (Sneath and Sokal, 1973). An example of utilization of methods similar to the one discussed here in literary analysis can be found in Usher and Najock (1982). The specific implementation of cluster analysis used in the first phase of the project was BMDP (Dixon and Brown, 1979), in its version of November 1979. BMDP is a biomedical statistical package developed at the Health Sciences Computing Facility of the University of California (Los Angeles).

Statistical Analysis of Melodies. The basic concept involved in cluster analysis is quite simple. If a number of points are scattered on a sheet of paper, it may be possible to observe the fact that two of these points, A and B, are "closer" to each other than C and D, etc. Similarly, if A, B, and C are close together in an area of the sheet, it may be said that they form a "cluster." A point on a sheet of paper is an entity characterized by its two Cartesian coordinates. These coordinates (in cluster analysis often called "indicators") can be interpreted in many different ways, for example as musical parameters, as we shall show in the following example.

Let us consider five extremely simple melodies (P, Q, R, S, and T) all consisting of just two notes, say Cs and Gs. Let us suppose that we have statistics on the percentages of occurrences of Cs and Gs in the five melodies, as follows:

	C	G
P	20%	40%
Q	40%	30%
R	50%	40%
S	40%	10%
T	50%	10%

For example, 20% of the notes appearing in P are Cs, and 40% are Gs (the remaining 40% being rests). We can now apply clustering by seeing each melody as a point with two coordinates in a two-dimensional space. If we take the number of occurrences of Cs as the first coordinate, and the number of occurrences of Gs as the second one, we get the following picture (for simplicity, we reduce our scale by a factor of 10):

```
4 |       P           R
3 |              Q
2 |
1 |_____S___T_____
    1    2    3    4    5    6
```
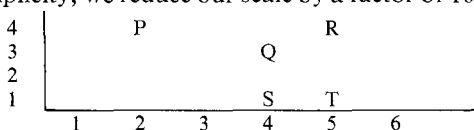
Clearly, the mutual distances of points in this picture reflect to some extent mutual similarities of the related melodies from the point of view of tonal structure. The picture easily reveals the two clusters (S,T) and (Q,R), the second made up of two melodies mutually less closely related than the first. Further, melody P is clearly detached from the others, an "outlier" according to statistical terminology. These facts are obvious without computer assistance, because we considered a space of only two dimensions. Our task would quickly become impossible if we start considering songs of three, four, five or more notes, which would require three, four, five or more dimensional spaces. Fortunately, the procedure can easily be automated, and with computer assistance even spaces of hundreds of dimensions present no problem.

We will continue the example above to show how this can be done. We begin by calculating the "distance" between points in our space. For example, the distance between melodies S and T is 1 (the 10% difference in the use of Cs) and between Q and S is 2. In order to find the distance between Q and T, we can apply the Pythagorean Theorem, which yields a distance of 2.23. Continuing in this way, we obtain the complete distance table which is:

```
     ------
Q  |2.23|
     -----------
R  | 3  |1.41|
     -----------------
S  |3.60| 2  |3.16|
     -----------------------
T  |4.24|2.23| 3  | 1  |
     -----------------------
      P    Q    R    S
```

We then proceed to find the clusters. We first note the smaller distances, then increasingly larger ones until all the melodies are classified. We observe that the two melodies in the pair (S,T) are mutually the closest, at a distance of 1, and this constitutes the first cluster. Next, we note that the second shortest distance in the table is 1.41, and this gives us the second cluster (Q,R). The third closest distance is 2, between S and Q. This enables us to join together the two clusters we have already found, and gives us (Q,R,S,T). Finally P joins this latter cluster at a distance of 2.23 from Q. This completes the classification of all the elements. Our results can be shown in the following "amalgamation" (or "classification") tree:

(the scale at the left shows the approximate distances at which the amalgamation occurred).

Therefore, the clustering algorithm proceeds in two steps:

1) First it computes a "coefficient of similarity," or "distance" between the entities in consideration, in our case, the melodies.

2) The distances are then used in the second step: the amalgamation process. This step will group the songs together in "clusters," starting with those that are mutually closest, and continuing with progressively more distant ones. The results of the amalgamation process can be shown in several forms, of which the two we have found the most useful are the amalgamation tree, discussed above, and the "sorted shaded distance matrix" introduced in the section Similarity of Melodies.

A note of caution is necessary here. Some readers might balk at our choice of "distances" between songs, which might seem arbitrary from a musicological point of view. For example, can we really justify the fact that Q and R, with a 10% difference in Cs and Gs, are mutually less different than Q and S, with a 20% difference in Gs? Before these methods become widely used, such questions will have to be satisfactorily answered. Research in this direction is already under way, as is noted by Krumhansl (1985). Tentatively, we propose the following two answers. From the statistical point of view, there are many different methods of calculating distances and performing the amalgamation. The "euclidian distance" method presented above is simply one possibility, and it was used in the example because it is the most easily understood. For the research discussed in the following section and the one on Combined Tonal Interval Density Analysis we preferred the chi square method as being the most appropriate: this choice was both recommended by the literature (Dixon and Brown, 1979) and verified empirically. Different methods, however, will satisfy different needs, just as one method may uncover relationships not shown by others. Experimentation is therefore not only possible but also desirable. (Some examples are discussed in Logrippo, Pelinski and Stepien, 1981.) Our second answer is more pragmatic: that our method can be justified by the results that it yields. It enabled us to discover relationships that were relevant for musicological analysis, and in this sense the method has proved itself.

In the first phase of our project, we used the method on each corpus of songs individually (i.e., Rankin Inlet, Eskimo Point and Thule) to understand the internal organization of the songs of a given Inuit settlement. Subsequently, we mixed the three corpora to verify the influence of geographic dispersion on the Inuit singing style. Finally, we tried to correlate classifications according to different indicators: i.e., tonal systems vs. interval systems, etc. For reasons of space, we are not able to discuss all this work in detail, and we will limit our discussion to the analysis of the Rankin Inlet corpus.

Classification of Songs by Tonal Density

The term "density" will be used to denote the proportion of notes or intervals in a song. The density of notes will be called "tonal density." In order to classify the songs by tonal density, we proceeded as follows. First, we performed on the encoded songs a simple frequency count program (not part of BMDP). For each song and note, we calculated the percentage of playing time of the note in relation to the total playing time of the song. This gave us a table of percentages of time values for each note in each song. In rearranged form this is Table 1. On this table, we ran BMDP, which produced the amalgamation tree in Fig. 1. Another simple program (also not part of BMDP) rearranged the table of percentages according to the results in Fig. 1, and this yielded the "sorted indicator table" shown in Table 1. This table, which shows the songs sorted by similarity together with their indicators, is not provided by the standard BMPD package.

In Fig. 1, the numbers at the top are song numbers, sorted in such a way as to group together similar songs as much as possible. The numbers at the left are "amalgamation distances," i.e., the distances at which the songs or groups of songs were clustered together. For example, from Fig. 1 it appears that songs 19 and 38 are very similar from the point of view of tonal density, since they were clustered together at the small distance of 1.085. On the other hand, songs 34 and 36, which are quite dissimilar (having been clustered together at the relatively large distance of 12.993), are also very dissimilar from the other songs in the corpus, since they join the closest cluster—the one consisting of all the songs appearing at the right side of the figure —only at the large distance of 24.067. The tonal density-sorted indicator table (Table 1) shows song numbers and song authors at the left. For each song and each note, it also shows the percentage of time value

of the note in the song (these percentages are the "indicators" discussed above). For example, note F accounts for 10.5% of the total played length of song 1, whose author is Kavik. Notes are represented according to the external representation discussed in Hickey and others (1979), i.e., -C is middle C, etc. Note G always appears very frequently because all songs were transposed so that G corresponds to the main recitation tone. This table gives us a clear picture of the tonal relationships among songs.

With Fig. 1 and Table 1 in hand, the classification of the songs by tonal density does not present difficulties (see Table 2). It is clear that songs 34, 36, 11, and 28 are tonally "unusual" songs, or "outliers." Songs 31 and 17 are not classified as outliers because they will also appear together in the interval density analysis, and therefore deserve to belong to a class of their own. Further, we detect a fairly large class of songs (class 1), characterized by the use of notes F, G, A, and (with the exception of two songs) C. This class does not seem to have any particularly interesting subclasses, with the exception of (29,27), two songs by Tiktak, which exhibit a low percentage of A and will reappear together at the interval analysis stage. Other subclasses appear, but they are either small or fail to be characterized by the strong occurrence (or non-occurrence) of certain notes. In fact, the internal structure of class 1 seems to be due mainly to variations in frequency, rather than to note replacements. It is quite possible that a larger sample would reveal a somewhat more interesting picture.

In class 3, characterized by notes E, G, A, and +D, the substructure is more sharply defined, giving rise to the two subclasses 3 ' (using C), and 3 " (using B and D instead). Similar criteria enable us to identify class 4, which is clearly the most strongly characterized of all and in fact will reappear intact in the interval analysis.

Looking at Table 1 from a more global point of view, one discovers a yet deeper organization that encompasses the whole set of songs. Schematically, this is shown in Fig. 3. There are five notes that appear in the greater majority of songs, and these are D, G, A, C, and +D. Class 1 is characterized by a strong F, class 2 is characterized by the replacement of A sharp for A, class 3 is characterized by a strong E, and class 4 is characterized by a strong F sharp and the absence of D and +D. Within class 3 we have the following structure: class 3 ' is characterized by the weakness

of D, and class 3 " by the replacement of B for C.

Classification of Songs by Interval Density

The basic information on interval density is provided in Fig. 2 (Interval Density Amalgamation Tree). Fig. 2 is to be read in the same way as Fig. 1. Table 3 is similar to Table 1, though the indicators here are percentages of intervals. Intervals are represented by the number of semitones they contain, e.g., -2 is the descending tone, O is the prime interval, and 4 (also written + 4) is the ascending major third. Therefore descending tones are 13.5% of all the intervals in song 1 by Kavik. It should be noted that, due to space limitations, only the central most important part of the table is shown.

According to the method used for the classification by tonal density, we obtain Table 4, which shows the classification by interval density. It should be noted that Table 4 contains more classes than Table 2, as is to be expected since there are more intervals than notes.

A global look at Fig. 2 again reveals some interesting structural facts, as shown by Fig. 4. The analysis of Fig. 4 is less clear-cut than the one of Fig. 3, because of the greater complexity of the picture. For example, class 2 represents a sort of "transition" between class 1 and class 3 (it is shown as related to class 1 in Fig. 4). The modal analysis to be given in the next section will show that class 2 consists of the two subclasses (17,31) and (38,19), of which only the first relates to class 1 (because of the -4 and + 4).

Combined Tonal Interval Density Analysis

The next logical step is to attempt to combine the previous two analyses. Songs that have both a similar tonal density and a similar interval density will therefore be classified together. The resulting classes of songs can be said to identify "modes," in the sense that two songs belonging to the same mode will have similar tonal and interval structures. (The authors are aware of the fact that the usual definition of modes in musicology involves much more than the distribution of notes and intervals, the only aspects considered here.)

It is by pure coincidence that we end up with eight modes, just as many as the medieval church modes. These are shown in Table 5. For example, songs (1, 13,20,2) which belong together in tonal class 1 ' and

interval class 1, form mode 1. Eight songs (i.e., approximately one quarter of the total), do not belong to any of these eight modes. As we have mentioned already, this analysis helps reveal some interesting subclasses of the tonal classes: such are (38,19), two very similar songs, (27,29), the aforementioned two songs by Tiktak, and (15,35,23), a subclass of tonal class 3″ characterized by the presence of +D.

Similarity of Melodies

At this point it would be legitimate to ask when two melodies are similar. In musicology this is a vexed question which recently has even been attacked by computer-assisted methods (Dillon and Hunter, 1982, and Stech, 1981). We use a different approach, based on cluster analysis techniques. The simplest way to do this would be to consider each note in a melody as an indicator, and define a distance on this basis. For example, two melodies that are note-by-note identical except for the fact that one uses E instead of C at some point, would be considered to be at a distance of 4, the number of semitones between C and E. Unfortunately, this method is too simplistic: for example, two melodies that are completely identical except for an extra note at the beginning of one of them might turn out to be at very large distances. Therefore, it is wise to begin by determining a distance by using a pattern-matching method, and then use this distance to perform amalgamation.

In this method, the proportion of in-sequence notes between two melodic lines (henceforth also called "modules") is used as a distance. By way of example, we may consider the following two modules:

module 1:	C	E		G	A	B
module 2:	A	C	F	E	B	A	G

They have only three in-sequence notes in common: C - E - G, which is 60% of common notes relative to the number of notes in the smallest module. We take 60 to be the distance between these two modules. The length of the smaller phrase is used as a reference because we are using the concept of degree of inclusion of one phrase into a larger phrase. Unfortunately, this percentage alone is not sufficient to establish a realistic distance, as shown by the following example. Consider the three following modules:

module 1:	C	D	B	G		
module 2:	D	C	A	B	G	D
module 3:	E	C	A	E	G	E	F	C

Both 1 and 2 are at a distance of 50% from 3. However, since in the first case a sequence of two notes is involved (C-G), while in the second case a sequence of three notes is involved (C-A-G), we must conclude that module 2 is more similar than module 1 to module 3. Therefore our concept of percentage distance must be corrected by a "secondary ranking" whereby the length of the common sequence is taken into consideration when the percentage distance is the same. This method enables us to establish directly the distance between any two melodic lines, without having to resort to concepts such as "euclidian distance" or "chi square distance" mentioned in the section Cluster Analysis.

In our current research, all notes are given the same weight. However, it would be possible to include a hierarchy of weights based on rhythmic values, position, and others.

Once all pairs of modules have been compared and the distance table is established, the amalgamation process is performed similarly. The output is a classification of the fragments using various graphics, which emphasize the relationships that have been found. One major difference between this work and the one discussed in the previous sections is that previously we were classifying whole songs, while here we are classifying short melodic fragments manually extracted from the songs. The following examples are taken from a set of French nursery rhymes.

One example is shown in the amalgamation tree of Fig. 5. Each "leaf" of the tree is a module and the module itself is given on the right. Specifically, the information on the right of the tree is to be read as follows: module number, cluster number (see below), and sequence of notes in the module. For example, module 8 belongs to cluster 0 and consists of the five notes shown. By looking at the tree, we see immediately that module 8 is quite dissimilar from all others and so it makes a class of its own. (Here we use the convention that cluster 0 is the cluster of all the outliers.)

There are, instead, other fragments that are mutually closely related. By aligning their common notes, one may be able to discover "root phrases" in a cluster. For example, cluster 1, consisting of modules (21,23,16), is characterized by the "root phrase" C-B-A-G. This may be done in the following way. The user specifies a "cutoff distance" for which all melodies being mutually closer than the distance are to be considered as belonging to the same cluster. In this example, the cutoff distance was 69%. The system then

proceeds to group the modules into clusters. For each cluster, the "root" of the cluster is taken to be the module that has the smallest distance from all the points in the cluster (the "center" of the cluster). The notes of the modules in the cluster are then aligned with the corresponding notes in the root. Two additional examples of this are given in Fig. 6, which is a small portion of an analysis involving 220 modules grouped in 42 clusters (no relation with those of Fig. 5). Clearly, a critical role in this process is played by the choice of the cutoff distance. Smaller cutoff distances will give smaller clusters and longer root sequences, hence there is again much room for experimentation. There are several open issues in this area. For example, a cluster could have several roots, etc.

The diagram shown in Fig. 7 is the "sorted and shaded distance matrix" mentioned in the section Cluster Analysis and is obtained by appropriately sorting and darkening the distance table discussed in that section. Fig. 7 refers to the example of Fig. 5 and is an alternative way of presenting the clusters of that figure. More closely related modules are grouped together, and their distances are represented by darker symbols. For example, the diagram indicates two main classes of mutually similar fragments: (22, 2, 13, 7, 6, 1, 18, 11) and (11, 19, 5, 14, 17, 15, 3, 4, 12, 10). This second class contains a subclass which is (5, 14, 17, 15, 3, 4), etc. We were now able to discover that module 11 belongs to the intersection of two classes, a fact not visible from the amalgamation tree, that demonstrates a main advantage of this type of display. Another instance of this phenomenon appeared already in the Cluster Analysis section, where the amalgamation tree did not show the fact that the pairs (P, Q) and (Q, T) have equal distances. In general, we can say that overlapping classes are hidden in the amalgamation tree but show up in the sorted shaded distance matrix. These matrices were found to be important tools in the research discussed in the sections on classification of songs by tonal density and the two following sections as well; however, they were not mentioned in that context for reasons of space.

More recently, we have been experimenting with an interactive system, in which the musicologist is able to intervene in the way clustering and alignment are performed. Therefore, human judgment can be used to solve some of the ambiguities that the computer can only solve in arbitrary ways.

Conclusions

We have presented some computer-assisted methods for classifying melodies. Even though our analysis could, in principle, have been done by hand, in practice this would have been extremely laborious and error-prone. The same software packages could be used equally well to classify hundreds of melodies, which would be most useful. Of course, large quantities would require large computers, and in addition some different type of processing may be necessary in order to further compress the information.

We would stress that our methods do not yield a "cut-and-dry" musicological analysis. They only provide tools with which the musicologist can experiment, especially in order to detect relationships that could not be easily found by the naked eye (or ear). In our view, one of the most important aspects of the methods is that they are simple to use and the results are easily checked by hand, by studying the various types of displays and relating them back to the initial data. Therefore, there is little chance that the methods may force the careful musicologist to draw false conclusions.

As the reader may have noticed from the description of the method given in Cluster Analysis, cluster analysis has a wide range of applicability. Any collection of objects having a set of precisely identifiable indicators (whatever they might be) can be classified in this way. Our method could then be used, with only slight adaptations, to classify musical excerpts by criteria such as the chords used, the rhythmic structures that appear in them, etc.

On the basis of their experience, the authors believe that these methods may one day be common tools in some types of musicological research. We realize that this claim may appear to be unsubstantiated in this paper, where we have shown only two ethnomusicological applications. However, the same methods could be applied to any corpus of melodies at hand, whether ethnomusicological or not. Gregorian chant, for example, seems to be an ideal area of application. It is true that each area of musicological research has its own established methods that are probably quite different from the ones suggested here. However, our methods should not be viewed as replacing existing ones, rather as augmenting them by providing additional information. As we have already mentioned above, the methods have a high degree of flexibility and are capable of much adaptation and customization. The

problem of adapting and customizing them to the various existing areas of musicological research is a major one, and may require much experimentation and research.

We must here raise the question of the manual pre-processing required in order for the method to be applied. In our research, the musical text was entered as a straightforward encoding of the musical notation. Only in the work discussed in the section Similarity of Melodies, some manual pre-processing was performed in order to segment the given melodies into "modules," which were then processed in the fashion described above. However, considerable additional preprocessing will probably be required in most cases in order to obtain meaningful results. The function of this preprocessing might be to include information such as relative "importance" of various notes (as determined for example by meter or accents), exclude some irrelevant groups of notes, etc. Clearly, the type and amount of preprocessing would have to vary according to the type of repertoire and the results desired.

There are, of course some obvious limitations to these methods as we have used them, the most obvious one being the inability to deal with complex contexts and several different parameters at once, as is common in musicology. To what extent these limitations can be overcome, perhaps by combining these techniques with others, is an open research question.

REFERENCES

Dillon, M. and Hunter, M. (1982). Automatic Identification of Melodic Variants in Folk Music. *Computers and the Humanities*, 16, 107-117.

Dixon, W.J., and Brown, M.B. (Ed.) (1979). BMDP-79, Biomedical Computer Programs P-Series. University of California Press, Berkeley, Los Angeles.

Hauser, M. (1977). Formal structure in polar Eskimo drumsongs. *Ethnomusicology*, XXI, 33-53.

Hickey, P.J., Logrippo, L., Pelinski, R., and Strong, E.C. (1979). Internal and external representation of musical notation for an ethnomusicological project. *Proceedings of the Tenth Ontario Universities Computing Conference*, University of Ottawa, 215-228.

Krumhansl, C.L. (1985). Perceiving Tonal Structure in Music. *American Scientist*, 73, 371-378.

Logrippo, R., Pelinski, R., and Stepien, B. (1981). Typologies des contours mélodiques de chants personnels inuit réalisées à l'aide de l'ordinateur. In: Pelinski (1981) 103-155.

Morando, B. (1979). Construction et analyse de nuages associés à des parallelipèdes de correspondance: une contribution à l'analyse musicale. Thèse de 3ème cycle, Université Paris VI.

Pelinski, R. (1981a). A Typology of Melodic Contours Implemented by Computer. In: G.E. Lasker (ed.) *Applied Systems and Cybernetics*, Vol. V, 2750-2775.

Pelinski, R. (1981b). *La musique des inuit du caribou*. Presses du l'Université de Montréal, Montréal.

Sneath, P.H.A., and Sokal, R.R. (1973). *Numerical Taxonomy*. W.H. Freeman and Co., San Francisco.

Stech, D.A. (1981). A Computer-Assisted Approach to Micro-Analysis of Melodic Lines. *Computers and the Humanities*, 15, 211-221.

Steinbeck, W. (1976). The Use of Computers in the Analysis of German Folk-Songs. *Computers and the Humanities*, 10, 287-298.

Usher, S. and Najock, D. (1982). A Statistical Study of Authorship in the Corpus Lysiacum. *Computers and the Humanities*, 16, 85-105.

Wenker, J. (1977). A computer-aided analysis of Canadian folk-songs. In: Lusignan, S., and North, J.S. (Ed.) *Proceedings of the Third International Conference on Computing in the Humanities*, The University of Waterloo Press.

Cluster groups: 1, 2, 3', 3", 4

SONG #: 1 18 29 27 3 26 13 20 30 24 2 39 32 21 34 36 11 28 31 17 19 38 40 37 33 15 35 23 14 12 9 16 25 4

AMALG. DISTANCE:
1.085, 1.690, 1.939, 2.350, 2.634, 2.680, 3.508, 3.521, 3.588, 3.716, 3.785, 3.860, 3.930, 4.599, 4.794, 5.312, 5.545, 5.871, 6.153, 6.961, 7.118, 7.514, 7.985, 9.790, 10.601, 10.788, 11.592, 12.195, 12.993, 14.535, 20.166, 24.067, 25.263
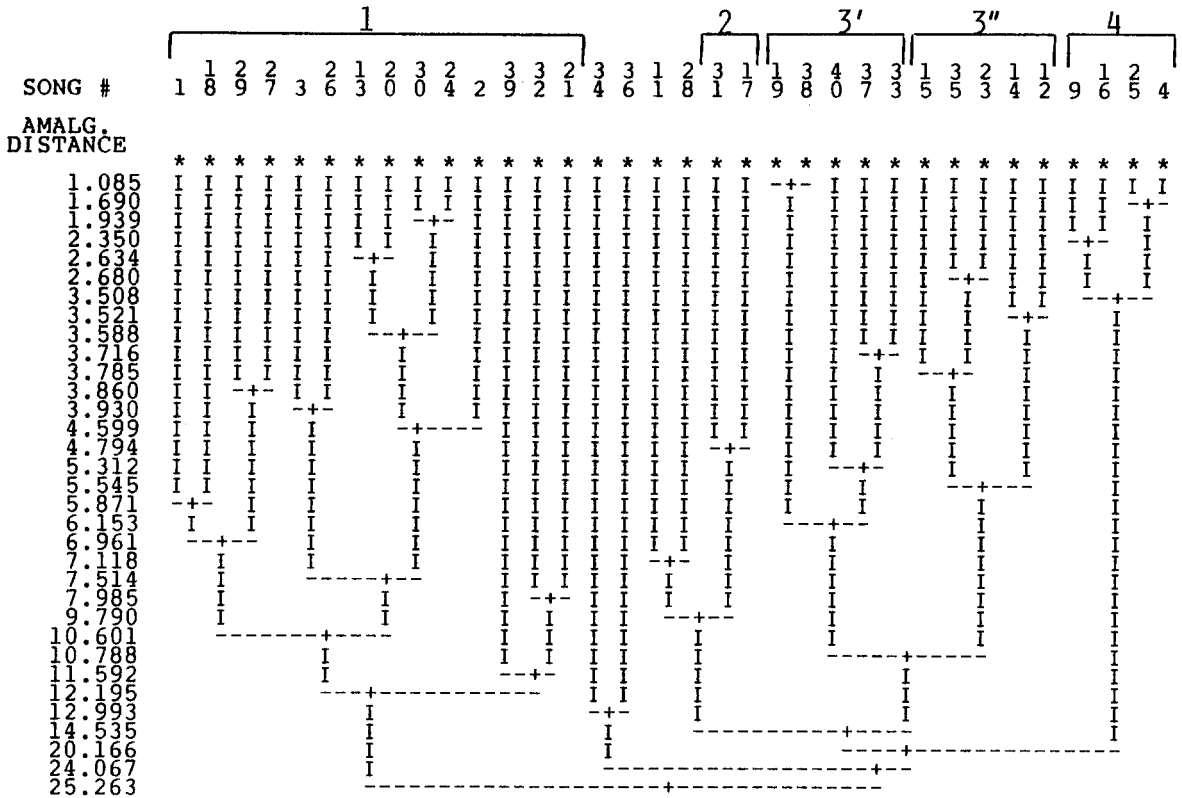
FIGURE 1. RANKIN INLET TONAL DENSITY AMALGAMATION TREE

PROPORTIONS OF NOTES USED FOR RANKIN INLET ON CLUSTERED SONGS

	SONG #	COMPOSER	-C	-C#	D	D#	E	F	F#	G	G#	A	A#	B	C	C#	+D	+D#
1	RK 1	KAVIK						10.5		73.3		1.3	12.3					
	RK 18	ULUJUK					3.4	25.8		59.2		11.6						
	RK 29	TIKTAK	3.5		1.6			13.9	4	66.0	1.4	1.6	2.8	.4	9.4			
	RK 27	TIKTAK						20.6		68.8		.6		4.1	5.9			
	RK 3	KAVIK			1.6		1.3	18.0		51.4		15.9		6.7	6.4		8.5	
	RK 26	KOLIT			2.8		.5	13.4		46.5		28.3			1.9		6.6	
	RK 13	TAUTUNNG						16.3		58.1		16.9			8.7			
	RK 20	MIRANAA					.7	14.4		60.1		10.6	3.7		9.6		1.6	
	RK 30	MANILAK						4.7		65.2		11.7			13.2		2.2	
	RK 24	USSAK			1.6			7.3		62.9		17.3			8.5		2.4	
	RK 2	KAVIK			2.4			17.7		42.1		21.9			15.6			
	RK 39	T.KIMALA	2.8		7.5		4.5	59.4		13.5		11.2			5.1			
	RK 32	ANGUTIQT						11.0		39.3		10.0	22.5		15.4		1.9	
	RK 21	NILAULA						4.0		21.0		49.9		1.7	16.4		3.4	
OUTL.	RK 34	KASIGIAK	8.9		3		65.2			19.2		6.4						
	RK 36	MCCASKIL						3.2		6.3					48.7		33.3	
	RK 11	PANIGONIA							12.2	55.9	2.7	2.7		18.5	8.1			
	RK 28	TIKTAK			23.8				7	63.5		2.1		9.9				
2	RK 31	ANGUTIQT			13.9					68.6			12.2		2.9		2.5	
	RK 17	ULUJUK			.5		5.4			62.9			17.6		7.3		5.1	
3'	RK 19	TAUTUNNG	4					7.5	2	68.5		5.5			15.3		2.6	
	RK 38	T.KIMAL					10.2	3.1		65.4		5.6			15.5		2.4	
	RK 40	T.KIMAL					15.3			46.4		25.6			5.4		1.7	
	RK 37	KIMALAR			4.4		8.2			49.1		9.5		4.2	13.4		12.3	
	RK 33	KASIGIAK	1.3		4.2		8.5			57.2		15.1		9	7.8		4.0	
3"	RK 15	ANARUAK			5.6		13.4			50.0		9.1		11.1			6.9	
	RK 35	TAPARTI			6.2		3.1			61.8		7.4		15.4			6.2	
	RK 23	USSAK			2.6		7.7			55.8		15.3		15.1			3.6	
	RK 14	TAUTUNNG			5.6	1.9	14.8			59.9		13.0		4.9				
	RK 12	TAUTUNNG			4.8		8.3		1.7	49.7		24.8		10.4				
4	RK 9	PANIGONIAK						37.5		53.9		7.9			.7			
	RK 16	ULUJUK					2.0	25.2		60.4		11.2			1.2			
	RK 25	USSAK						15.8		76.2		8.1						
	RK 4	IRKOOTI						21.9		69.1		7.5			1.5			

TABLE 1. RANKIN INLET TONAL DENSITY SORTED INDICATOR TABLE
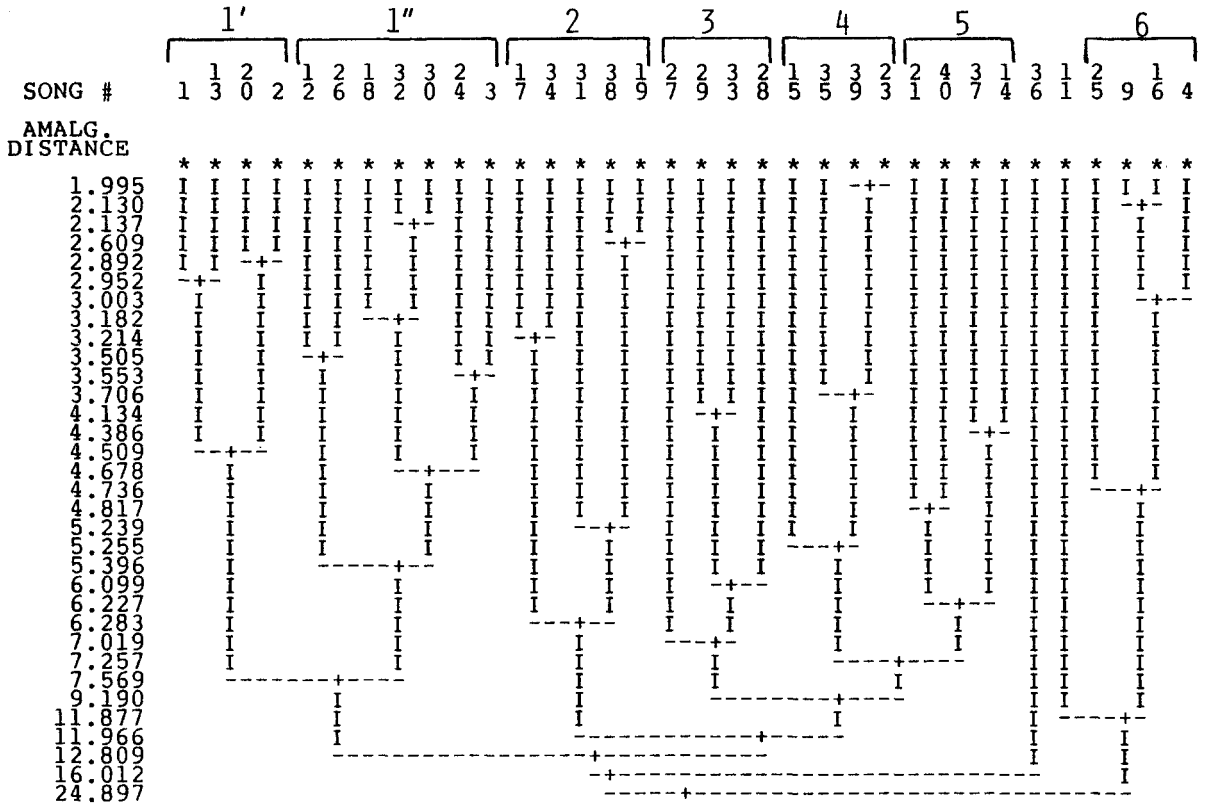
FIGURE 2. RANKIN INLET INTERVAL DENSITY AMALGAMATION TREE

CLASS			CHARACTERISTIC NOTES		INCLUDES SONGS
1			F,G,A,[C]		1,18,29,27,3,26,13,20,30,24,2,39,32,21
2			[D],G,A#,C,+D		31,17
3	3'	E,G,A	C,+D		19,38,40,37,33
	3"		B,D,[+D]		15,35,23,14,12
4			F#,G,A,[C]		9,16,25,4
OUTLIERS					34,36,11,28

TABLE 2. CLASSIFICATION OF SONGS BY TONAL DENSITY

MOST COMMON NOTES: D,G,A,C,+D

USES F — CLASS 1

A → A# — CLASS 2

USES E

USES F# NO D,D+ — CLASS 4

WEAK D — CLASS 3'

C → B — CLASS 3"
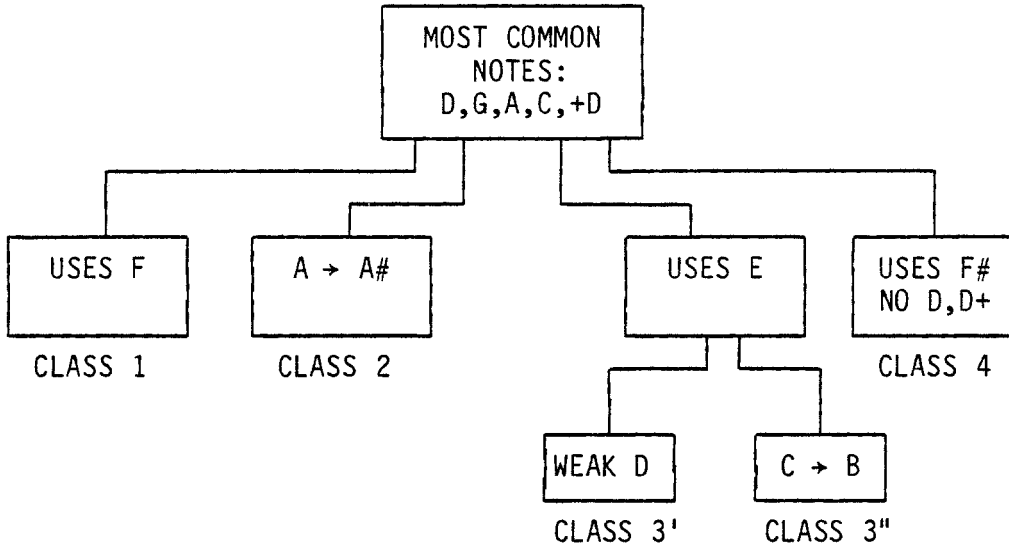
FIGURE 3

PROPORTIONS OF INTERVALS USED FOR RANKIN INLET ON CLUSTERED SONGS

	SONG #	COMPOSER	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7
1'	RK 1	KAVIK						5.6	13.5		61.8		14.6	3.4				
	RK 13	TAUTUNNG					4.9	4.9	7.4		59.0		18.9	4.9				
	RK 20	MIRANAA				1.4	4.1	4.1	12.9		54.4	7	13.6	4.8	4.1			
	RK 2	KAVIK				.7	6.7	8.1	11.4		43.0		17.4	6.0	3.4			
1"	RK 12	TAUTUNNG			1.6	2.4		8	14.6		47.2	8	26.8	3.3				
	RK 26	KOLIT				2.5	5.0	1.3	13.8		48.4		25.2	1.9	1.9			
	RK 18	ULUJUK					6.7		10.4		60.6		21.8		.5			
	RK 32	ANGUTIQT		1.1		3.8	4.9		7.7		57.9		19.1	3.8	1.6			
	RK 30	MANILAR		2.0			5.2		6.5		58.2		20.9	4.6	2.6			
	RK 24	USSAK		4.0		1.5			9.5		63.2		16.4	3.0		2.5		
	RK 3	KAVIK		1.8	4	2.7		2.2	5.8		58.3	1.8	16.6	4.5	3.1	9		
2	RK 17	ULUJUK	5					5.5	16.0	1.8	51.6		7.3	13.7	1.4	1.8		
	RK 34	KASIGIAK						6.6	10.5	6.1	53.6		6.1	10.5	6.6			
	RK 31	ANGUTIQT				3.8		1.9	5.7	6.1	73.3		3.8	7.6		3.8		
	RK 38	T.KIMAL	3.3			.7		7.9	3.9		56.6		9.9	17.8				
	RK 19	TAUTUNNG	1.8			6		5.3	7.1		57.6		10.6	12.9	6	2.9		
3	RK 27	TIKTAK					4.2		16.8	5.0	56.3		12.6			5.0		
	RK 29	TIKTAK		.5 2.7		4.5	5	3.6	10.8	9	52.3	2.3	8.1	1.4	5	4.1		3.2
	RK 33	KASIGIAK	2.0	2.0		5.4		4.1	7.4		51.4	1.4	11.5	4.7		10.1		
	RK 28	TIKTAK		.7		11.9	1.3		2.6		63.6		3.3		4.0	11.3		
4	RK 15	ANARUAK		1.4		4.2			14.1		38.0		11.3	4.2	7.0	12.7		
	RK 35	TAPARTI		2.5			8	2.5	9.3		55.1		10.2	5.9	5.9	3.4		
	RK 39	T.KIMALA	7	2.1		2.8		2.1	9.1		46.2		16.8	9.8	1.4	4.9		
	RK 23	USSAK		2.1		2.1		2.1	9.6		54.1		13.0	6.2	2.7	5.5		
5	RK 21	NILAULA				1.6		3.2	14.4	2.4	45.6	3.2	17.6	8.8	8			
	RK 40	T.KIMAL	3.9	1.1		8.3		.6	9.4		43.9	6	11.7	13.3	5.6	6		
	RK 37	KIMALAR				4.4	2.2	4.4	5.8		46.7		14.6	10.2	2.2	5.1		
	RK 14	TAUTUNNG		3.0	3.0	7.8		8.0	6.0		43.0	3.0	15.0	11.0		4.0		
ATL.	RK 36	MCCASKIL				7.8			32.5		16.9		32.5		2.6			
	RK 11	PANIGONIA				1.9	11.5	3.8		13.5	36.5	17.3		15.4				
6	RK 25	USSAK							3.4	9.7	73.7	9.7	3.4					
	RK 9	PANIGONIAK						6.9	5.2	10.3	44.8	17.2	5.2	8.6	1.7			
	RK 16	ULUJUK						9.7	4.1	9.7	49.8	12.4	8.3	6.0				
	RK 4	IRKOOTI						7.4	2.9	6.3	64.6	8.6	5.1	5.1				

TABLE 3. RANKIN INLET INTERVAL DENSITY SORTED INDICATOR TABLE

CLASS		CHARACTERISTIC INTERVALS		INCLUDES SONGS
1	1'	[-4],-2,0,2,3	-3	1,13,20,2
	1"			12,26,18,32,30,24,3
2		-3,[-2],0,2,3		17,34,31,38,19
3		[-7],[-5],-2,0,2,5		27,29,33,28
4		-7,[-5],[-3],-2,0,2,3,4,5		15,35,39,23
5		-5,-3,-2,0,[1],2,3,[4],[5]		21,40,37,14
6		[-3],-2,-1,0,1,2,[3]		25,9,16,4
OUTLIERS				36,11

TABLE 4. CLASSIFICATION OF SONGS BY INTERVAL DENSITY
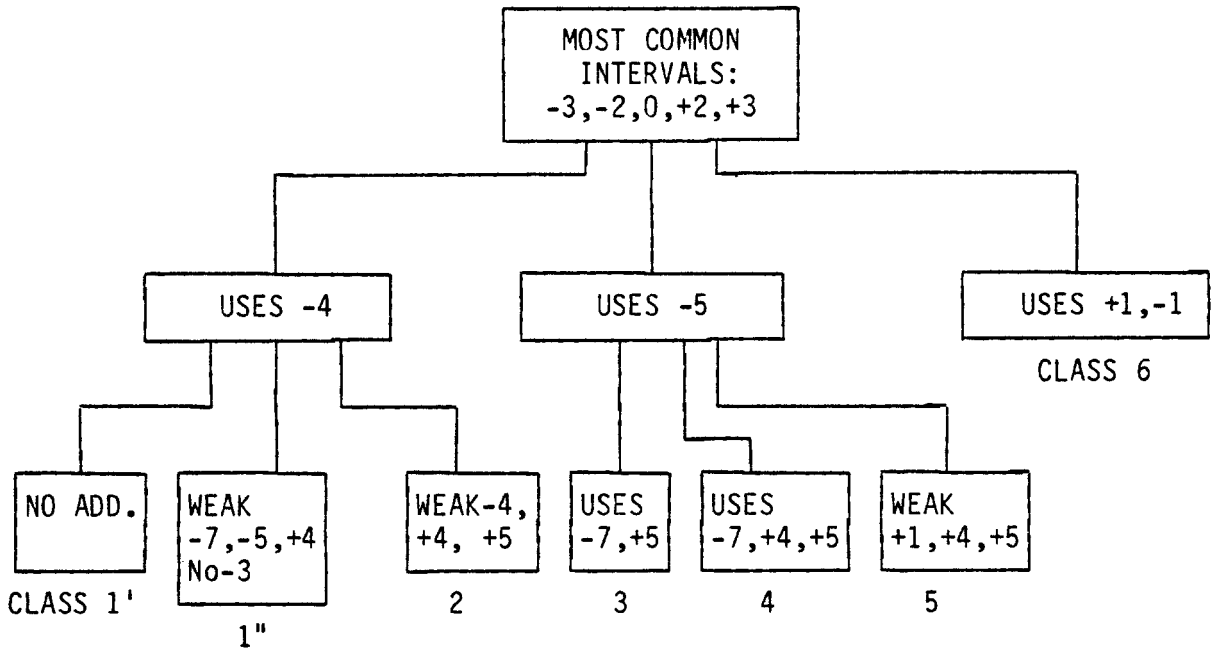
FIGURE 4

MODE	Tonal Class	Interval Class	SONGS
1	1	1'	1, 13, 20, 2
2	1˙	1"	26, 18, 32, 30, 24, 3
3	1	3	27, 29
4	2	2	31, 17
5	3'	2	38, 19
6	3	5	40, 37, 14
7	3"	4	15, 35, 23
8	4	6	9, 16, 25, 4
OUT-LIERS			21, 39, 33, 12, 34, 36, 11, 28

TABLE 5. SONG CLASSIFICATION BY "MODES"

```
I---------------------------------   8  0  G   A   G   F   E
I
I                             I------  24  0  C  -B   C   D  -A  -B   C
I        I----------------I
I        I                I  I----  21  1  D   E   D   C  -B  -A  -G
I        I                I-I
I        I                  I  I--  23  1              C  -B  -A  -G
I        I                  I-I
I        I                    I--  16  1              C  -B  -A  -G
I        I
I        I  I--------------------  20  0  -G  -A  -B   C   E   C
I        I  I
I  I-I  I                     I----   9  2              F   G   F   E       D
I  I  I  I                 I-I
I  I  I  I                 I  I I--  22  2      D   E   F   G   F   E
I  I  I  I                 I I-I
I  I  I  I             I-I   I--   2  2      D   E   F   G   F   E       D
I  I  I  I             I  I
I  I  I  I             I  I  I--  13  2  E   D   E           F   E
I  I  I  I           I-I  I---I
I  I  I  I             I  I     I--   7  2  E   D   E       G   F   E   C
I  I  I  I      I-----I  I
I  I  I-I  I           I  I I--------   6  3  G   F   E   D       G
I  I      I      I       I
I  I      I      I       I----------   1  3      F   E   D   C   E   G   D
I  I      I  I-I
I  I      I  I  I                 I--  18  4  C   D       F   E   D   C
I  I      I  I  I  I-----------I
I  I      I  I  I  I             I--  11  4  C   D   E   F   E   D   C
I-I      I  I  I-I
   I      I  I     I  I-----------  19  5              C   D   C   D       G   C
   I      I  I     I I
   I      I  I     I-I  I----------   5  5  E   D   C   D   C   D   C   G
   I      I  I       I  I
   I      I-I       I-I  I--------  14  6              C               E   C  -G
   I        I         I  I
   I        I         I  I     I--  17  6  C   D   C   D           E   C
   I        I       I-I    I-I
   I        I         I  I-I  I--  15  6      D   C   D           E   C
   I        I           I  I I
   I        I         I-I  I----   3  6      D   C   D           E   C  -G
   I        I           I
   I        I           I------   4  6  E   D   C   D   C   D   E   C
   I        I
   I        I-----------------  12  6  -G          C   D           E
   I
   I-----------------------------  10  0  E   F   E   D   C   D  -G
```

FIG. 5. EXAMPLE OF AMALGAMATION TREE OF MODULES

DETAILED EXAMINATION OF CLUSTER # 21

ID												
184				G				E				C
130	D	E	F	G			F	E	D	E		C
93		E	F	G			F	E	D			C
18	G	E	F	G			F	E	D			C
122	D	E	F	G			F	E				
2	D	E	F	G			F	E	D			
85		E	F	G	E	A	F	E	D		G	C
83		E	F	G	E	A	F		D			
63		E	F	G	E	A						

DETAILED EXAMINATION OF CLUSTER # 24

ID																	
185						E										C	-G
109		C				E										C	-G
191						E	C	F	E	D	C	D	E			C	
143						E				D	C	D	E			C	
163										D	C	D	E			C	
110										D	C	D	E			C	
3										D	C	D	E			C	-G
175		C		D	C					D			E			C	
112		C		D	C					D			E			C	
4	E	C		D	C					D	C	D	E			C	
73						E				D			E			C	
9		C	E	D		E										C	
99		C		D		E				D	C	D	E			C	-G
88		C		D		E					C	D	E	F	E		
38		C		D		E					C	D	E				

FIG. 6. TWO EXAMPLES OF MODULE ALIGNMENT

```
  8  0
 24  0
 21  1
 23  1   =*
 16  1   =**
 20  0   .
  9  2   .
 22  2   .         |=
  2  2   .         |**
 13  2             | ==
  7  2   .         |.$=*
  6  3   .         |=.= .
  1  3             |. . . . =
 18  4   . -      .|. .-.-.-
 11  4   .        .|.-===..|*
 19  5   ..        .   .|.--
  5  5   .            ..|-.$
 14  6   .=..=     .  .|==..
 17  6   ..    .   .  .--$$=
 15  6   ..    .   .  .|=====*
  3  6   .-...     .  .----*$*
  4  6   .    .   ..  .-.$$=**$
 12  6   ..    =  ....  .|==...====
 10  0    ...  ......=|-=..=..-..
```
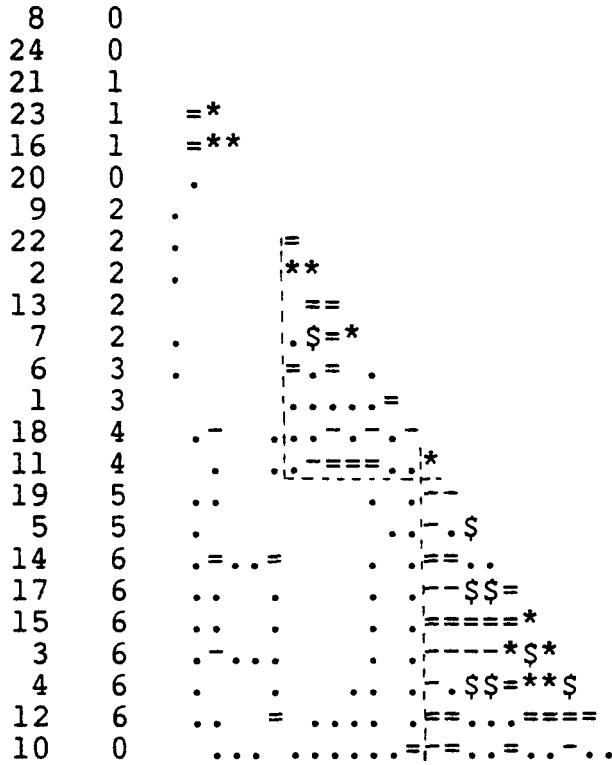
THE ABOVE SHADES REPRESENT THE FOLLOWING VALUES

```
0.10=
0.20=
0.30=
0.40=
0.50=  .
0.60=  -
0.70=  =
0.80=  $
0.90=  #
1.00=  *
```

FIG. 7.   SORTED AND SHADED DISTANCE MATRIX