Applications of Multidimensional Scaling in Psychometrics

Yoshio Takane

Department of Psychology
McGill University
1205 Dr. Penfield Avenue
Montréal Québec, Canada
email: takane@takane2.psych.mcgill.ca

1 Introduction

Multidimensional scaling (MDS) is a set of data analysis techniques for analysis of proximity data. Both narrow and broad definitions of MDS exist. Because of space limitation, we take a narrow view of MDS in this paper. According to this view, MDS is a collection of techniques that represent proximity data by spatial distance models. The space is usually Euclidean, although other forms of distance functions (e.g., the city-block distance model) have also been considered in the literature. Again due to space limitation, we focus on the Euclidean case in this paper. See Arabie (1991), and Hubert, Arabie, and Hesson-Mcinnis (1992) for applications of the city-block distance model.

MDS represents a set of stimuli as points in a multidimensional space in such a way that those points corresponding to similar stimuli are located close together, while those corresponding to dissimilar stimuli are located far apart. To illustrate, many road maps have a matrix of inter-city distances. Put simply, MDS recovers a map based on the set of inter-city distances. Given a map it is relatively straightforward to measure the Euclidean distance between the cities. However, the reverse operation, that of recovering a map from inter-city distances (or finding relative locations of the cities so that their mutual distances best agree with a given set of distances) is not as easy. MDS, roughly speaking, is a method to perform this reverse operation.

To illustrate further, let us look at Table 1. This is a matrix of airplane distances between 10 US cities (Kruskal and Wish, 1978). Given a matrix of this sort it would be difficult to find geographic locations of the 10 cities. The story may be totally different, if the cities are identified as in Table 1. Most of us know that the 10 cities should be located as in Figure 1. This is because we already have a fairly good internalized map of North America. It would still be very difficult, if not impossible, for those who are unfamiliar with the geography of North America to construct a map like this based exclusively on the set of inter-city distances. The role of MDS is to construct a map like the one in Figure 1 from a table like Table 1 for those who do not know the "geography" in certain areas. Figure 1 was in fact derived from Table 1 by MDS. In the remaining portions of this paper I would like to elaborate this role of MDS. This will primarily be done by showing various examples of application of MDS.

*** Insert Table 1 and Figure 1 about here ***

This paper consists of three major sections. In the next section (Section 2) we give examples of MDS by the simple Euclidean distance model. In Section 3 we discuss applications of MDS by the weighted Euclidean distance model to capture certain individual differences in proximity judgments. In Section 4 we illustrate applications of unfolding analysis, a special kind of MDS to represent individual differences in preference judgments. Throughout this paper, we keep the amount of technical information minimal. Readers who are interested in a more comprehensive account of MDS are referred to the textbooks listed in the conclusion section of this paper.

2 MDS with the Simple Euclidean Model

In this section, we present some results from MDS with the simple Euclidean model. Simple MDS is typically applied to a single (dis)similarity matrix. We first present a series of examples of application, followed by a brief account of simple MDS. Most of the results in this section were derived by least squares (LS) MDS. In the final subsection, however, we discuss maximum likelihood (ML) MDS, and illustrate what it can do in the analysis of proximity data.

2.1 Examples of application of simple MDS

Example 1: Confusion among Morse Code signals. Although the example in the previous section involved stimuli (cities) which had an intrinsic geometrical structure, this is not the case in most applications of MDS. Let us first look at Figure 2. This is called confusion matrix (Rothkopf, 1957). Stimuli are 36 Morse Code signals. Two signals were presented in each trial, and subjects were asked to judge whether the two stimuli presented were the "same" or "different". (This is called same-different judgment.) The relative frequencies with which row stimuli were judged the "same" as column stimuli were entered in this table. For example, when the signal indicating letter A was followed by the same signal, 92% of the subjects correctly responded that they were the same. When it was followed by the signal indicating letter K, 22% incorrectly responded that they were identical, and so on. This is an overwhelming array of numbers (36 by 36), and it is almost impossible to find any structure underlying the confusion process by merely inspecting a table like this. MDS is helpful in finding some meaningful structure in the data.

*** Insert Figures 2 and 3 about here ***

Figure 3 presents the result of MDS of these data (Shepard, 1963). MDS locates the stimulus points in such a way that more confusable (or more similar) stimuli are close together, while less confusable stimuli are far apart. By looking at the figure, it is readily apparent that:

1. A process mediating the confusions between the signals is roughly two dimensional.
2. One is the total number of components in a signal (i.e., the number of dots plus the number of dashes), and the other is the ratio of the number of dots to the number of dashes in a signal (that is, which type of components is more predominant).

Signals having more components tend to be located toward the top of the figure. Signals with the same number of components are more confusable than those having different numbers of components. Within the same number of components, symbols on the right have more dashes and fewer dots than those on the left. Signals with the same number of dots and the same number of dashes are relatively more confusable, particularly when they are mirror images of one another. Two signals are said to be mirror images when they have exactly the same number of components of each kind, and the components are arranged in exactly reverse orders. In the figure, stimuli that are mirror images of each other are connected by line segments. The configuration is more crowded toward the top. This is because a more variety of signals can be formed with more components. A more variety of signals in turn lead to more confusions.

The role of MDS in this example is rather clear. MDS extracts useful information from an overwhelming array of proximity data in a way easily assimilable to human eyes. This example is perhaps the single most famous example in MDS, and readers may have seen it elsewhere. Buja and his collaborators (Buja and Swayne, 2002; Buja, Swayne, Littman, Dean, and Hoffman, 2001) present further analyses of this data set by XGvis, a very flexible computer program for simple MDS with interactive visualization features.

Example 2: Japanese Kana characters. Figure 4 displays an MDS configuration of 46 Japanese Kana characters (phonetic symbols). Twenty students at the University of Hawaii who did not know Japanese at all were asked to classify these symbols into as many groups as they liked in terms of their similarities in shape (Dunn-Rankin and Leton, 1975). The frequencies with which two symbols were classified into the same groups were used as similarity measures between the symbols for MDS analysis.

*** Insert Figure 4 about here ***

Unlike the previous example, this configuration does not seem to permit a straightforward dimensional interpretation, though perhaps the horizontal direction roughly represents the complexity in shape. More complicated symbols tend to be located toward the left hand side of the configuration. The vertical axis is difficult to interpret, however. A cluster analysis, which groups similar stimuli together, was applied to the same set of data, and the groupings of the stimuli obtained by the cluster analysis were superposed on the stimulus configuration derived by MDS. Circular contours enclosing subsets of stimulus points indicate the groupings. Apparently another kind of interpretation (called configural interpretation) is possible of the derived stimulus configuration. (In a dimensional interpretation we seek for meaningful directions in the space, whereas in a configural or neighborhood interpretation we look for meaningful regions in the space). Our interpretation of the six clusters obtained by the cluster analysis are as follows: 1. an angular form, 2. a curved feature, 3. discrete components, 4. a zigzag feature, 5. a round feature, and 6. crossed features. Of the six clusters, the sixth one is the biggest that includes three sub-clusters labelled (6), (7) and (8). Note that at the left end of this cluster, characters having a "double cross" are located. The organizing principle underlying perceived similarities between these symbols seems to be the "distinctive features" that subsets of symbols share in common, and that symbols in other clusters do not.

Example 3: Have words. In what is called stimulus sorting method, subjects are asked to sort a set of stimuli according to the similarity among them. Typically, a group of subjects are given a deck of cards on which stimuli are printed, and are asked to sort the cards into as many piles as they want, so that the stimuli within the same piles are more similar to each other than those grouped into distinct piles. Because of its simplicity, it is a very popular data collection method for similarity data among social scientists. The previous example (Example 2) used the sorting method. We give another example.

Stimuli were 29 Have words used originally by Rapoport and Fillenbaum (1972): 1. accept, 2. beg, 3. belong, 4. borrow, 5. bring, 6. buy, 7. earn, 8. find, 9. gain, 10. get, 11. get rid of, 12. give, 13. have, 14. hold, 15. keep, 16. lack, 17. lend, 18. lose, 19. need, 20. offer, 21. own, 22. receive, 23. return, 24. save, 25. sell, 26. steal, 27. take, 28. use, and 29. want. Ten university students were asked to classify them in terms of their similarity in meaning. A somewhat specialized MDS method (Takane, 1980) was applied to these data to derive the stimulus configuration displayed in Figure 5. (This specialized MDS is analytically similar to dual scaling (Nishisato, 1980) and multiple correspondence analysis (Greenacre, 1984). See Takane (1980) for more details.)

*** Insert Figure 5 about here ***

This configuration permits a straightforward dimensional interpretation. The horizontal direction contrasts two possible future states, either possession or nonpossession, following the current state of possession. Words such as 13. have, 21. own, and 3. belong are located on the left side, while 12. give, 25. sell, 18. lose, etc. are placed on the opposite side. Thus, the horizontal direction contrasts a stable state of possession (left) with the state of losing possession (right). Similarly, the vertical direction distinguishes two states of current nonpossession, a stable state of nonpossession at the top and an unstable state of nonpossession at the bottom. It seems that the vertical direction represents subtle gradients of sureness of change in the current state. Words such as 16. lack, 19. need, and 29. want are located at the top, which indicate little prospect of change, while 22. receive, 10. get, 8. find, etc. are located at the bottom, which indicate that the change is most probable. Interestingly, 2. beg is located at about the middle, which indicates that some action has been taken to change the state, but the prospect of change is uncertain. Four circular contours in the figure show the four sorting clusters elicited by one of the subjects. The stars indicate the centroids of the four clusters. This subject is most likely to have sorted the stimuli into the four groups by following the dimensional structure described above.

Example 4: Color space for the pigeon. MDS is not restricted to human subjects. The next example shows an instance of MDS with infra-human subjects (Schneider, 1972). Six pigeons were trained to discriminate between colors varying primarily in hue. A pair of colors were presented as two halves of a circle in each trial, and the pigeons were trained to peck the left lever when the two colors were identical, and peck the right lever when they are different. This is a kind of same-different judgment similar to the one used in the Morse Code confusion data. The correct discrimination probabilities between distinct colors were used as dissimilarity measures for MDS analysis. Two experiments were conducted with varying numbers of stimuli (12 in Experiment I, and 15 in Experiment II).

*** Insert Figures 6 and 7 about here ***

Figure 6 displays the derived stimulus configuration. In this figure, the results of the two experiments were superposed; the stimuli used in Experiment I are indicated by squares, and those in Experiment II by circular dots. Colors are labelled by their wave lengths. A familiar color wheel is observed in pigeon's color space. Starting from purple (associated with the shortest wave length) at the left bottom corner, the configuration goes clockwise through blue, green, and yellow to red (associated with the longest wave length) in a circular manner. For comparison, a color wheel typically found with human subjects is presented in Figure 7 (Ekman, 1954; Shepard, 1964). The two configurations are strikingly similar to each other. This means that pigeons have a color perceiving system similar to that of human subjects.

2.2 A brief account of simple MDS

Before proceeding further, a brief account of simple MDS is in order. In simple MDS, we are given a matrix of observed (dis)similarities data on a set of n stimuli. The data may be put in the form of an n by n matrix $\boldsymbol{O}$ whose ij^{th} entry (the element in the i^{th} row and the j^{th} column), o_{ij}, designates the (dis)similarity between stimuli i and j. By applying MDS to the data, we obtain an n by A matrix $\boldsymbol{X}$ of stimulus coordinates, where A is the dimensionality of the space in which the stimuli are represented as points. The ia^{th} element of $\boldsymbol{X}$, x_{ia}, indicates the coordinate of stimulus i on dimension a. The matrix of stimulus coordinates, $\boldsymbol{X}$, is further converted into a graphic representation (like the ones we saw in the previous subsection) by introducing a Cartesian coordinate system. Since the Euclidean distance is invariant over the shift of origin and the rotation of axes, we may remove the coordinate axes, once stimulus points are located.)

Inter-point distances are calculated from $\boldsymbol{X}$. In the simple Euclidean model, the distance between stimuli i and j is defined by

$$d_{ij} = \{\sum_{a=1}^{A} (x_{ia} - x_{ja})^2\}^{1/2}. \tag{1}$$

MDS determines the locations of the stimulus points in such a way that the set of d_{ij} calculated from their coordinates best agree with the set of observed (dis)similarities between the stimuli.

When the observed dissimilarity data are measured on a ratio scale, there exists a method to obtain closed form solutions (Torgerson, 1952). This method first transforms the data into scalar (inner) products (Young and Householder, 1938), which are then subjected to the spectral (eigenvalue and vector) decomposition. However, dissimilarity data collected in psychology are rarely measured on a ratio scale, but only on an ordinal scale. This means that the observed dissimilarity data and inter-point distances are only monotonically related (or in the case of similarity data, inversely monotonically related) with each other. Nonmetric MDS, initiated by Shepard (1962) and Kruskal (1964a, b), monotonically transforms the observed (dis)similarity data (to make them more in line with the underlying distances), and at the same time it finds the best spatial representation of the stimuli. Let g denote the monotonic transformation of the data. (The transformed data are often called "disparities.") Then, nonmetric MDS finds both g and a set of stimulus coordinates x_{ij} in such a way that

$$\phi(g, \boldsymbol{X}) = \sum_{i,j>i} (g(o_{ij}) - d_{ij})^2 \tag{2}$$

is minimized under a suitable normalization restriction on distances. All the examples given so far were analyzed by nonmetric MDS, except the one involving airplane distances, in which actual distances were used. When nonmetric MDS was applied to ranked airplane distances, we obtained a configuration virtually indistinguishable from the one derived from the actual distances. (The correlation between the two configurations was .999 with the largest displacement of a city of about 70 miles.) This indicates that the the ordinal information is often sufficient to derive stable stimulus configurations.

Figure 8 presents the plot of observed confusion probabilities in Morse Code signal confusion data (Example 1) against fitted distances. (This is called Shepard diagram.) The connected line segments in the figure shows the best (inverse) monotonic transformation of the data obtained by a nonmetric MDS procedure. The transformation looks like a negative exponential function of distances. Of course, we had no *a priori* knowledge of the form of this function before MDS was applied. Nonmetric MDS can be carried out without knowledge of the exact form of the function relating observed (dis)similarity data to underlying distances. The transformation was derived solely on the basis of the (inverse) monotonicity relationship assumed between the data and the distances.

*** Insert Figure 8 about here ***

A number of computer programs exist for simple MDS. KYST (Kruskal, Young, and Seery, 1978) is still a good and reliable program for nonmetric MDS. Other available programs are described in Schiffman, Reynolds, and Young (1981). See also other textbooks on MDS listed at the end of this article. PROXSCAL (Busing, Commandeur, and Heiser, 1997) employs a majorization algorithm for MDS (de Leeuw, 1977, 1988; de Leeuw and Hesier, 1977; Heiser, 1991). PROXSCAL is being incorporated in SPSS.

2.3 Maximum likelihood MDS

. The next and final example in this section uses maximum likelihood (ML) MDS. This type of MDS allows certain kinds of hypothesis testing as well as assessment of reliability of derived stimulus configurations under some distributional assumptions on observed (dis)similarity data. We briefly discuss ML MDS in preparation for the next example. The exposition is rather terse. For more detailed accounts of ML MDS, see Ramsay (1977, 1978, 1982) and Takane (1978, 1981, 1994; Takane and Carroll, 1981; Takane and Sergent, 1983; Takane and Shibayama, 1986, 1992).

There are three essential ingredients in ML MDS; the representation model, the error model, and the response model. The representation model is the model that represents proximity relations between stimuli. In MDS, this is typically a distance model like (1). Distances are then assumed to be error-perturbed. The error model describes the nature of this error-perturbation process. The error-perturbed distances are further transformed into specific forms of (dis)similarity judgments. The response model prescribes the mechanism of this transformation process. Because a variety of data collection methods exist that elicit a variety of proximity judgments, the response model must be specified for each specific type of data collection method.

Suppose, for simplicity, that we use (1) as the representation model, and that the distance is error-perturbed by

$$\delta_{ijr} = d_{ij} e_{ijr}, \tag{3}$$

where $\ln e_{ijr} \sim \mathrm{N}(0, \sigma^2)$, and r is the index of replication. Suppose further that proximity judgments are obtained on rating scales with a relatively small number of observation categories. We may assume that the dissimilarity between stimuli i and j is rated in category m whenever δ_{ijr} falls between the lower and the upper thresholds of this category, denoted by b_m and b_{m+1}, respectively. Then

$$p_{ijm} \equiv \Pr(b_m < \delta_{ijr} < b_{m+1}). \tag{4}$$

Under the distributional assumption made on δ_{ijr}, this probability can be explicitly evaluated as a function of model parameters ($\{x_{ia}\}$ in the representation model, σ^2 in the error model, and $\{b_m\}$ in the response model). Let f_{ijm} denote the observed frequency with which the dissimilarity between stimuli i and j is rated in category m. Then, the likelihood of the entire set of observed frequencies can be stated as

$$L \propto \prod_{i,j<i,m} p_{ijm}^{f_{ijm}}. \tag{5}$$

Parameters in the model are estimated so as to maximize the likelihood. For response models for other types of proximity judgments, see the papers cited above. We also see another example of response models in the final section of this paper.

The likelihood ratio (LR) statistic can be used to test the significance of the difference between two nested models, one hypothesized under the null hypothesis (H_0) and the other under the alternative hypothesis (H_1). The LR statistic, λ, is defined as

$$\lambda = \frac{L_0^*}{L_1^*}, \tag{6}$$

where L_0^* and L_1^* are the maximum likelihoods obtained under H_0 and H_1, respectively. Under some regularity conditions, $-2\ln\lambda$ asymptotically follows the chi-square distribution with degrees of freedom equal to the difference in the number of parameters between the two models, when the null hypothesis is true. Caution should be exercised, however, when the LR test is used for dimensionality selection. One of the regularity conditions is violated in the comparison between the A dimensional true model and the $A+1$ dimensional model (Takane, van der Heijden, and Browne, 2003), and no direct comparison can be made between them. Solutions of specific dimensionality should always be compared against the saturated model. In the categorical rating data described above, the saturated model takes $f_{ijm}/\sum_m f_{ijm}$ as the maximum likelihood estimate of p_{ijm} in (4). The whole testing procedure proceeds as follows: First, the one-dimensional model is compared against the saturated model, and if the difference is significant, the dimensionality is at least two. Then the two-dimensional model is compared against the saturated model, and so on until a solution with no significant difference from the saturated model is found. The first dimensionality with no significant difference from the saturated model is considered the best dimensionality.

Alternatively, AIC may be used to compare goodness of fit (GOF) of various models. The AIC is defined by

$$\mathrm{AIC}_\pi = -2\ln L_\pi^* + n_\pi, \tag{7}$$

where L_π^* is the maximum likelihood of model π, and n_π is the effective number of parameters (Takane, 1994). The model associated with the smallest value of AIC is considered the best fitting model (the minimum AIC criterion). AIC may be used to compare any number of models simultaneously. To use AIC for dimensionality selection, the saturated model should always be included as one of the models in the comparison. In the example we discuss shortly, we mostly use AIC for model selection.

One potential problem with the LR tests or AIC is that it is based on the large sample behavior of the ML estimators. At the moment there is no practical guideline as to what sample size is big enough to rely on the asymptotic properties of the ML estimators. Some small sample behavior of the LR statistic has been investigated by Ramsay (1980).

The Moore-Penrose inverse of the information matrix evaluated at the maximum likelihood estimates of model parameters provides asymptotic variance-covariance estimates of the estimated parameters (Ramsay, 1978). The information matrix, $\boldsymbol{I}(\boldsymbol{\theta})$, is defined as

$$\boldsymbol{I}(\boldsymbol{\theta}) = \mathrm{E}\left[\left(\frac{\partial \ln L}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \ln L}{\partial \boldsymbol{\theta}}\right)'\right], \tag{8}$$

where $\boldsymbol{\theta}$ is the parameter vector. Under the asymptotic normality of the maximum likelihood estimates, confidence regions may be drawn that indicate how reliably the point locations are estimated. We will see examples of confidence regions in the final section of this paper.

Example 5: Similarity of rectangles. Takane (1981) investigated the psychological dimensions of similarity between rectangles using ML MDS. The rectangles were constructed by factorial combinations of several levels each of the physical area and shape dimensions (Krantz and Tversky, 1975). A single subject made dissimilarity judgments on all pairs of rectangles on seven-point rating scales. This was replicated six times, and frequencies of category responses were cumulated and used as input to the ML MDS procedure.

Table 2 gives a summary of GOF statistics for fitted models. The unconstrained two- and three-dimensional solutions fit about equally well according to the minimum AIC criterion, both of which in turn fit substantially better than the saturated model. For ease of presentation, we display the two-dimensional configuration in Figure 9. (The three-dimensional solution is obtained by slightly bending the two-dimensional configuration along the third dimension.) In this figure, the horizontal direction roughly corresponds with the area dimension, while the vertical direction with the shape dimension. However, some qualifications are needed.

*** Insert Table 2 and Figure 9 about here ***

In the figure, rectangles having the same area are connected by line segments. If the horizontal direction indeed corresponds with the area dimension, the connected line segments should all be perpendicular to the horizontal direction. Essentially the same should hold for the vertical direction. We may formally test the validity of the area and shape hypothesis by comparing the GOF of this hypothesis against that of the unconstrained solution. Figure 10 presents the stimulus configuration derived under the area and shape hypothesis. In this figure, rectangles of the same area take equal coordinate values on the horizontal axis, and those with the same shape take equal coordinate values on the vertical axis. Table 2 indicates, however, that the GOF of this configuration is much worse than the unconstrained two-dimensional solution.

*** Insert Figure 10 about here ***

Since the area of rectangles is defined as height times width, and the shape is defined as height divided by width, the same set of rectangles can also be characterized by the combination of the (log) height and width dimensions. Again we may formally test how good this characterization is as subjective dimensions of rectangles. Figure 11 displays the stimulus configuration derived under the height and width hypothesis. Table 2 indicates that this hypothesis fits to the data much better than the area and shape hypothesis. It captures one important aspect of the unconstrained solution; the shape difference is not so important when the rectangles are small, but it becomes more important as the rectangles get larger. However, the GOF of the height and width hypothesis is still substantially worse than that of the unconstrained two-dimensional solution.

*** Insert Figure 11 about here ***

Schönemann (1977) formalized the hypothesis that the perceived shape difference increases as the area of the rectangles increases. This hypothesis is depicted in Figure 12. This hypothesis was fitted and the GOF of this hypothesis is reported in Table 2. It can be observed that this hypothesis still does not account for the data as well as the unconstrained two-dimensional solution. In Schönemann's hypothesis, area levels were still assumed to be independent of the shape dimension, whereas in the unconstrained two-dimensional configuration, the area dimension is curved along the shape dimension. Curves connecting the rectangles with the same area levels look like arcs drawn from a common focal point, where both the area and the shape of rectangles vanish. Rectangles with the same area levels are equally distant from this focal point. An hypothesis implementing this idea is depicted in Figure 13. Unfortunately, no ML MDS is yet capable of fitting this hypothesis.

*** Insert Figures 12 and 13 about here ***

3 Individual Differences MDS

In this section we discuss individual differences (ID) MDS. We primarily focus on a particular kind of ID MDS, called INDSCAL (Carroll and Chang, 1970a), to capture a specific kind of individual differences in (dis)similarity judgments, namely differential weighting of dimensions by different subjects. Again we discuss some examples first, followed by a brief account of the model used to represent the individual differences.

3.1 Examples of individual differences MDS

Example 6: Helm's color data. We have so far been focussing on MDS which derives a single stimulus configuration from a single matrix of observed proximity data. Suppose we have N (≥ 1) such matrices, each obtained from a different subject. If no systematic individual differences exist, we may analyze them simultaneously and derive a single common stimulus configuration. Alternatively, when some systematic individual differences are suspected, we may apply MDS separately to each single matrix. In this case we obtain N separate stimulus configurations. The question is whether there is a better way to represent differences among matrices of (dis)similarities than applying MDS separately to these data. The answer is "yes", and the technique is called individual differences MDS (Carroll and Chang, 1970a).

Figure 14 helps understand the kind of analysis that can be carried out by the individual differences (ID) MDS. Dissimilarity judgments among 10 colors of varying hue were obtained from 12 subjects (Helm, 1964). Two of the 12 subjects replicated the same experiment twice. The whole data set is given in Borg and Groenen (1997, p. 360). The stimulus configuration displayed in Figure 14 was obtained by applying ID MDS to these data. In this figure, stimuli are labelled as: RP (red-purple), RO (red-orange), Y (yellow), GY1 (green-yellow 1), GY2 (green-yellow 2), Gr (green), Blu (blue), PB (purple-blue), P2 (purple 2), and P1 (purple 1). We can see the familiar color wheel with: 1. Vertical axis roughly contrasting red and green (R-G dimension), and 2. Horizontal axis roughly contrasting yellow and blue (Y-B dimension). Figure 15 represents the weights attached to these two dimensions by different subjects. Subjects N6, for example, put about equal emphasis on both dimensions, while subjects D3 put excessively heavy emphasis on the horizontal axis. This means that N6's judgments of dissimilarities reflect the two dimensions about equally, while D3's judgments were almost exclusively based on the Y-B dimension. In fact, subjects labelled by D are color deficient subjects in the G-R dimension, who are lacking the sensitivity along the G-R dimension. (There was one normal and one color deficient subject who replicated the experiment twice. They are distinguished by symbols a and b in Figure 15.) ID MDS thus attempts to capture individual differences in (dis)similarity judgments by the differential weights attached to different dimensions by different individuals.

*** Insert Figures 14 and 15 about here ***

The stimulus and weight configurations presented above were derived by a nonmetric ID MDS procedure developed by the author, but they are virtually indistinguishable from those obtained by INDSCAL (Carroll and Chang, 1970b), which assumed a ratio level of measurement. (Stimuli P1 and P2 are somehow interchanged; we followed Borg and Groenen's labelling of the stimuli.)

Example 7: Digits. Not all types of individual differences in (dis)similarity judgments can be represented by differential weighting of dimensions. This example shows an instance of more complicated types of individual differences. The stimuli are 10 digits often used to display time in digital clocks. They were constructed by picking appropriate subsets of seven line segments arranged in the shape of 8 (Figure 16). For example, digit 1 is composed of segments 3 and 6, digit 2 is composed of segments 1, 3, 4, 5, and 7, and so on. In each trial a pair of digits were presented, and subjects were instructed to respond as quickly as possible whether the two presented stimuli were the "same" or "different". In half the trials the same digits were presented, and in the other half, different digits were presented. Reaction times that took the subjects to discriminate two different digits were used as input to ML MDS, especially designed to analyze reaction time data (Takane and Sergent, 1983).

*** Insert Figure 16 about here ***

There were two subjects, and the two subjects' data were separately analyzed by simple MDS. Figures 17 and 18 present the results of MDS obtained from the two subjects (Takane, 1994). The two configurations seem completely unrelated. For subject 1 (Figure 17), the stimuli are organized according to the patterns of presence and absence of line segments. The horizontal axis divides (roughly at the vertical dotted line) the digits having two top (or three) horizontal line segments (on the left) from those lacking them (on the right). The vertical axis divides (roughly at the horizontal dotted line) the digits having three vertical line segments 2, 3 and 6 (toward the bottom) from those lacking at least one of them. For subject 2 (Figure 18), on the other hand, the configuration (the unconstrained solution) roughly circles around, starting from 0 and ending with 9. To verify that the configuration is really circular, another solution was obtained under the constraint that all the points lie on a circumference of a circle. The constrained solution was found to fit better according to the minimum AIC criterion. This subject thus seems to perceive the stimuli as numbers (rather than collections of line segments arranged in a specific way), since he took much longer time to discriminate two digits which are closer in numerical value.

*** Insert Figures 17 and 18 about here ***

The way the same set of stimuli were perceived was entirely different for the two subjects. Clearly, this type of individual differences cannot be captured by the kind of ID MDS described above. Separate analyses by simple MDS are in order, as has been done here. Alternatively, an MDS that allows more general kinds of individual differences may be called for (e.g., Meulman and Verboon, 1983).

Example 8: MDS of line drawings of faces. For multiple sets of (dis)similarity data to be adequately described by the kind of ID MDS described earlier, they must exhibit certain patterns when analyzed separately by simple MDS. This example shows such patterns. The stimuli are eight faces (Figure 19) constructed by factorially combining two levels each of three features: Hair (long or short), Jaw (smooth or rugged), and Eye (bright or dark). As in Example 7, reaction times were used as input data to ML MDS. Again there were two subjects.

*** Insert Figure 19 about here ***

Figures 20 and 21 display derived stimulus configurations for the two subjects. Four dimensions were extracted and interpreted as follows: Dimension 1. Hair, Dimension 2. Jaw, Dimension 3. Eye, and Dimension 4. Sex consistency. The first three dimensions correspond with the three defining features of the stimuli. The fourth dimension is the additional dimension used by the subjects in performing the task. On this dimension (the coordinates of the stimuli on the fourth dimension are given in the right margin of the two figures), the stimuli most consistent with the most typical sex profiles (short hair, rugged jaw and dark eyes for males, and long hair smooth jaw and light eyes for females) are located at the top, while the remaining stimuli are located downward from the top according to their degrees of inconsistency with these typical sex profiles. The most striking observation is that the two configurations are remarkably similar, in fact almost identical. The two subjects use the same set of dimensions, only their saliency differs across the two subjects. ID MDS discussed earlier is most appropriate to describe this kind of individual differences.

*** Insert Figures 20 and 21 about here ***

Example 9: Facial expression data. We give a couple of more examples of ID MDS. A set of stimuli employed in the first example are shown in Figure 22. They were constructed by factorially combining two of the most important determinants of facial expressions, namely the curvature of eyes and the curvature of lips (Inukai, 1981). Dissimilarity judgements were obtained on these faces from twenty subjects (10 males and 10 females) using rating scales. Figure 23 shows the two-dimensional common stimulus configuration obtained by INDSCAL (Carroll and Chang, 1970a), assuming that the observed dissimilarity data are measured on a ratio scale. This configuration indicates that subjects' dissimilarity judgments are organized around the two physical attributes (the lips and the eyes dimensions) used to construct the stimuli. Note, however, that the two attributes are not strictly orthogonal in the psychological space. Concave (downward) eyes and convex lips seem to go together, and so do convex eyes and concave lips (lower right corner). The vertical direction roughly corresponds with the eyes dimension and the horizontal direction with the lips dimension.

*** Insert Figures 22 and 23 about here ***

In addition to the dissimilarity data, the experimenter also collected rating data on the same set of stimuli regarding the extent to which they expressed certain emotional dispositions. Directions (designated by arrows in Figure 23) indicate the directions with which certain emotional dispositions are most highly correlated. For example, "smile" is in quadrant 2, and this is the direction in which this emotional disposition is most clearly represented. Similarly, we may characterize other emotional dispositions as: 1. happiness - moderately concave eyes and moderately convex lips, 2. alertness - high concave eyes and straight lips, 3. anger - straight eyes and high concave lips, 4. weeping - moderately convex eyes and moderately convex lips, 5. sleeping - deep convex eyes and straight lips. Numbers in parentheses indicate multiple correlation coefficients.

Interestingly, there seem to be fairly clear sex differences in the evaluation of the two dimensions. Figure 24 shows individual differences weights. Squares represent weights for male subjects and circular dots those for female subjects. Generally speaking, the female subjects tend to put more emphasis on the lips dimension, while the male subjects on the eyes dimension.

*** Insert Figure 24 about here ***

Example 10: Body-parts data. The final example in this section involves a developmental change in conceptual relations among words (Takane, Young, and de Leeuw, 1977). Dissimilarity judgments between various body-parts were obtained from a group of adult subjects and a group of 6-year-old children, using the conditional rank-order method. (In this data collection method, one of the stimuli is chosen as the standard stimulus in turn, and the remaining stimuli are rank-ordered in terms of their similarity with the standard.) Figure 25 presents the derived stimulus configuration. The solution is three-dimensional, where the three dimensions are interpreted as: Dimension 1. Contrasts between face terms and limbs terms (both lower and upper limbs) with "body" in between. Dimension 2. Contrasts between upper and lower limbs with "body" and all face terms in the middle. Dimension 3. Represents the whole-part hierarchy (The term "body" is right in front, and more peripheral parts of the body are located toward the back of the configuration.)

*** Insert Figure 25 about here ***

Interestingly, there is a fairly systematic developmental difference in the weighting of these dimensions. Figure 26 depicts individual differences weights. White cubes represent

6-year-old children's weights, while black cubes represent those for adults. As can be seen, 6-year-old children tend to put more emphasis on dimension 2 than dimension 1 or 3, while adults are more heterogeneous. They split into two groups, one group placing more emphasis on dimension 1 and the other on dimension 3, but no adults put most emphasis on dimension 2. The distinction between upper and lower limbs is very important for young children. As they grow older, they come to realize certain parallelism between upper and lower limbs.

*** Insert Figure 26 about here ***

3.2 A brief account of ID MDS

In ID MDS, we are given N matrices of proximity data between n stimuli, denoted by n by n matrices, $\boldsymbol{O}_k$ $(k = 1, \cdots, N)$, the ij^{th} element, o_{ijk}, of which indicates the (dis)similarity between stimuli i and j obtained from subject k. In ID MDS used in the above examples, we represent o_{ijk} by the weighted Euclidean distance model,

$$d_{ijk} = \{\sum_{a=1}^{A} w_{ka}(x_{ia} - x_{ja})^2\}^{1/2}, \tag{9}$$

where x_{ia} is, as before, the coordinate of stimulus i on dimension a, and w_{ka} is the weight attached to dimension a by subject k. To eliminate scale indeterminacies between the stimulus coordinates and individual differences weights, x_{ia}'s are usually scaled so that $\sum_{i=1}^{n} x_{ia}^2 = n$. By applying ID MDS to the set of $\boldsymbol{O}_k$'s, we obtain the n by A matrix of stimulus coordinates, $\boldsymbol{X}$, and the N by A matrix of individual differences weights, $\boldsymbol{W}$, whose ka^{th} element is w_{ka}. These quantities are then plotted in graphical forms.

One important feature of the weighted Euclidean model is that the orientation of coordinate axes is uniquely determined. This means that we interpret the axes derived by the ID MDS, rather than look for some meaningful directions in the space to interpret as in the simple Euclidean distance model (1). This removes the burden of rotating the stimulus configurations in the right orientation as typically required in simple MDS.

Various algorithms have been developed for fitting ID MDS, of which INDSCAL (Carroll and Chang, 1970a) is still the most frequently used algorithm for ID MDS. In this method, observed dissimilarity data were first transformed into inner products (Schönemann, 1972; Young and Householder, 1938), which are then approximated by a weighted inner product model using a special iterative minimization algorithm. (The weighted inner product model is derived from the weighted Euclidean distance model with the same transformation by which the observed dissimilarity data were transformed into inner products.) INDSCAL, however, requires the ratio level of measurement for the observed dissimilarity data. ALSCAL (Takane, Young, and de Leeuw, 1977), on the other hand, allows monotonic transformations of observed (dis)similarity data. However, it fits squared distances to monotonically transformed (dis)similarity data. Consequently, it puts more emphasis on large dissimilarities. PROXSCAL (Busing, et al., 1997) is the first workable algorithm that directly fits the weighted Euclidean model to (monotonically transformed) (dis)similarity data.

The weighted distance model poses some difficulty for ML MDS. The number of parameters (individual differences weights) in the weighted distance model increases as more subjects are added to increase the number of observations (unless replicated observations are made within subjects). These are called incidental parameters. Asymptotic properties of the maximum likelihood estimators do not hold in this case. Clarkson and Gonzales (2001) resolved this problem by incorporating the subject weights as random effects, which were then integrated out to define the marginal likelihood. The marginal likelihood is maximized to estimate other (structural) parameters. Takane (1996, 1997) proposed a similar approach in unfolding analysis of multiple-choice questionnaire data.

4 Unfolding analysis

In this section we discuss MDS of preference data. Individual differences are prevalent in preference judgments. Unfolding analysis provides a way of representing individual differences in preference judgments. In unfolding analysis, preference data are regarded as representing proximity relations between subjects' ideal stimuli and actual stimuli.

4.1 Some examples of unfolding analysis

Example 11: Family composition preference. MDS is not restricted to the usual kind of (dis)similarity data. This example pertains to preference data on family compositions, namely how many boys and girls the subjects would like to have in their family. Stimuli are various family compositions constructed by factorially combining several levels of the number of boys (0 to 3) and the number of girls (0 to 3). Preference rankings were obtained on the 16 family compositions from 82 Belgian students (Delbeke, 1978). The data are presented in Figure 27. The 16 stimuli are numbered so that the index for the number of boys changes more rapidly. Specifically, 1. (0, 0), 2. (1, 0), 3. (2, 0), 4. (3, 0), 5. (0, 1), 6. (1, 1), 7. (2, 1), 8. (3, 1), 9. (0, 2), 10. (1, 2), 11. (2, 2), 12. (3, 2), 13. (0, 3), 14. (1, 3), 15. (2, 3), and 16. (3, 3), where the first number in the parentheses indicates the number of boys, and the second number indicates the number of girls. In the figure, the smaller numbers indicate more preferred combinations. For example, for the first subject, stimulus 6 (the (1, 1) combination) is the most preferred combination, stimulus 7 (the (2, 1) combination) is the second most preferred combination, and so on. One way to analyze these data is an extension of MDS, called unfolding analysis, in which it is assumed that:

1. Each subject has an ideal point in a multidimensional space corresponding to his or her ideal stimulus.
2. The closer a stimulus point is to one's ideal point, the more preferred that stimulus is by that subject.

Unfolding analysis locates both subjects' ideal points and stimulus points in a joint multidimensional space in such a way that more preferred stimuli are located close to the ideal points, while less preferred stimuli are located far from the ideal.

*** Insert Figure 27 about here ***

Figure 28 (Heiser 1981) shows the result of the unfolding analysis of these data. In this figure, stimulus points are indicated by a pair of numbers (e.g., 1, 3), while subjects' ideal points are indicated by x's. There is a strong tendency to prefer large families among the Belgian students. There is also some boy bias (tendency to prefer boys to girls). It is also evident that the difference between two small families (e.g., (1, 0) and (1, 1)) is much larger than the difference between two large families (e.g., (2, 2) and (2, 3)). Adding one girl has more pronounced effects when there is no girl in the family, but it has relatively small impact when there are already two girls in the family. The important point is that these observations, rather obvious in Figure 28, can hardly be made by merely inspecting the data table in Figure 27.

*** Insert Figure 28 about here ***

Example 12: A marketing research application. The next example shows importance of unfolding analysis in marketing research (DeSarbo and Rao, 1984). Stimuli are 12 brands of telecommunication devices (e.g., transceivers) characterized by seven descriptor variables such as calling feature (receive only or both send and receive) and style (military or cradle). Preference ratings were obtained from a group of 50 subjects characterized by 11 demographic variables such as age and sex. In the previous example, the stimulus and ideal point configuration was derived based on the observed preference data alone. In contrast, the background information about the stimuli and the subjects was explicitly incorporated in the present analysis, and was utilized in developing useful marketing strategies.

Figure 29 displays the derived stimulus and ideal point configuration. Stimuli are labelled by alphabet (A - I). Ideal points are indicated by circular dots. (White ones indicate *anti-ideal* points, meaning that stimuli further away from the points are more preferred by these subjects). It can be immediately observed that products A, B, and H are rather unpopular (few ideal points are located close to them), while the others (e.g., I, J, K, and L) are better accepted.

*** Insert Figure 29 about here ***

The background information is incorporated by constraining the coordinates of stimulus and ideal points as linear combinations of the background variables. An advantage of this is that the dimensional interpretation becomes easier. Products with high price, walkie-talkie style, and send and receive capability tend to be located toward the top of the configuration. Products with send and receive capability, high range, repertory dialing feature, speaker-phone option and stand up set-up tend to be located toward the right. Similarly, subjects who own single dwelling houses with many phone, and are married males having no children living at home, tend to be located toward the top, while older subjects who are educated, own multi-dwelling houses, but are lower in occupational status tend to be located toward the right. Thus, we immediately know what sort of products and subjects are located in which parts of the configuration, and what sort of people prefer, or do not prefer, what sort of products. This provides valuable information regarding which products should be selectively promoted to what sort of people.

Other valuable information can also be obtained by "manipulating" the configuration. Suppose we are interested to know how well new products will sell, if they are put on the market. The prospective products are defined by a set of values that take on the descriptor variables. By applying the weights used to derive the coordinates of the existing products, we can map the new products into the same space in which the existing products as well as subject points are already represented. The location of five new potential products (numbered from 1 to 5) are superposed in the figure. (Product 1 is slightly out of the frame of Figure 29.) It can be seen that products 1, 2 and 4 will not be popular. Products 3 and 5 will be better accepted, although whether they would sell well is rather questionable, since similar products are already on the market.

Essentially the same idea may be utilized for repositioning existing products. For example, product A is currently not very popular. We would like to make it more attractive to more people by changing some of its features. In the current example, product A is modified in three different ways. Symbols a, b, and c in the figure indicate the predicted relocation of product A after these changes: a. Provide send and receive capability for an additional cost of \$40; b. Increase the range by additional 200 feet for an additional cost of \$20; c. Add repertory dialing and redesign military style at no additional cost. Modifications involved in b and c are almost totally useless to make product A more competitive. a is the best strategy, but it still falls short of attracting many buyers.

A point can be located in the space which will attract the largest number of people. This point is such that the average (squared Euclidean) distance to all subject points is minimal. Thus, it might be called ideal product. This point is indicated by symbol "&" in the figure. This point, although having the minimal average squared distance to all the subjects points, may not be realizable by manipulating the defining features of the product.

The realizable ideal product is indicated by symbol "$" in the figure. This product has send and receive capability, no repertory dialing, range =1000 ft., speaker-phone option, price = 239.95, lie-down set-up and cradle style. Products I and J come fairly close to this profile, although I is probably better selling than J because the latter has a lot of competitors.

As it turned out, the ideal product is not radically better than some of the existing products. We might give up the idea of developing a product having the best overall popularity, but a product having a strong appeal to a certain group of people, and which does not suffer from strong competitions from other products. For example, a product located in the upper left portion of the configuration will attract people in that region, and will not have major competitors. Thus, it may prove to be a better product to sell. Furthermore, we know what sort of people will like this product. A sales strategy for promoting this product may take this into account.

4.2 A brief account of unfolding analysis

In unfolding analysis, we are given an N by n data matrix obtained from N subjects making preference judgments on n stimuli. By subjecting the data matrix to unfolding analysis, we obtain two coordinate matrices, one for subjects' ideal points (an N by A matrix, $\boldsymbol{Y}$), and the other for stimulus points (an n by A matrix, $\boldsymbol{X}$). Let y_{ka} denote the coordinate of subject k's ideal point on dimension a (the ka^{th} element of $\boldsymbol{Y}$), and x_{ia} the coordinate of stimulus i on dimension a (the ia^{th} element of $\boldsymbol{X}$). The Euclidean distance between ideal point k and stimulus point i is calculated by

$$d_{ki} = \{\sum_{a=1}^{A}(y_{ka} - x_{ia})^2\}^{1/2}. \quad (10)$$

The coordinates of ideal and stimulus points are determined in such a way that observed preference relations between subjects and stimuli are as inversely (monotonically) related to the distances as possible. The distance model used in unfolding analysis is often called ideal point model. In most cases, the coordinates are derived based on observed preference data alone as in Example 11. In some cases, however, additional information about the stimuli and the subjects are available, which may be incorporated in the analysis, as in Example 12. This is often done by constraining the coordinates as linear functions of the background variables.

Unfolding analysis is thus a special kind of MDS. It appeared as if a fitting procedure for unfolding analysis could easily be developed by simple modifications of standard MDS procedures. This expectation, however, turned out to be too optimistic (DeSarbo and Rao, 1984; Kruskal, et al., 1978; Takane, et al., 1977). The problem is that in many cases, so-called degenerate solutions are obtained. The degenerate solutions are those that fit to the data very well, but are substantively meaningless. Most often, ideal points and stimulus points are completely separated in the configuration, with distances between them being equal or nearly equal.

Kruskal's (1964a, b) original fitting criterion, called stress 1, normalized the (squared) raw stress (like (2)) by the sum of squared distances. This normalization convention does not discourage equal distances. Kruskal and Carroll (1969) proposed to normalize the (squared) raw stress by the sum of squared deviations of distances from the mean to induce more variability among distances. This, however, introduced variability of distances across subjects, while keeping the within-subject distances relatively homogeneous. It was then proposed to assess the fits within subjects, normalized by the within-subject sum of squared deviations, which are then averaged across subjects. This, however, led to another kind of degenerate solution; both ideal points and stimulus points are split into two, and so are the within-subject distances. Other attempts have also been made to prevent degenerate solutions by introducing other kinds of constraints like smoothness in the data transformations (Heiser, 1981, 1989; Kim, Rangaswamy, and DeSarbo (1999). More recently, Busing, Groenen, and Heiser (2005) proposed to penalize the stress function by an inverse of the variation (the ratio of variance to the squared mean) defined on disparities. This strategy, implemented in PREFSCAL, seems to be working. A potential danger (of the penalty method in general) is, of course, that we never know the right amount of nondegeneracy in the solution. For example, by imposing a penalty, we may be getting "too" nondegenerate a solution for a given data set. Some kind of objective criterion such as cross validation is required to choose appropriate values of penalty parameters.

Another interesting idea has been presented recently by Van Deun, Groenen, Heiser, Busing, and Delbeke (2005), who advocate not to try to avoid degenerate solutions, which may occur with good reasons. The ideal point (distance) model to account for individual differences in preference may be adequate only for a subset of subjects. By inspecting some of the degenerate solutions, they have identified that there are at least two other groups of subjects for whom the distance model is not necessarily adequate. For one of these groups, the vector preference model is more appropriate, and for the other group a signed compensatory distance model was more appropriate. In the former, subjects are represented as vectors, and their preferences are predicted by the projections of stimulus points to the vectors. This model is considered as a special case of the ideal point model, where the ideal points are located infinitely far away from the origin (Carroll, 1972). In the signed compensatory distance model, on the other hand, stimulus points are first projected onto the subject vectors, and distances are measured between ideal points and the projected stimulus points. These distances are then assumed to be inversely related to observed preference data. Van Deun et al. proposed to represent the entire data set by a mixture of the three models.

Another possibility is to use the idea of external unfolding (Carroll, 1972). In external unfolding, a stimulus configuration is derived first, and then subjects' ideal points are mapped into the stimulus configuration in such a way that distances between the stimulus points and ideal points best agree with observed preference relations. (This is contrasted with the "internal" unfolding, in which both stimulus and ideal point configurations are derived simultaneously.) A technique developed by Gower (1968) and Carroll (1972) may be used to map the ideal points into the stimulus configuration. A similar approach has been proposed by Meulman (1992), although in her case, the model is PCA (principal component analysis), and the subject points are estimated first rather than the other way round.

This last suggestion entails asymmetric treatment of subjects and stimuli. This may be more natural in unfolding analysis because subjects are typically treated as replications in multivariate analysis. For example, Takane (1996, 1997) proposed models for multidimensional analysis of multiple-choice questionnaire data that combine an item response theory (IRT) model and the ideal point model. It is assumed that both item categories and subjects are represented as points in a multidimensional space, and that the probability of a subject choosing a particular response category is stated as a decreasing function of the distance between them. Subject points are treated as random effects, which are then integrated out for the estimation of the item category points. The subject points may subsequently be mapped into the item category configuration, once the latter is determined. This is akin to external unfolding. A similar idea may be exploited in unfolding analysis in general. The idea is also similar to that of Clarkson and Gonzales (2001) for dealing with individual differences weights in the weighted distance model.

4.3 Ideal Point Discriminant Analysis

Ideal Point Discriminant Analysis (IPDA), proposed by Takane (1987), is a method for ML unfolding analysis of two-way contingency tables. Entries in contingency tables are frequencies of joint occurrences of row and column categories, and thus indicate degrees of similarities between them. In IPDA, both row and column categories of contingency tables are represented as points in a multidimensional space. The probabilities of column categories given a specific row category are decreasing functions of the distances between the corresponding row and column points. The row and column points are located in such a way that these probabilities best agree with the observed frequencies of co-occurrences of rows and columns. By incorporating information regarding row categories, a variety of model comparisons are possible, and the best fitting model identified through the minimum AIC criterion. IPDA combines the spatial representation feature of dual scaling (Nishisato, 1980) or correspondence analysis (Greenacre, 1984), and the detailed model evaluation feature of log-linear models (e.g., Bishop, Fienberg, and Holland, 1975) for analysis of contingency tables. (See Takane (1980) and Heiser (1981) for more details of the relationship between unfolding analysis and correspondence analysis. See also Takane, Yanai, and Mayekawa

(1991) for methods of correspondence analysis with linear constraints.)

Let $\boldsymbol{F} = (f_{ki})$ denote an R by C contingency table, where f_{ki} denotes the observed frequency of row k and column i. Let $\boldsymbol{G}$ denote an R by P matrix of "predictor" variables on rows of the table. When no special structures are assumed on rows of the table, we may set $\boldsymbol{G} = \boldsymbol{I}$. We let $\boldsymbol{Y}$ denote an R by A matrix of the coordinates of the row points, where A indicates, as before, the dimensionality of the representation space. We assume that this $\boldsymbol{Y}$ is obtained by linear combinations of $\boldsymbol{G}$, that is, $\boldsymbol{Y} = \boldsymbol{GB}$, where $\boldsymbol{B}$ is a P by A matrix of weights. Let $\boldsymbol{X}$ denote the C by A matrix of coordinates of column points. We assume that $\boldsymbol{X}$ is obtained by a weighted average of $\boldsymbol{Y}$, namely $\boldsymbol{X} = \boldsymbol{D}_C^{-1}\boldsymbol{F}'\boldsymbol{Y}$, where $\boldsymbol{D}_C$ is a diagonal matrix with column totals of $\boldsymbol{F}$ in its diagonals. (This constraint on $\boldsymbol{X}$ is useful for avoiding degenerate solutions often encountered in unfolding analysis.)

Distances between row and column points can now be calculated by (10), where y_{ka} and x_{ia} are the ka^{th} and the ia^{th} elements of $\boldsymbol{Y}$ and $\boldsymbol{X}$, respectively. Let $p_{i|k}$ denote the probability of column i given row k. We assume that this probability is given by

$$p_{i|k} = \frac{w_i \exp(-d_{ki}^2)}{\sum_{j=1}^{C} w_j \exp(-d_{kj}^2)}, \tag{11}$$

where $w_i(\geq 0; \sum_{i=1}^{C} w_i = 1)$ indicates the bias for column i. The likelihood for the entire set of observed frequencies in a contingency table is now stated as

$$L \propto \prod_{k=1}^{R} \prod_{i=1}^{C} p_{i|k}^{f_{ki}}. \tag{12}$$

Parameters in the model ($\boldsymbol{B}$ and $w_j, j = 1, \cdots, C$) are estimated in such a way that (12) is maximized. Once the maximum likelihood is obtained, AIC can be calculated by (7) for various model comparisons. The information matrix, defined by (8), can be evaluated at the maximum likelihood estimates of the model parameters. The Moore-Penrose inverse of the information matrix provides variance-covariance estimates of the estimated parameters.

Example 13: Seriousness of traffic accidents. Table 3 (Agresti, 1996) presents observed frequencies of traffic accidents, cross-classified by the seriousness (3 column categories; A. not injured or injured but not transported by emergency medical services, B. injured and transported by emergency medical services but not hospitalized, and C. injured and hospitalized, or injured and died), and 8 accident profiles defined by three binary variables: 1. gender of the driver (male or female), 2. location of accidents (urban or rural), and 3. whether seat-belt was on (yes or no). We apply IPDA to find out how the accident profile variables affect the seriousness of accidents.

*** Insert Table 3 about here ***

Table 4 summarizes the GOF statistics (AIC) for fitted models. The independence model refers to the model in which $p_{i|k} = f_{.i}/f_{..}$, and the saturated model to the model in which $p_{i|k} = f_{ki}/f_{k.}$, where $f_{k.}$, $f_{.i}$, and $f_{..}$ are the row total, the column total, and the grand total of $\boldsymbol{F}$, respectively. These non-spatial models serve as bench-mark models against spatial distance models. Table 4 indicates that neither the independence model nor the saturated model fit to the data as well as the best spatial distance model.

*** Insert Table 4 about here ***

Assuming that a spatial representation is adequate, what is the most appropriate dimensionality of the representation space, and is there any simple and meaningful relationship among the row categories? The dimensionality comparison was carried out for both with and without assuming a special relationship among the row categories. Table 4 indicates that in both cases the two-dimensional representation provides a better fit. To see if there is any simple relationship among the row categories, the model without any constraints on rows ($\boldsymbol{G} = \boldsymbol{I}$) and the model with the main effects of the three profile variables as the predictor variables for the row categories were compared. Table 4 indicates that the latter fits to the data better. This means that the derived coordinates of the row categories can be well approximated by linear functions of the main effects of the three profile variables. The model without any constraints on rows, on the other hand, is equivalent to having all seven possible predictor variables that can be formed from the three profile variables (three main effects, three two-way interactions, and one three-way interaction effect). The result thus indicates that there are no significant interaction effects among the profile variables affecting the seriousness of accidents.

Given that no interaction effects are necessary, are all three main effects significant? To see this, one of the three profile variables was removed at a time, and the reduced models were fitted in turn. The results are reported at the bottom portions of Table 4. It is found that none of these reduced models fit as well as the two-dimensional three-main-effect model. This means that the three main effects are all significant. From the size of the increase in the value of AIC, the location variable is the most significant, while the seat-belt variable is a close second.

Figure 30 displays the configuration of the row and column points derived under the best fitting model. Accidents tend to be more serious when they occur in rural areas and when passengers wore no seat belts. Female drivers are slightly more prone to serious accidents than male drivers. Ellipses surrounding the points show 99% confidence regions, indicating how reliably the point locations are estimated. They are all small, indicating that the point locations are very reliably estimated. This, of course, is due to the large sample size. (Note that the confidence region for column category A is so small that we cannot see it in the figure.)

*** Insert Figure 30 about here ***

Figure 31 depicts a similar configuration obtained under no constraints on the row categories. Confidence regions are still small, but are nonetheless uniformly larger than the corresponding ones in the previous figure, indicating that point locations are less accurately estimated. This shows the importance of having proper constraints to obtain more reliable estimates of the point locations.

*** Insert Figure 31 about here ***

We have quickly run through the kind of analyses that can be performed with IPDA. It was impossible to cover all important aspects of IPDA in this paper. We refer to Takane (1987, 1989, 1998), Takane, Bozdogan, Shibayama (1987), van der Heijden, Mooijaart, and Takane (1994) for more detailed accounts of IPDA. De Rooij and Heiser (2005) recently proposed a similar method for analysis of two-way contingency tables.

5 Concluding Remarks

Almost a quarter of a century has passed since two definitive articles on MDS (de Leeuw and Heiser, 1982; Wish and Carroll, 1982) appeared earlier in this series. Many significant events took place since then, of which the most significant one was perhaps the so-called computer revolution. The advent of personal computers with massive computational power has had profound impact on the way we think about statistics and data analysis. MDS is no exception. The number of stimuli that can be realistically handled by MDS has been scaled up considerably. There has been a report of MDS with 3,648 stimuli (Buja et al., 2001), which was totally unthinkable 30 years ago. People tended to downplay importance of the local optimum problem. (See Arabie (1973) for some critique). However, with today's computer it is not at all unrealistic to obtain 1,000 MDS solutions starting from 1,000 random initials to investigate the seriousness of the local optimum problem. Some attempts have also been made to ensure global optimum solutions algorithmically (Groenen and Heiser, 1996; Groenen, Heiser, and Meulman, 1999). XGvis (Buja et al., 2001) allows to investigate, interactively and visually, the nature of locally optimal solutions. Interaction and visualization seem to be the key words for future software developments in MDS.

Nonlinear dimension reduction constitutes another area of computer intensive methods. See Tenenbaum, de Silva, and Langford (2000), and Roweis and Saul (2000) for recent developments in this area. Proximity-based multivariate analysis (e.g., kernel methods) typically requires a large amount of computation because of its repetitive operations (inversion, spectral decomposition, etc.) applied to N by N matrices (kernel matrices), where N is the number of cases (subjects) in the data set.

The methodology of MDS is in a maturing stage. Counting only those published after 1990, at least four books (Borg and Groenen, 1997; Cox and Cox, 2001; Everitt and Rabe-Hesketh, 1997; Young and Hamer, 1994), and at least three review papers (Carroll and Arabie, 1998; Nosofsky, 1992; Wood, 1990) have been published. From an application perspective, Kruskal Wish (1978) is becoming a classic in MDS. With the development of more flexible and more reliable algorithms for MDS and unfolding analysis, it is expected that its applications will grow faster both in number and variety.

References

Agresti, A. (1996). *An introduction to categorical data analysis*. New York: Wiley.

Arabie, P. (1973). Concerning Monte Carlo evaluation of nonmetric multidimensional scaling algorithms. *Psychometrika*, **38**, 607-608.

Arabie, P. (1991). Was Euclid an unnecessarily sophisticated psychologist? *Psychometrika*, **56**, 567-587.

Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975). *Discrete multivariate analysis: Theory and practice*. Cambridge, MA: The MIT Press.

Borg, I., and Groenen, P. (1997). *Modern multidimensional scaling: Theory and applications*. New York: Springer.

Buja, A., Swayne, D. F. (2002). Visualization methodology for multidimensional scaling. *Journal of Classification*, **19**, 7-43.

Buja, A., Swayne, D. F., Littman, M. L., Dean, N., and Hoffman, H. (2001). XGvis: Interactive data visualization with multidimensional scaling. To appear in *Journal of Computational and graphical Statistics*.

Busing, F. M. T. A., Commandeur, J. J. F., and Heiser, W. J. (1997). PROXSCAL: A multidimensional scaling program for individual differences scaling with constraints. In W. Bandilla and F. Faulbaum (Eds.), *Softstat '97 Advances in statistical software* (pp. 237-258). Stuttgart, Germany: Lucius.

Busing, F. M. T. A., Groenen, P. J.F., and Heiser, W. (2005). Avoiding degeneracy in multidimensional unfolding by penalizing on the coefficient of variation. *Psychometrika*, in press.

Carroll, J. D. (1972). Individual differences and multidimensional scaling. In R. N. Shepard, A. K. Romney, and S. B. Nerlove (Eds.), *Multidimensional scaling. Volume 1: Theory* (pp. 105-155). New York: Seminar press.

Carroll, J. D., and Arabie, P. (1998). Multidimensional scaling. In M. H. Birnbaum (ed.), *Handbook of perception and cognition, Vol. 3: Measurement, judgment and decision making* (pp. 179 250). San Diego, CA: Academic Press.

Carroll, J. D., and Chang, J. J. (1970a). Individual differences and multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika*, **35**, 282-319.

Carroll, J. D., and Chang, J. J. (1970b). Reanalysis of some color data of Helm's by INDSCAL procedure for individual differences multidimensional scaling. *Proceedings of the* 78^{th} *Annual Convention of the American Psychological Association*, 137-138.

Clarkson, D. B., and Gonzales, R. (2001). Random effects diagonal metric multidimensional scaling models. *Psychometrika*, **66**, 25-43.

Cox, T. F., and Cox, M. A. A. (2001). *Multidimensional scaling.* 2nd ed. Boca Raton: Chapman and Hall.

Delbeke, L. (1978). Enkele analyses op voorkeuoordelen voor gezinssamenstellingen. *Centrum voor Mathematishe Psychologie en Psychologische Methodologie.* Katholieke Universiteit, Leuven.

de Leeuw, J. (1977). Applications of convex analysis to multidiemnsional scaling. In J. R. Barra, F. Brodeau, G. Romier, and B. van Cutsem (Eds.), *Recent development in statistics* (pp. 133-145). Amsterdam: North Holland Publishing.

de Leeuw, J. (1988). Convergence of the majorization method for multidimensional scaling. *Journal of Classification*, **5** 163-180.

de Leeuw, J., and Heiser, W. (1977). Convergence of correction-matrix algorithms for multidimensional scaling. In J. C. Lingoes (Ed.) Geometric representations of relational data. Ann Arbor, MI: Mathesis Press.

de Leeuw, J., and Heiser, W. (1982). Theory of multidimensional scaling. In R. P. Krishnaiah and L. N. Kanal, (Eds.), *Handbook of statistics 2: Classification, pattern recognition, and reduction of dimensionality* (pp. 285-316). Amsterdam: North Holland.

De Rooij, M., and Heiser, W. (2004). Graphical representations and odds ratios in a distance-association models for the analysis of cross classified data. *Psychometrika*, in press.

DeSarbo, W. S., and Rao (1984). GENFOLD2: A set of models and algorithms for the GENeral unFOLDing analysis of preference and dominance data. *journal of Classification*, **1**, 147-186.

Dunn-Rankin, P., and Leton, D. A. (1975). Differences between physical template matching and subjective similarity estimates of Japanese letters. *Japanese psychological Research*, **15**, 51-58.

Ekman, G. (1954). Dimensions of color vision. *Journal of Psychology*, **38**, 467-474. .

Everitt, B. S., and Rabe-Hesketh, S. (1997). *The analysis of proximity data.* New York: Wiley.

Gower, J. C. (1968). Adding a point to vector diagrams in multivariate analysis. *Biometrika*, **55**, 582-585.

Greenacre, M. J. (1984). *Theory and applications of correspondence analysis.* London: Academic Press.

Groenen, P. J. F., and Heiser, W. J. (1996). The tunneling method for global optimization in multidimensional scaling. *Psychometrika*, **61**, 529-550.

Groenen, P. J. F., Heiser, W. J., and Meulman, J. J. (1999). Global optimization in least-squares multidimensional scaling by distance smoothing. *Journal of Classification*, **16**, 225-254.

Heiser, W. J. (1981). Unfolding analysis of proximity data. Unpublished doctral dissertation, Department of Data Theory, Leiden University, Leiden, The Netherlands.

Heiser, W. J. (1989). Order invariant unfolding analysis under smoothness restrictions. In G. De Soete, H. Feger, and K. C. Klauer (Eds.), *New developments in psychological choice modeling* (pp. 3-31), Amsterdam: North-Holland.

Heiser, W. J. (1991). A generalized majorization method for least squares multidimensional scaling of pseudodistances that may be negative. *Psychometrika*, **55**, 7-27.

Helm, C. E. (1964). Multidimensional ratio scaling analysis of perceived color relations. *Jpurnal of the Optical Society of America*, **54**, 256-262.

Hubert, L., Arabie, P., and Hesson-Mcinnis, M. (1992). Multidimensional scaling in the city-block metric: A combinatorial approach. *Journal of classification*, **9**, 211-236.

Inukai, Y. (1981). Analysis of perceptual dimensions of schematic facial expressions via three-way multidimensional scaling. *Behaviormetrika*, **9**, 1-20.

Kim, C., Rangswamy, A., and DeSarbo, W. S. (1999). A quasi-metric approach to multidiemnsional unfolding for reducing the occurrence of degenerate solutions. *Multivariate Behavioral Research*, **34**, 143-180.

Kruskal, J. B. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, **29**, 1-29.

Kruskal, J. B. (1964b). Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, **29**, 115-129.

Kruskal, J. B., and Carroll, J. B. (1969). Geometric models and badness of fit functions. In R. P. Krishnaiah (ed.), *Multivariate analysis-II* (pp. 639-671). New York: Academic Press.

Kruskal, J. B., and Wish, M. (1978). *Multidimensional scaling.* Beverly Hills, CA: Sage Publications.

Kruskal, J. B., Young, F. W., and Seery, J. B. (1978). How to use KYST, a very flexible program to do multidiemsnional scaling and unfolding (Tech Rep.). Murray Hill, NJ: Bell Laboratories.

Meulman, J. (1992). The integration of multidimensional scaling and multivariate analysis with optimal transformations. *Psychometrika*, **57**, 539-565.

Meulman, J., and Verboon, P. (1993). Points of view analysis revisited: Fitting multidimensional structures to optimal distance components with cluster restrictions on the variables. *Psychometrika*, **58**, 7-35.

Nishisato, S. (1980). *Analysis of categorial data: Dual scaling and its applications.* Toronto: University of Toronto Press.

Nosofsky, R. M. (1992). Similarity scaling and cognitive process models. *Annual Review of Psychology*, **43**, 25-53.

Ramsay, J. O. (1977). Maximum likelihood estimation in multidimensional scaling. *Psychometrika*, **42**, 337-360.

Ramsay, J. O. (1978). Confidence regions for multidimensional scaling analysis. *Psychometrika*, **43**, 145-160.

Ramsay, J. O. (1980). Some small sample results for maximum likelihood estimation in multidimensional scaling. *Psychometrika*, **45**, 141-146.

Ramsay, J. O. (1982). Some statistical approaches to multidimensional scaling data. *Journal of the Royal Statistical Society, Series A (General)*, **145**, 285-312.

Rapoport, A., and Fillenbaum, S. (1972). An experimental study of semantic structures. In A. K. Romney, R. N. Shepard, and S. B. Nerlove (Eds.), *Multidimensional scaling: Theory and applications in the Behavioral sciences, Vol II*, (pp. 93-131).

Rothkopf, E. Z. (1957). A measure of stimulus similarity and errors in some paired associate learning. *Journal of Experimental Psychology*, **53**, 94-101.

Roweis, S. T., and Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science*, **290**, 2323-2326.

Schiffman, S. S., Reynolds, M. L., and Young, F. W. (1981). *Introduction to multidimensional scaling*. New York: Academic Press.

Schneider, B. (1972). Multidimensional scaling of color difference in the pigeon. *Perception and Psychophysics*, **12**, 373-378.

Schönemann, P. H. (1972). An algebraic solution for the class of subjective metrics models. *Psychometrika*, **37**, 441-451.

Schönemann, P. H. (1977). Similarity of rectangles. *Journal of Mathematical Psychology*, **16**, 161-165.

Shepard, R. N. (1962). Analysis of proximities: Multidimensional scaling with an unknown distance function: I and II. *Psychometrika*, **27**, 125-140 and 219-246.

Shepard, R. N. (1963). Analysis of proximities as a technique for the study of information processing in man. *Human Factors*, **5**, 19-34.

Shepard, R. N. (1964). Extracting latent structure from behavioral data. In Proceedings of the 1964 Symposium on Digital Computing. Bell Telephone Laboratories.

Takane, Y. (1978). A maximum likelihood method for nonmetric multidimensional scaling: I. The case in which all empirical pairwise orderings are independent - theory and evaluations. *Japanese Psychological Research*, **20**, 7-17 and 105-114.

Takane, Y. (1980). Analysis of categorizing behavior by a quantification method. *Behaviormetrika*, **8**, 75-86.

Takane, Y. (1981). Multidimensional successive categories scaling: A maximum likelihood method. *Psychometrika*, **46**, 9-28.

Takane, Y. (1987). Analysis of contingency tables by ideal point discriminant analysis. *Psychometrika*, **52**, 493-513.

Takane, Y. (1989). Ideal point discriminant analysis: Implications for multi-way data analysis. In R. Coppi and S. Bolasco (Eds.), *Multiway data analysis* (pp. 287-299). Amsterdam: North Holland.

Takane, Y. (1994). A review of applications of AIC in psychometrics. In H. Bozdogan (Ed.), *Proceedings of the First US/Japan Conference on the Frontiers of Statistical Modeling: An informational approach* (pp. 379-403). Dortrecht: Kluver Academic Publisher.

Takane, Y. (1996). An item response model for multidimensional analysis of multiple-choice questionnaire data. *Behaviormetrika*, **23**, 153-167.

Takane, Y. (1997). Choice model analysis of the pick any/n type of binary data. *Japanese Psychological Research*, **40**, 31-39.

Takane, Y. (1998). Visualization in ideal point discriminant analysis. In J. Blasius, and M. J. Greenacre (eds.), *Visualization of categorical data* (pp. 441-459). New York: Academic Press.

Takane, Y., Bozdogan, H., and Shibayama, T. (1987). Ideal point discriminant analysis. *Psychometrika*, **52**, 371-392.

Takane, Y., and Carroll, J. D. (1981). Nonmetric maximum likelihood multidimensional scaling from directional rankings of similarities. *Psychometrika*, **46**, 389-405.

Takane, Y., and Sergent, J. (1983). Multidimensional scaling models for reaction times and same-different judgments. *Psychometrika*, **48**, 393-423.

Takane, Y., and Shibayama, T. (1986). Comparison of models for the stimulus recognition data. In J. de Leeuw, W. J. Heiser, J. Meulman, and F. Critchley (eds.), *Multidimensional data analysis* (pp. 119-138). Leiden: DSWO Press.

Takane, Y., and Shibayama, T. (1992). Structures in stimulus identification data. In F. G., Ashby (Ed.), *Probabilistic multidimensional models of perception and cognition* (pp. 335-362). Hillsdale, NJ: Earlbaum.

Takane, Y., van der Heijden, P. G. M., and Browne, M. W. (2003). On likelihood ratio tests for dimensionality selection. In Higuchi, T., Iba, Y., and Ishiguro, M. (Eds.), *Proceedings of Science of Modeling: The 30th Anniversary Meeting of the Information Criterion (AIC)*, (pp. 348-349). Report on Research and Education 17. Tokyo: The Institute of Statistical Mathematics.

Takane, Y., Yanai, H., and Mayekawa, S. (1991). Relationships among several methods of linearly constrained correspondence analysis. *Psychometrika*, **56**, 667-684.

Takane, Y., Young, F. W., and de Leeuw, J. (1977). Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. *Psychometrika*, **42**, 7-67.

Tenenbaum, J. B., de Silva, V., and Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, **290**, 2319-2322.

Torgerson, W. S. (1952). Multidimensional scaling: I. Theory and method. *Psychometrika*, **17**, 401-409.

Wish, M., and Carroll, J. D. (1982). Multidimensional scaling and its applications. In R. P. Krishnaiah and L. N. Kanal, (Eds.), *Handbook of statistics 2: Classification, pattern recognition, and reduction of dimensionality* (pp. 317-345). Amsterdam: North Holland.

Wood, P. (1990). Application of scaling to developmental research. In A. von Eye (ed.), *Statistical methods in longitudinal research, Vol I: Principles and structural change* (pp. 225-256). St. Paul, MN: Assessment Systems Corporation.

van der Heijden, P. G. M., Mooijaart, A., and Takane, Y. (1994). In M. J. Greenacre and J. Blasius (Eds.), *Correspondence analysis in the social sciences* (pp. 79-111). New York: Academic Press.

Van Deun, K., Groenen, P. J. F., Heiser, W. J., Busing, F. M. T. A., and Delbeke, L. (2004). Interpreting degenerate solutions in unfolding by use of the vector model and the compensatory distance model. *Psychometrika*, in press.

Young, F. W., and Hamer, R. M. (1994). *Theory and applications multidimensional scaling.* Hillsdale, NJ: Erlbaum Associates.

Young, G., and Householder, A. S. (1938). Discussion of a set of points in terms of their mutual distances. *Psychometrika*, **3**, 19-22.

Table 1: Flying Mileage between 10 US Cities (Kruskal and Wish, 1978)

	Atl	Chi	Den	Hou	LA	Mia	NY	SF	Sea	DC
Atlanta	0	587	1212	701	1936	604	748	2139	2182	543
Chicago	587	0	920	940	1745	1188	713	1858	1737	597
Denver	1212	920	0	879	831	1726	1631	949	1021	1494
Houston	701	940	879	0	1374	968	1420	1645	1891	1220
Los Angels	1936	1745	831	1374	0	2339	2451	347	959	2300
Miami	604	1188	1726	968	2339	0	1092	2594	2734	923
New York	748	713	1631	1420	2451	1092	0	2571	2408	205
San Franc.	2139	1858	949	1645	347	2594	2571	0	678	2442
Seattle	2182	1737	1021	1891	959	2734	2408	678	0	2329
Wash. D.C.	543	597	1494	1220	2300	923	205	2442	2329	0

Table 2: Summary statistics for MDS analyses of the rectangle data

Model	$-2\ln L^*$	AIC	n_π
Saturated model	1127.9	2759.9	816
Unconstrained solutions			
3-dimensional	1543.1	1645.1	51
2-dimensional	1571.7	1645.7	37
Constrained solutions			
Area and shape hypothesis	2037.2	2065.2	14
Height and width hypothesis	1910.8	1946.8	18
Schönemann's hypothesis	1707.2	1737.2	15

Table 3: Illustrative data for IPDA. The original data were provided in courtesy of Dr. Cristina Cook, Mental Care Development, Augusta, Maine, and were reported on page 225 of Agresti (1996).

				Seriousness			
Plotting Symbol	Gender	Location	Seat-Belt	A	B	C	Row Total
fun	Female	Urban	No	7462	720	101	8283
fuy			Yes	11713	577	56	12346
frn		Rural	No	3319	710	190	4219
fry			Yes	6228	564	99	6891
mun	Male	Urban	No	10517	566	110	11193
muy			Yes	11052	259	38	11349
mrn		Rural	No	6264	710	233	7207
mry			Yes	6767	353	86	7206
			Column Total	63322	4459	913	68694

Table 4: Goodness of fit statistics for various IPDA models

Model	Dimensionality	AIC	# of para.
Saturated		666.4	(16)
Independence		2594.3	(2)
No constraints on rows	2	664.5	(15)
	1	713.5	(9)
Three main effects	2	*656.9	(7)
	1	715.1	(5)
Gender eliminated	2	1054.8	**(4)
Location eliminated	2	1531.7	(5)
Seat-belt eliminated	2	1495.1	(5)

*Minimum AIC solution, **Solution almost unidimensional

Figure Captions

Figure 1. The map of 10 US cities derived from flying mileage.

Figure 2. Confusion data for 36 Morse Code signals (Rothkopf, 1957).

Figure 3. A stimulus configuration derived from Morse code signal confusion data.

Figure 4. An MDS configuration of Japanese Kana characters (phonetic symbols).

Figure 5. A stimulus configuration of Have words with four sorting clusters elicited by one of the subjects.

Figure 6. A two-dimensional color space for the pigeon.

Figure 7. A two-dimensional color space for humans derived by Shepard (1964) from the data collected by Ekman (1954).

Figure 8. Shepard's diagram and the best monotonic transformation of the observed confusion probabilities among Morse Code signals.

Figure 9. The unconstrained two-dimensional configuration of rectangles derived by ML MDS.

Figure 10. The stimulus configuration of rectangles derived under the area and shape hypothesis.

Figure 11. The stimulus configuration of rectangles derived under the height and width hypothesis.

Figure 12. Schönemann's hypothesis for similarity of rectangles .

Figure 13. A focal point hypothesis.

Figure 14. The two-dimensional common stimulus space derived from Helm's color data.

Figure 15. The plot of individual differences weights attached to the two dimensions of the color space.

Figure 16. The construction of segmented numerals (digits) used in Example 7.

Figure 17. The configuration of digits for Subject 1.

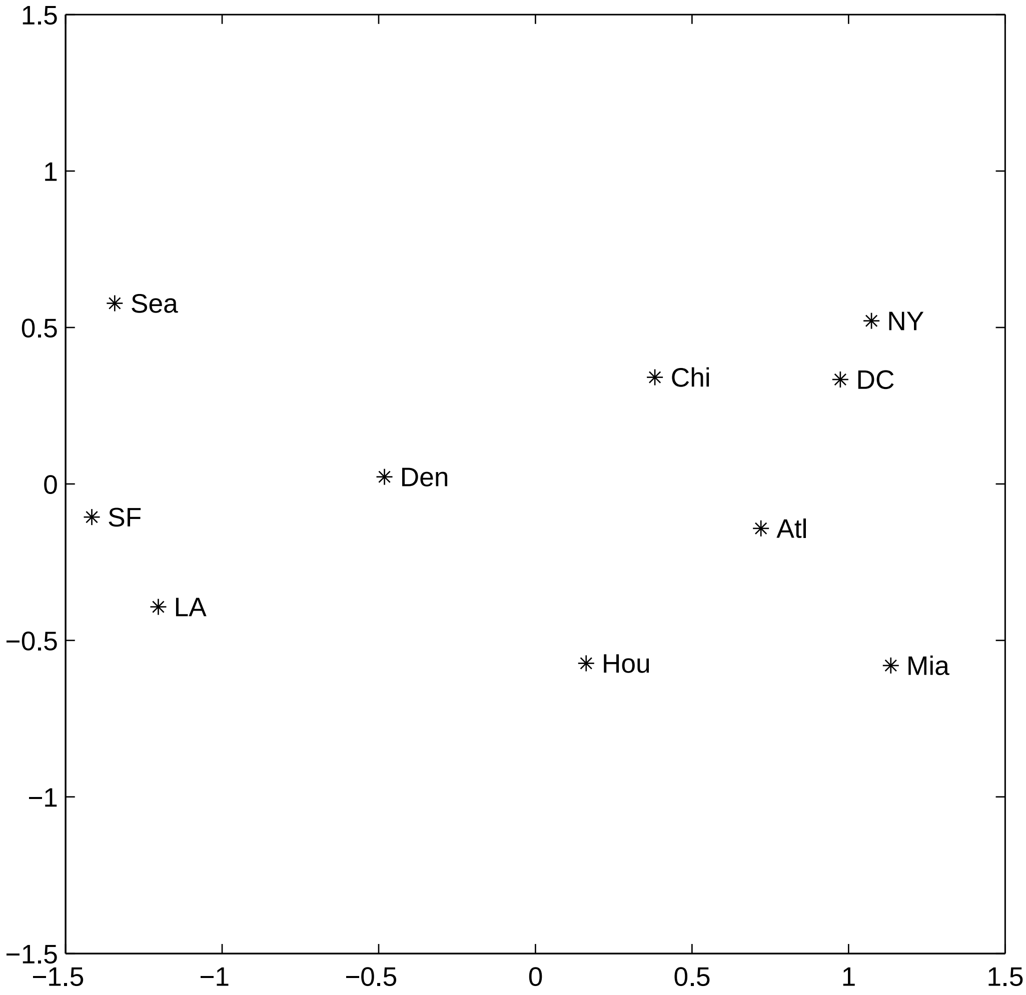
Figure 18. The configuration of digits for Subject 2.

Figure 19. Stimuli used in Example 8 (Takane and Sergent, 1983).

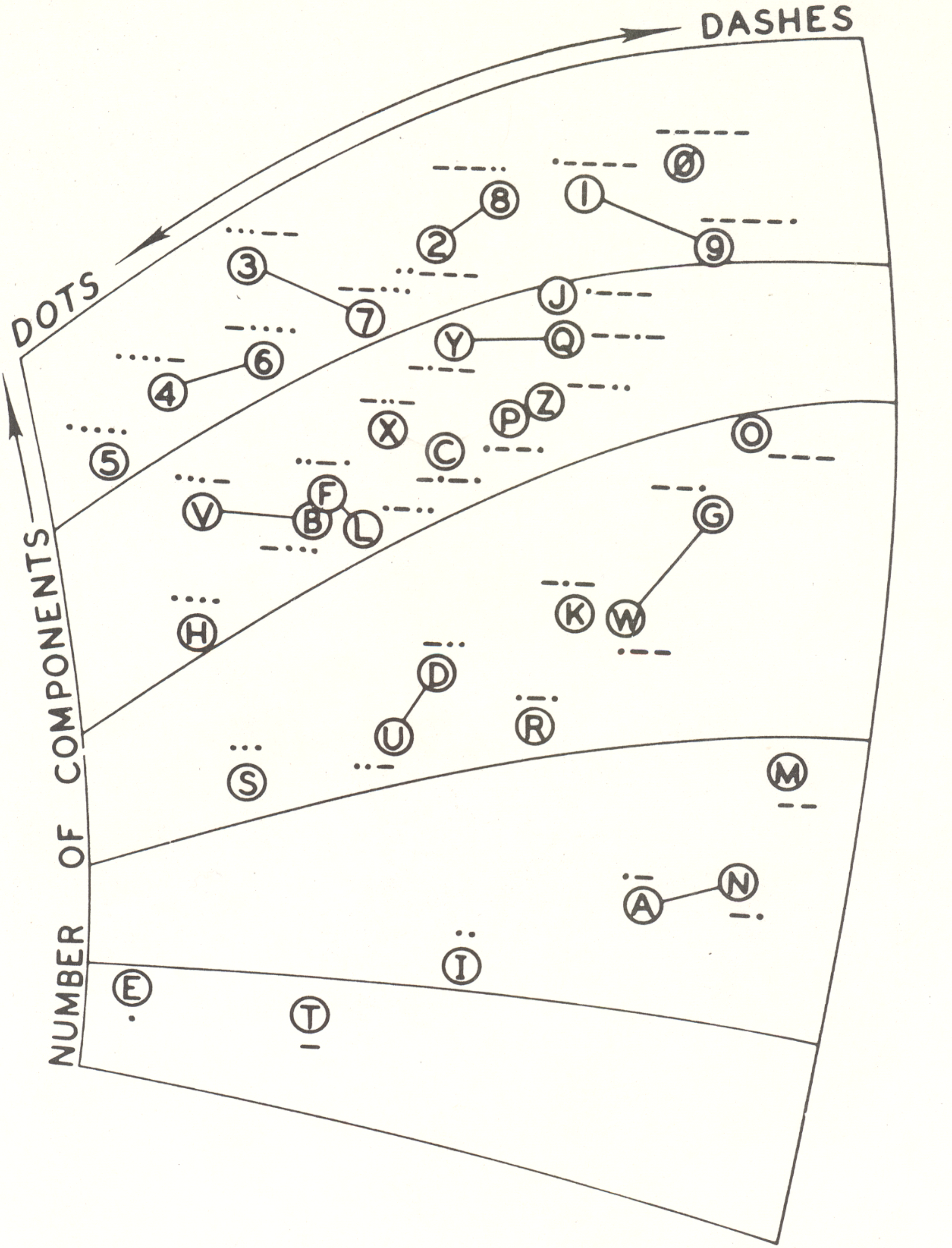
Figure 20. The four-dimensional configuration of faces for Subject 1.

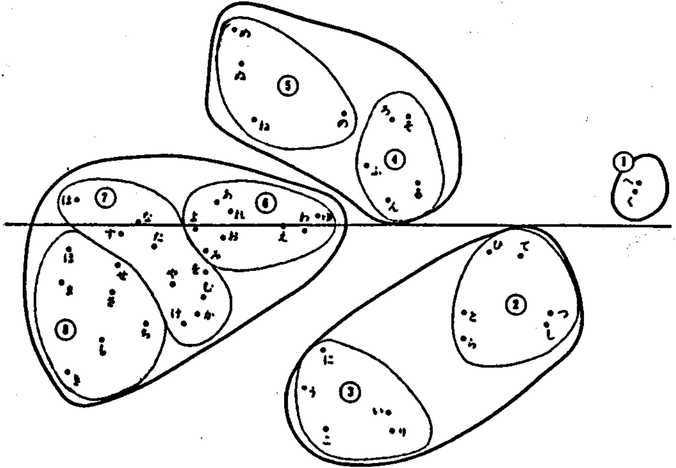
Figure 21. The four-dimensional configuration of faces for Subject 2.

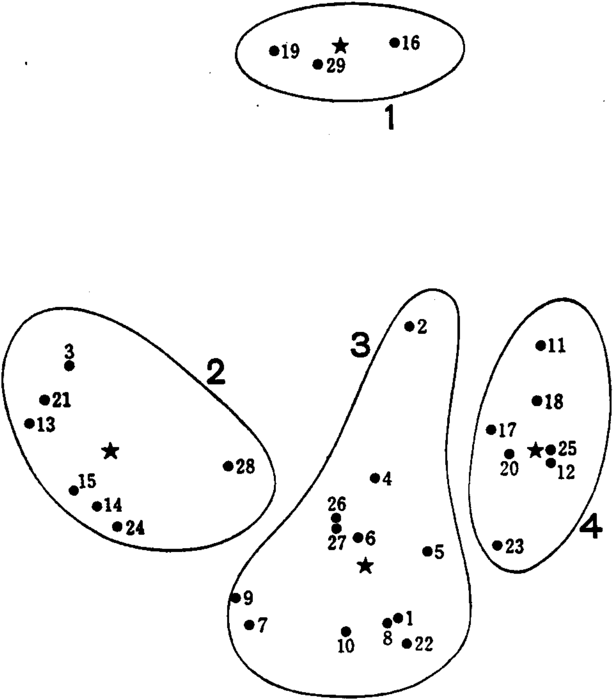
Figure 22. Stimuli used in Example 9 (Inukai, 1981).


Sea
NY
Chi
DC
Den
SF
Atl
LA
Hou
Mia
1.5
1
0.5
0
−0.5
−1
−1.5
−1.5
−1
−0.5
0
0.5
1
1.5

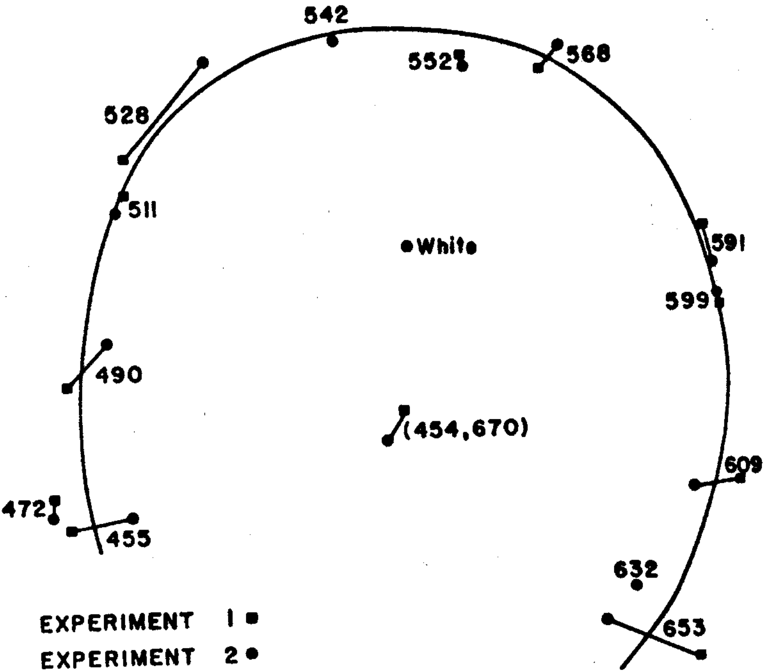
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	1	2	3	4	5	6	7	8	9	0	
A	92	04	06	13	03	14	10	13	46	05	22	03	25	34	06	06	09	35	23	06	37	13	17	12	07	03	02	07	05	05	08	06	05	06	02	03	A
B	05	84	37	31	05	28	17	21	05	19	34	40	06	10	12	22	25	16	18	02	18	34	08	84	30	42	12	17	14	40	32	74	43	17	04	04	B
C	04	38	87	17	04	29	13	07	11	19	24	35	14	03	09	51	34	24	14	06	06	11	14	32	82	38	13	15	31	14	10	30	28	24	18	12	C
D	08	62	17	88	07	23	40	36	09	13	81	56	08	07	09	27	09	45	29	06	17	20	27	40	15	33	03	09	06	11	09	19	08	10	05	06	D
E	06	13	14	06	97	02	04	04	17	01	05	06	04	04	05	01	05	10	07	67	03	03	02	05	06	05	04	03	05	03	05	02	04	02	03	03	E
F	04	51	33	19	02	90	10	29	05	33	16	50	07	06	10	42	12	35	14	02	21	27	25	19	27	13	08	16	47	25	26	24	21	05	05	05	F
G	09	18	27	38	01	14	90	06	05	22	33	16	14	13	82	52	23	21	05	03	15	14	32	21	23	39	15	14	05	10	04	10	17	23	20	11	G
H	03	45	23	25	09	32	08	87	10	10	09	29	05	08	08	14	08	17	37	04	36	59	09	33	14	11	03	09	15	43	70	35	17	04	03	03	H
I	64	07	07	13	10	08	06	12	93	03	05	16	13	30	07	03	05	19	35	16	10	05	08	02	05	07	02	05	08	09	06	08	05	02	04	05	I
J	07	09	38	09	02	24	18	05	04	85	22	31	08	03	21	63	47	11	02	07	09	09	09	22	32	28	67	66	33	15	07	11	28	29	26	23	J
K	05	24	38	73	01	17	25	11	05	27	91	33	10	12	31	14	31	22	02	02	23	17	33	63	16	18	05	09	17	08	08	18	14	13	05	06	K
L	02	69	43	45	10	24	12	26	09	30	27	86	06	02	09	37	36	28	12	05	16	19	20	31	25	59	12	13	17	15	26	29	36	16	07	03	L
M	24	12	05	14	07	17	29	08	08	11	23	08	96	62	11	10	15	20	07	09	13	04	21	09	18	08	05	07	06	06	05	07	11	07	10	04	M
N	31	04	13	30	08	12	10	16	13	03	16	08	59	93	05	09	05	28	12	10	16	04	12	04	06	11	05	02	03	04	04	06	02	02	10	02	N
O	07	07	20	06	05	09	76	07	02	39	26	10	04	08	86	37	35	10	03	04	11	14	25	35	27	27	19	17	07	07	06	18	14	11	20	12	O
P	05	22	33	12	05	36	22	12	03	78	14	46	05	06	21	83	43	23	09	04	12	19	19	19	41	30	34	44	24	11	15	17	24	23	25	13	P
Q	08	20	38	11	04	15	10	05	02	27	23	26	07	06	22	51	91	11	02	03	06	14	12	37	50	63	34	32	17	12	09	27	40	58	37	24	Q
R	13	14	16	23	05	34	26	15	07	12	21	37	14	12	12	29	08	87	16	02	23	23	62	14	12	13	07	10	13	04	07	12	07	09	01	02	R
S	17	24	05	30	11	26	05	59	16	03	13	10	05	17	06	06	03	18	96	09	56	24	12	10	06	07	08	02	02	15	28	09	05	05	05	02	S
T	13	10	01	05	46	03	06	06	14	06	14	07	06	05	06	11	04	04	07	96	08	05	04	02	02	06	05	05	03	03	03	08	07	06	14	06	T
U	14	29	12	32	04	32	11	34	21	07	44	32	11	13	06	20	12	40	51	06	93	57	34	17	09	11	06	06	16	34	10	09	09	07	04	03	U
V	05	17	24	16	09	29	06	39	05	11	26	43	04	01	09	17	10	17	11	06	32	92	17	57	35	10	10	14	28	79	44	36	25	10	01	05	V
W	09	21	30	22	09	36	25	15	04	25	29	18	15	06	26	20	25	61	12	04	19	20	86	22	25	22	10	22	19	16	05	09	11	06	03	07	W
X	07	64	45	19	03	28	11	06	01	35	50	42	10	08	24	32	61	10	12	03	12	17	21	91	48	26	12	20	24	27	16	57	29	16	17	06	X
Y	09	23	62	15	04	26	22	09	01	30	12	14	05	06	14	30	52	05	07	04	06	13	21	44	86	23	26	44	40	15	11	26	22	33	23	16	Y
Z	03	46	45	18	02	22	17	10	07	23	21	51	11	02	15	59	72	14	04	03	09	11	12	36	42	87	16	21	27	09	10	25	66	47	15	15	Z
1	02	05	10	03	03	05	13	04	02	29	05	14	09	07	14	30	28	09	04	02	03	12	14	17	19	22	84	63	13	08	10	08	19	32	57	55	1
2	07	14	22	05	04	20	13	03	25	26	09	14	02	03	17	37	28	06	05	03	06	10	11	17	30	13	62	89	54	20	05	14	20	21	16	11	2
3	03	08	21	05	04	32	06	12	02	23	06	13	05	02	05	37	19	09	07	06	04	16	06	22	25	12	18	64	86	31	23	41	16	17	08	10	3
4	06	19	19	12	06	25	14	16	07	21	13	19	03	03	02	17	29	11	09	03	17	55	08	37	24	03	05	26	44	89	42	44	32	10	03	03	4
5	08	45	15	14	02	45	04	67	07	14	04	41	02	00	04	13	07	09	27	02	14	45	07	45	10	10	14	10	30	69	90	42	24	10	06	05	5
6	07	80	30	17	04	23	04	14	02	11	11	27	06	02	07	16	30	11	14	03	12	30	09	58	38	39	15	14	26	24	17	86	69	14	05	14	6
7	06	33	22	14	05	25	06	04	06	24	13	32	07	06	07	36	39	12	06	02	03	13	09	30	30	50	22	29	18	15	12	61	85	70	20	13	7
8	03	23	40	06	03	15	15	06	02	33	10	14	03	06	14	12	45	02	06	04	06	07	05	24	35	50	42	29	16	16	09	30	60	89	61	26	8
9	03	14	23	03	01	06	14	05	02	30	06	07	16	11	10	31	32	05	06	07	06	03	08	11	21	24	57	39	09	12	04	11	42	56	91	78	9
0	09	03	11	02	05	07	14	04	05	30	08	03	02	03	25	21	29	02	03	04	05	03	02	12	15	20	50	26	09	11	05	22	17	52	81	94	0
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	1	2	3	4	5	6	7	8	9	0	

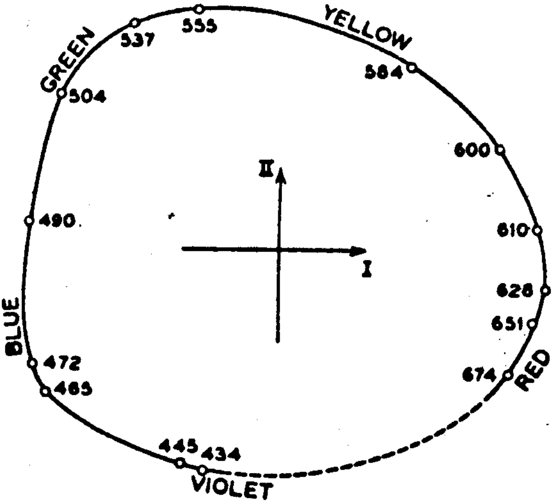

DASHES
DOTS
NUMBER OF COMPONENTS
Ø
1
8
2
9
3
7
J
Y
Q
6
4
Z
P
X
C
O
5
F
V
B
L
G
H
K
W
D
R
U
S
M
A
N
I
E
T


1
2
3
4
5
6
7
8

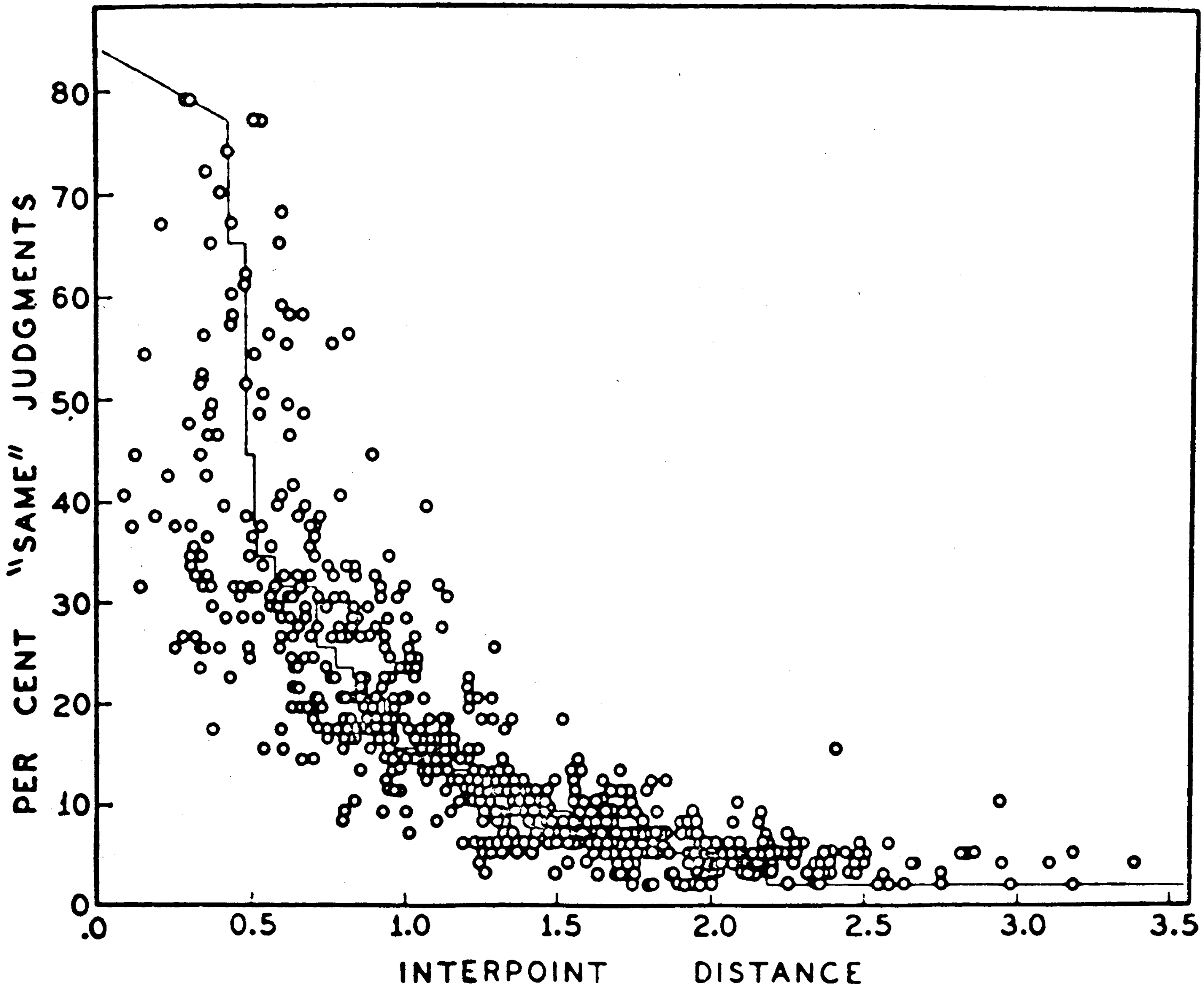


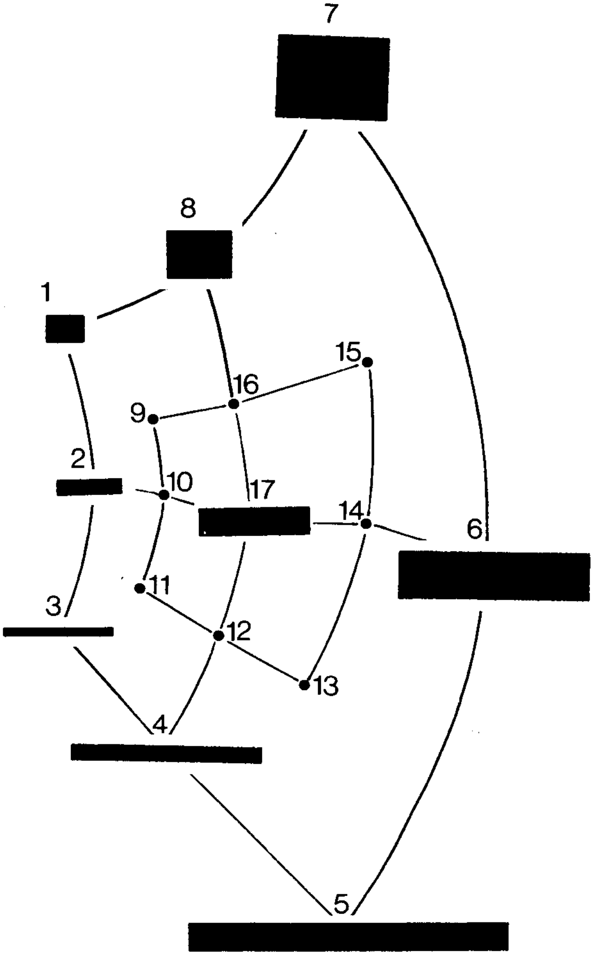
19
29
16
1
3
21
13
28
15
14
24
2
3
2
4
26
27
6
5
9
7
1
10
8
22
4
11
18
17
25
20
12
23

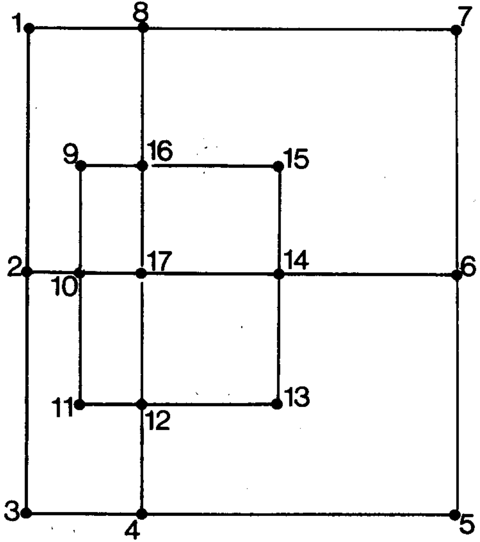

542
552
568
528
511
White
591
599
490
(454,670)
609
472
455
632
653
EXPERIMENT 1
EXPERIMENT 2



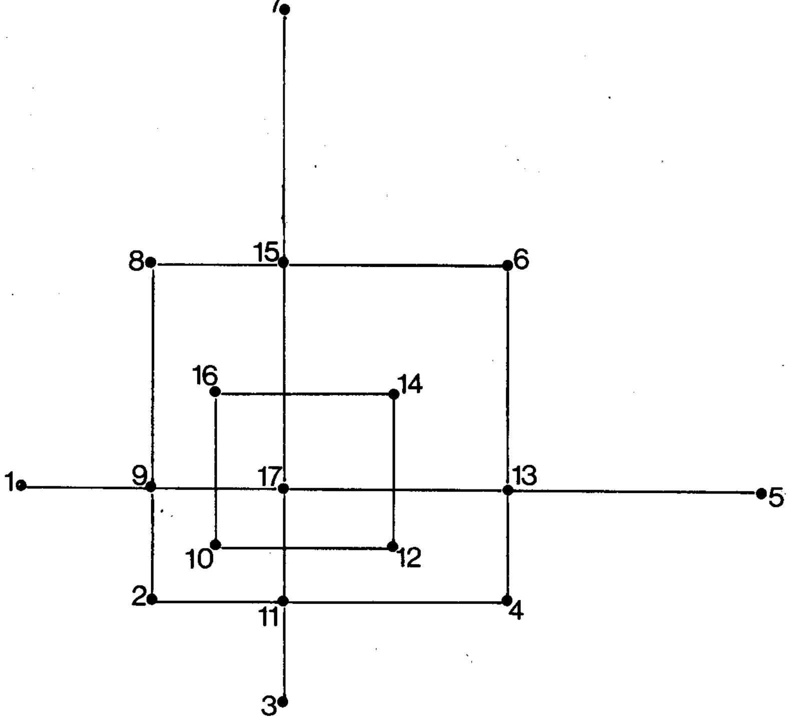
YELLOW
GREEN
555
537
584
504
600
II
490
610
I
628
651
BLUE
472
RED
674
465
445
434
VIOLET

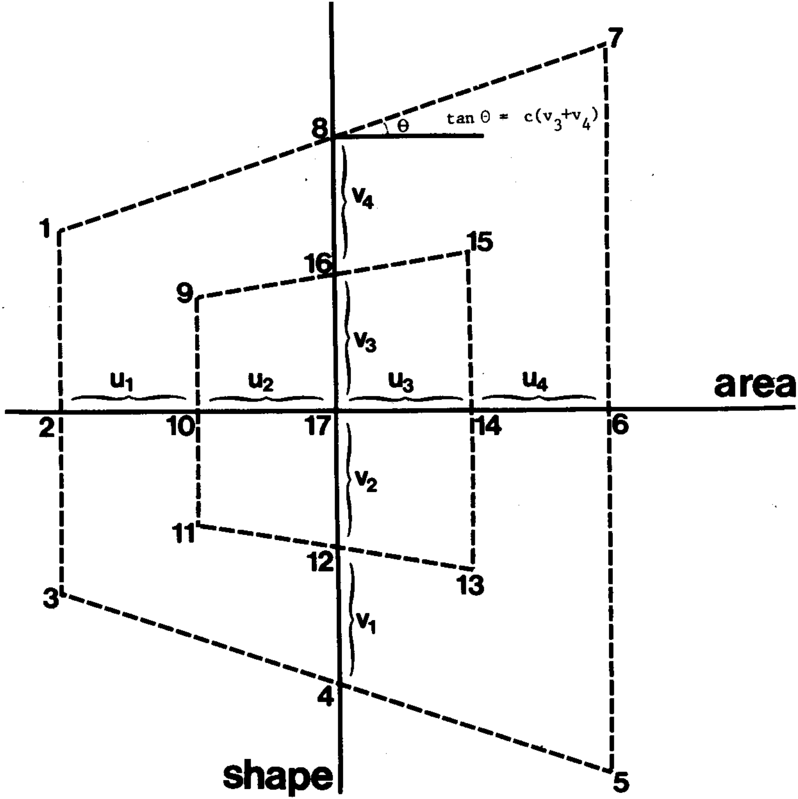

PER CENT "SAME" JUDGMENTS
80
70
60
50
40
30
20
10
0
.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
INTERPOINT DISTANCE


7
8
1
15
16
9
2
10
17
14
6
11
3
12
13
4
5

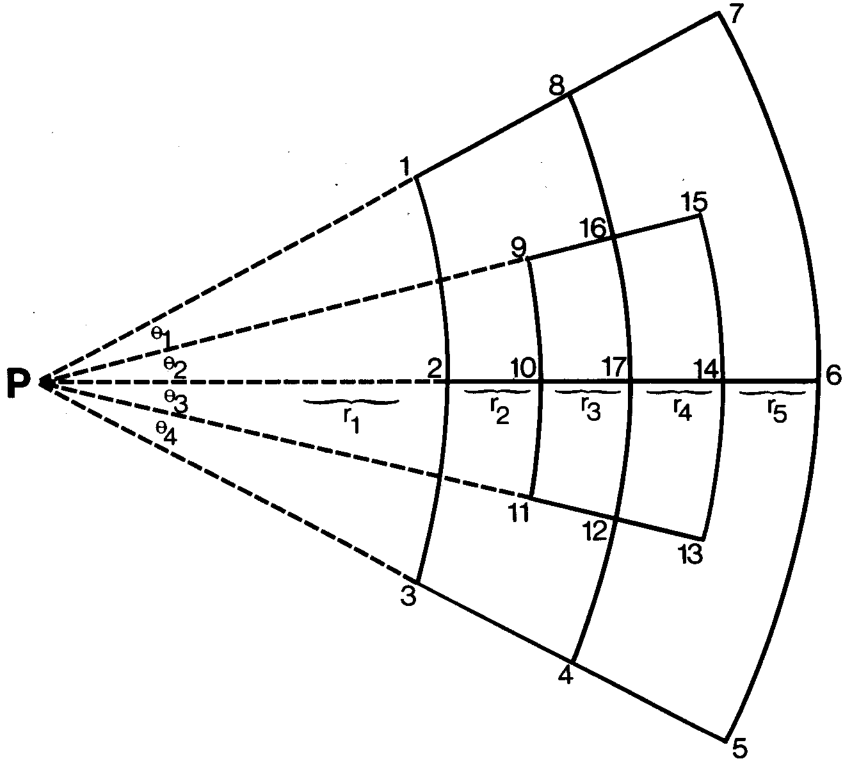


1
8
7
9
16
15
2
10
17
14
6
11
12
13
3
4
5

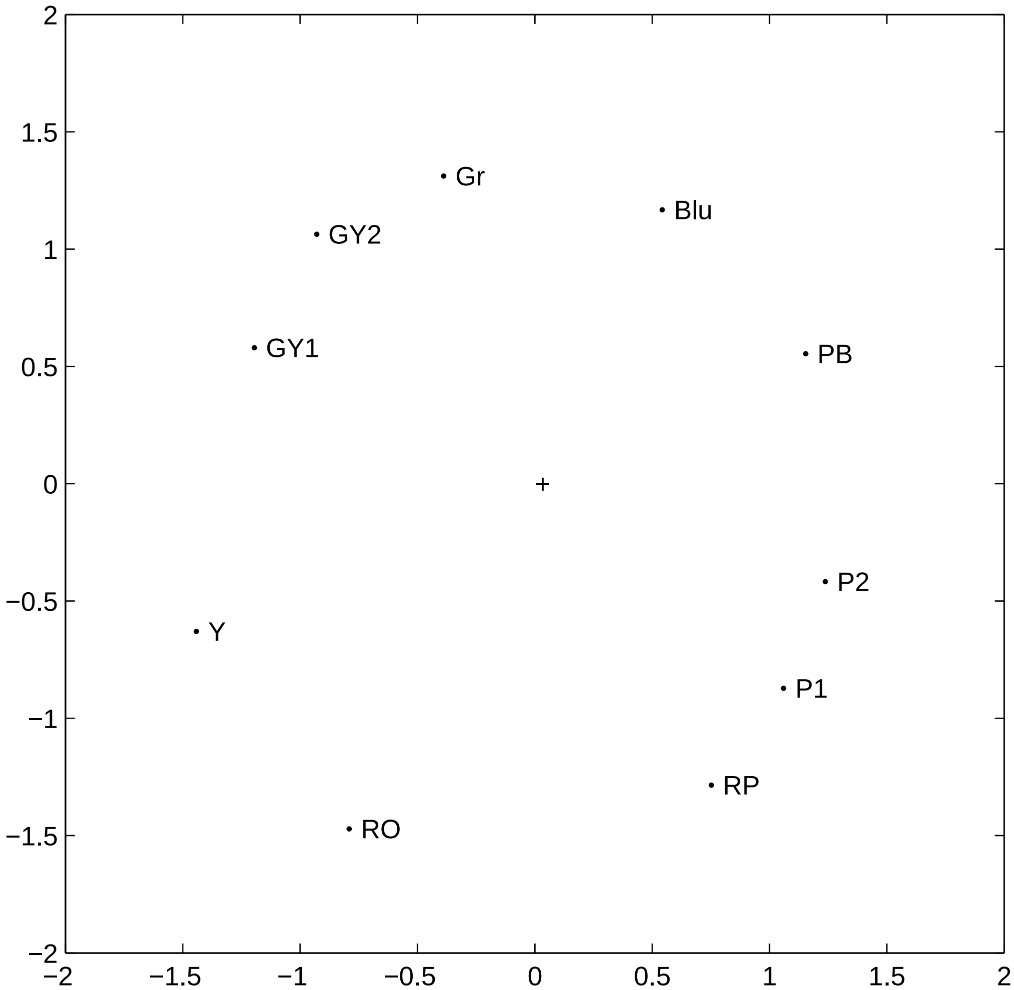

7
15
8
6
16
14
1
9
17
13
5
10
12
2
11
4
3

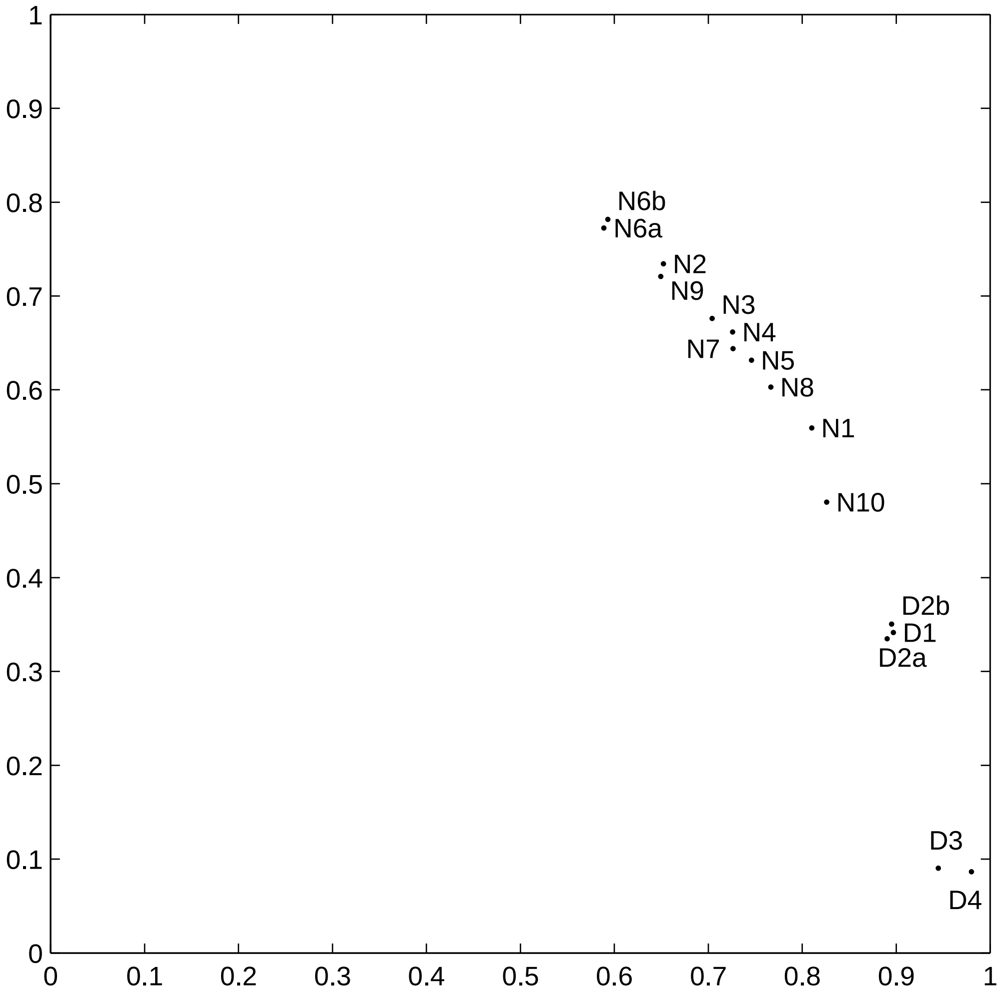


tan Θ = c(v_3+v_4)
Θ
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
v_4
v_3
v_2
v_1
u_1
u_2
u_3
u_4
area
shape

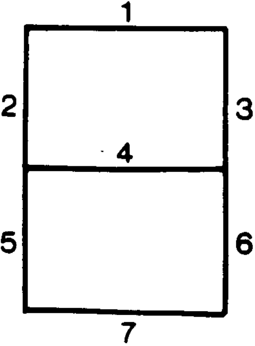


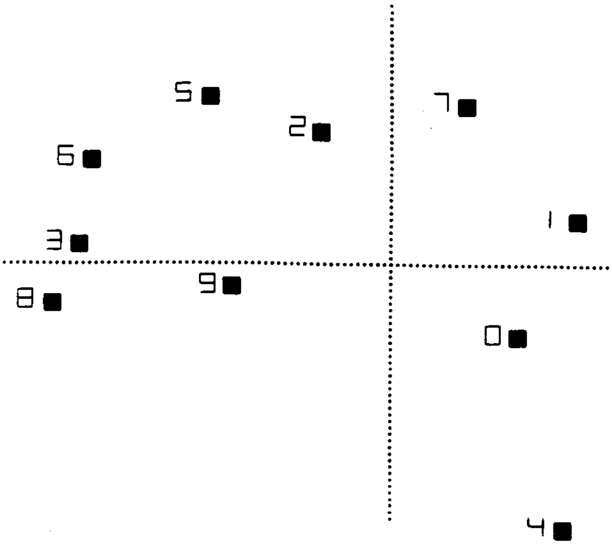
P
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
θ1
θ2
θ3
θ4
r1
r2
r3
r4
r5

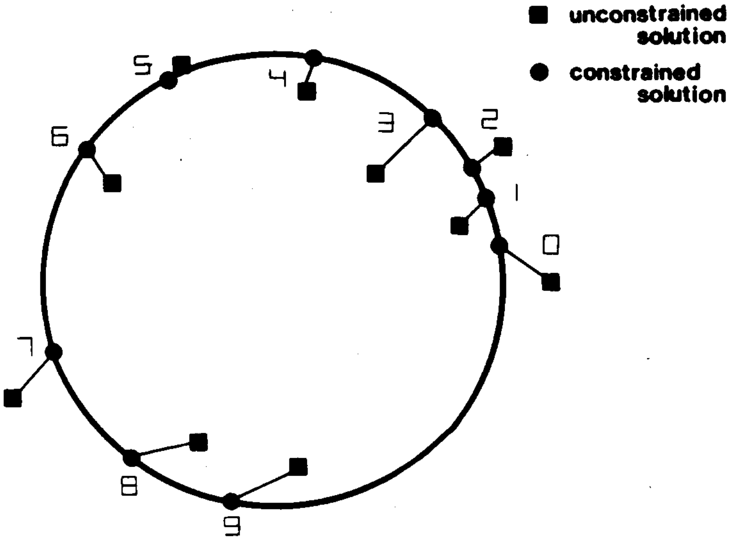

Gr
Blu
GY2
GY1
PB
+
P2
Y
P1
RP
RO
2
1.5
1
0.5
0
−0.5
−1
−1.5
−2
−2
−1.5
−1
−0.5
0
0.5
1
1.5
2

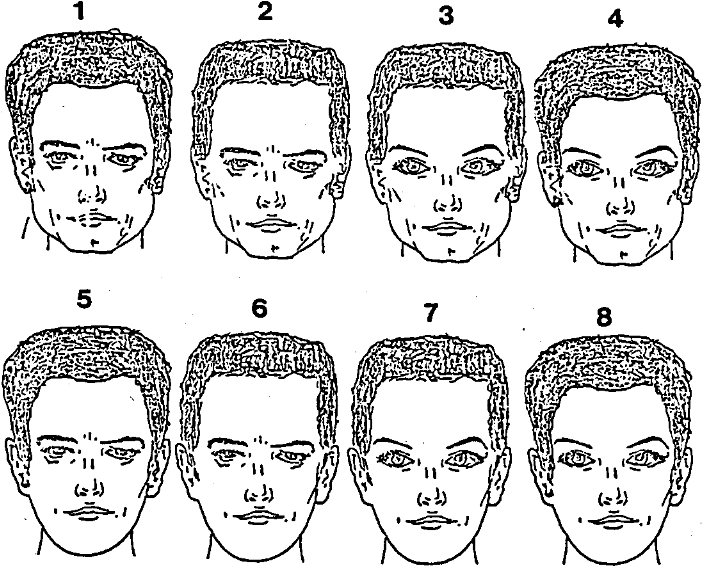


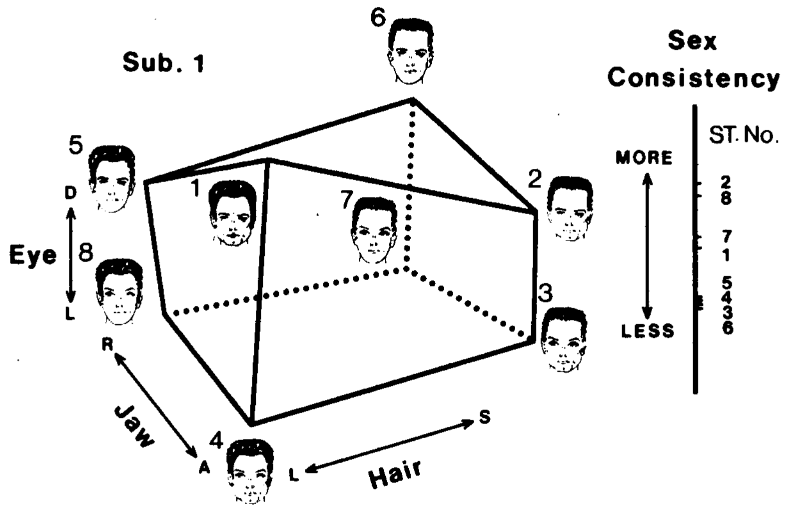
N6b
N6a
N2
N9
N3
N4
N7
N5
N8
N1
N10
D2b
D1
D2a
D3
D4
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1


1
2
3
4
5
6
7

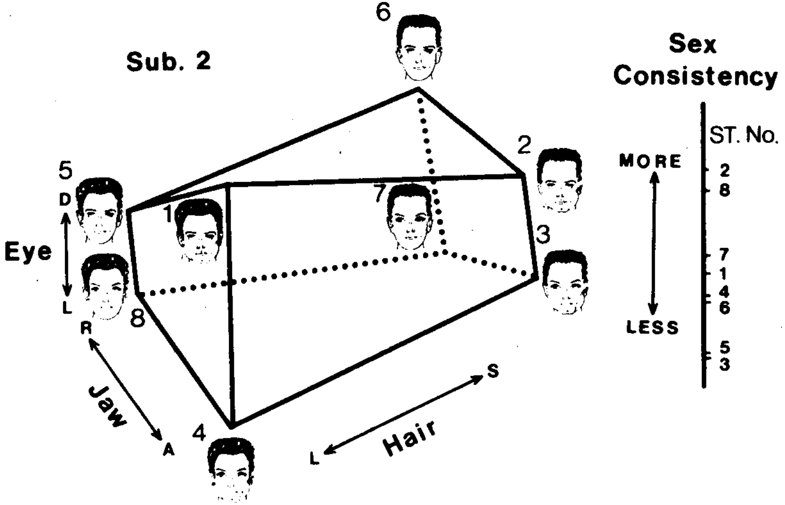

5
2
7
6
1
3
9
8
0
4


unconstrained solution
constrained solution
0
1
2
3
4
5
6
7
8
9

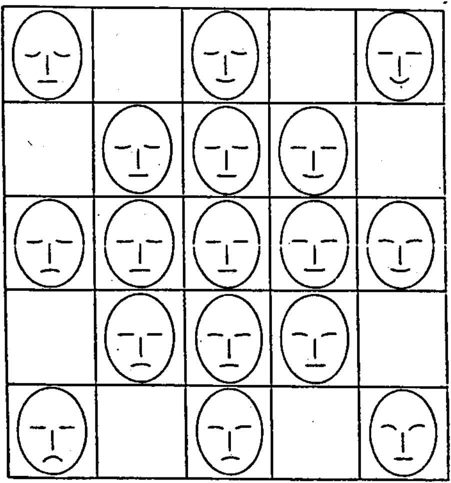

1
2
3
4
5
6
7
8

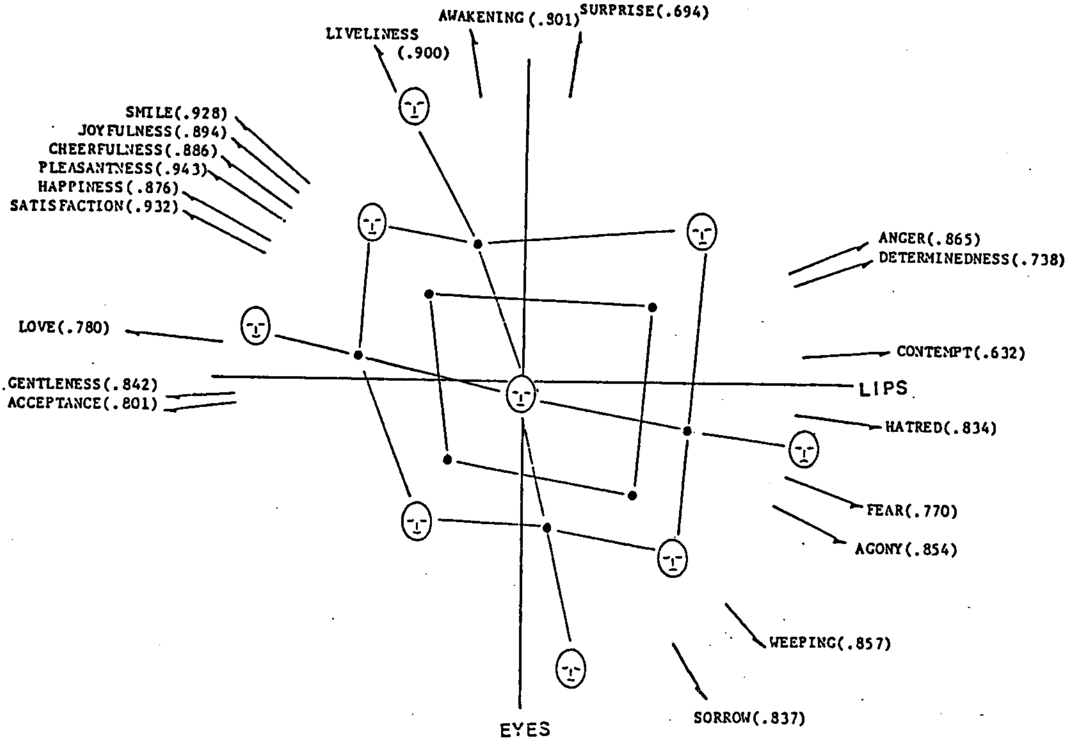


Sub. 1
6
5
1
7
2
3
8
4
D
Eye
L
R
Jaw
A
L
Hair
S
Sex
Consistency
ST. No.
MORE
LESS
2
8
7
1
5
4
3
6

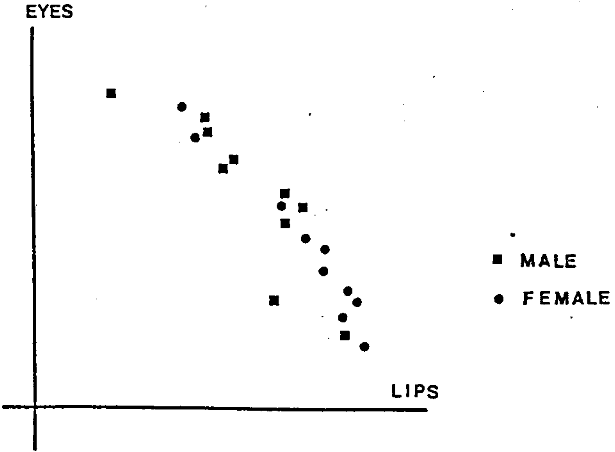


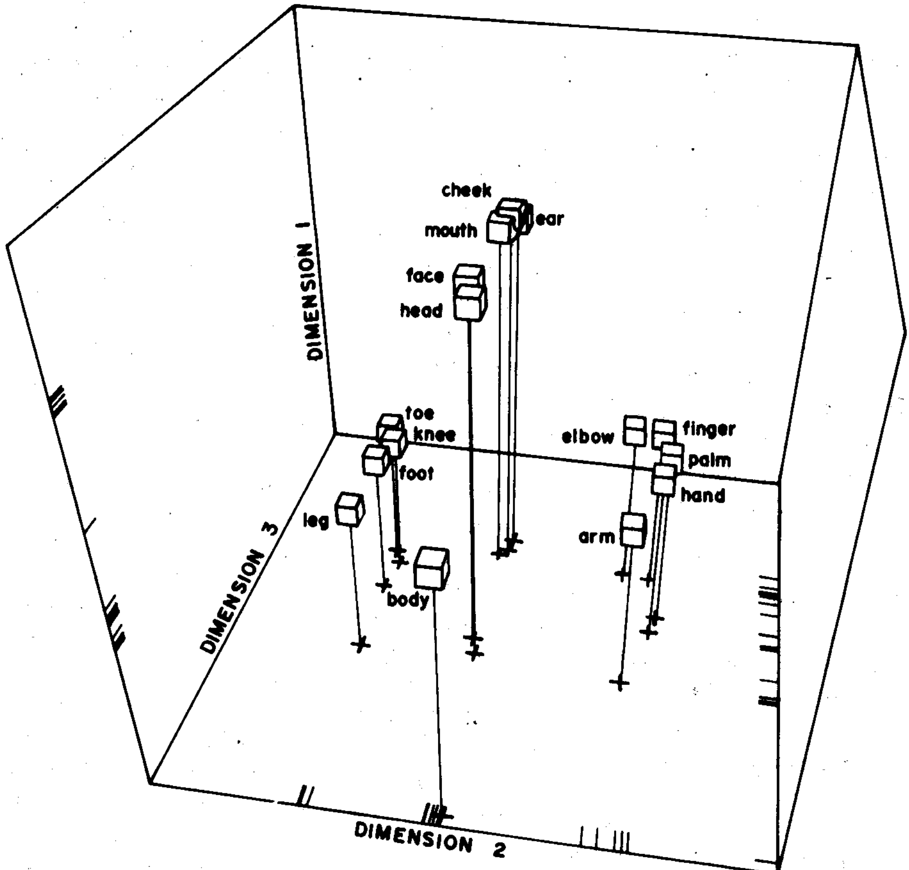
Sub. 2
6
2
7
5
D
1
3
Eye
L
8
R
Jaw
A
4
S
L
Hair
Sex
Consistency
ST. No.
MORE
LESS
2
8
7
1
4
6
5
3



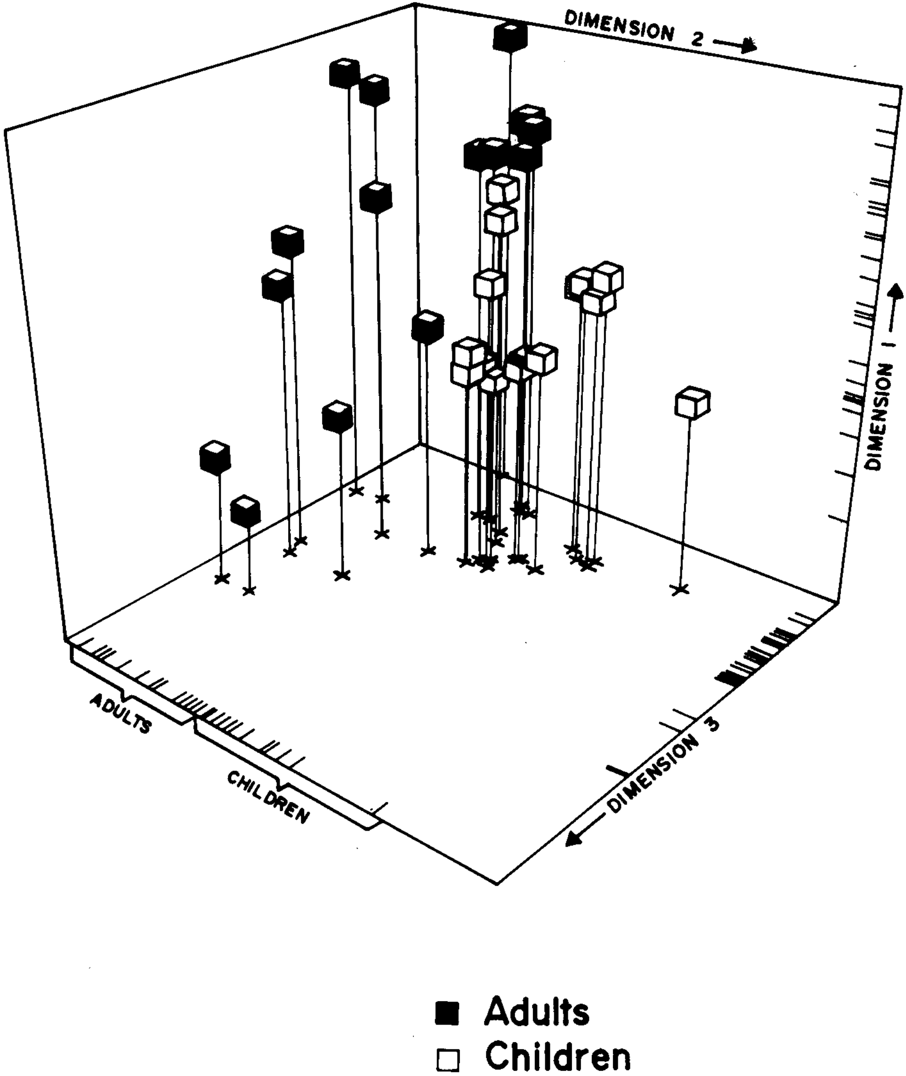


AWAKENING (.801)
SURPRISE(.694)
LIVELINESS (.900)
SMILE(.928)
JOYFULNESS(.894)
CHEERFULNESS(.886)
PLEASANTNESS(.943)
HAPPINESS(.876)
SATISFACTION(.932)
ANGER(.865)
DETERMINEDNESS(.738)
LOVE(.780)
CONTEMPT(.632)
GENTLENESS(.842)
ACCEPTANCE(.801)
LIPS
HATRED(.834)
FEAR(.770)
AGONY(.854)
WEEPING(.857)
SORROW(.837)
EYES

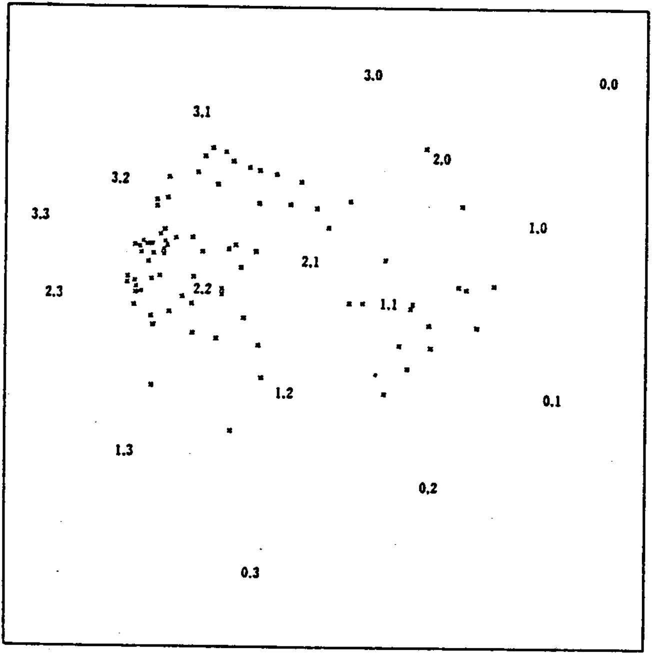

EYES
LIPS
MALE
FEMALE

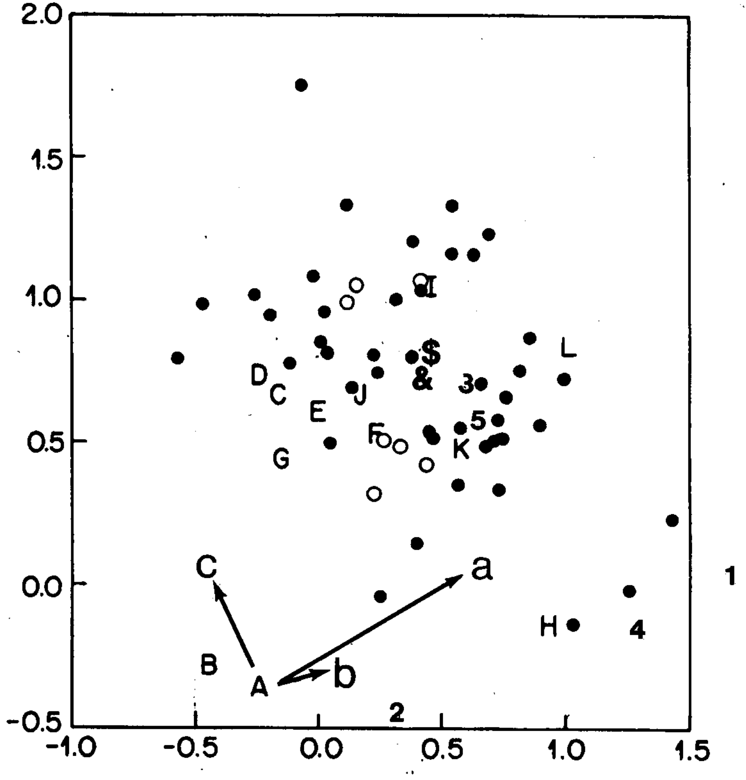


cheek
ear
mouth
face
head
DIMENSION 1
toe
knee
foot
leg
body
elbow
finger
palm
hand
arm
DIMENSION 3
DIMENSION 2

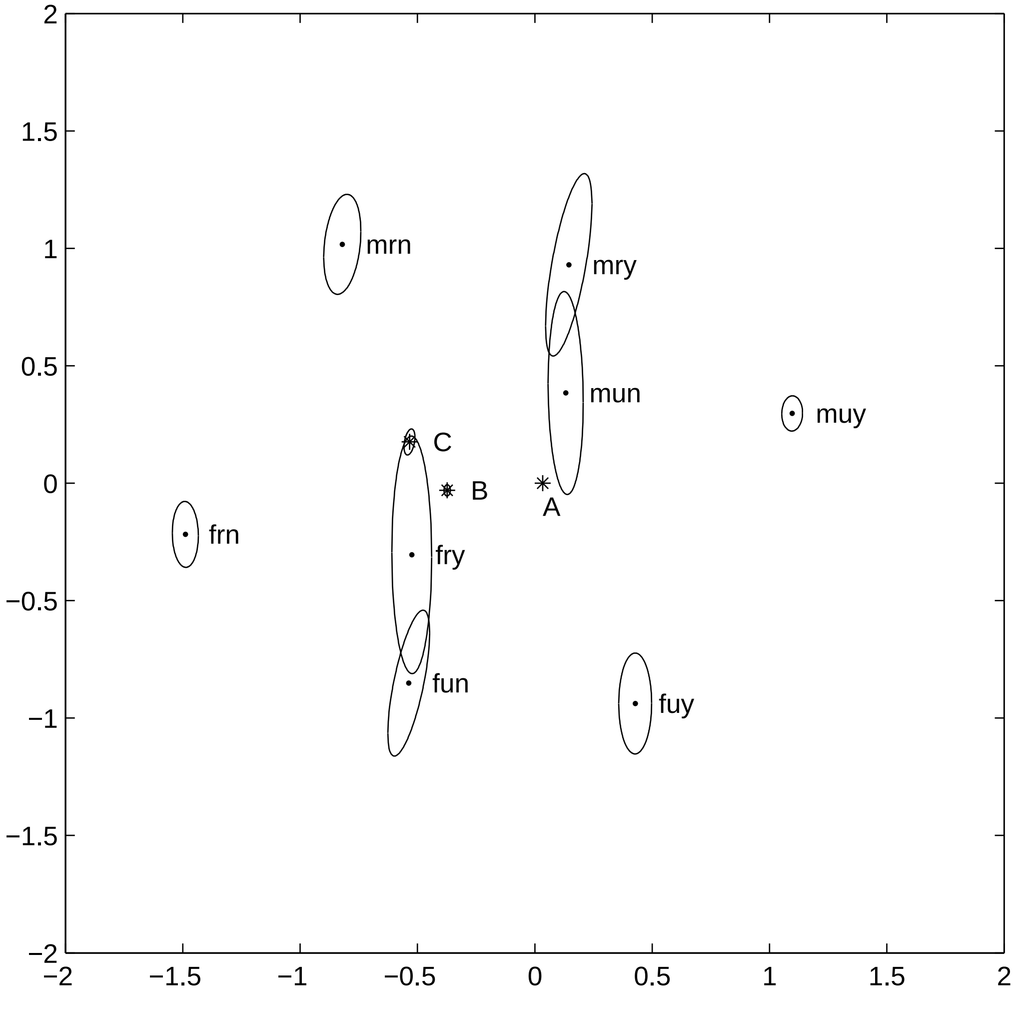

DIMENSION 2
DIMENSION 1
DIMENSION 3
ADULTS
CHILDREN
Adults
Children

NO	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	16	6	8	4	7	1	2	12	5	3	10	13	9	11	14	15
2	16	7	4	5	15	1	3	8	10	9	2	6	14	13	11	12
3	16	14	9	7	15	6	4	2	10	5	1	11	8	3	12	13
4	16	15	13	12	14	8	3	5	11	2	1	7	10	4	6	9
5	16	12	10	13	14	8	2	7	11	6	5	1	15	9	3	4
6	16	7	5	6	15	11	1	2	13	10	4	3	14	9	12	8
7	16	12	11	6	15	5	2	3	14	4	1	8	13	7	9	10
8	16	8	6	5	14	9	4	2	15	11	3	1	13	12	7	10
9	14	15	10	12	16	4	2	5	11	3	1	7	13	6	8	9
10	15	16	13	9	14	10	7	8	11	5	4	2	12	6	1	3
11	16	13	4	3	15	1	2	9	11	8	5	7	12	14	10	6
12	16	14	6	3	15	5	2	4	12	8	1	7	13	9	11	10
13	16	14	9	8	15	4	1	11	10	2	5	3	13	12	6	7
14	16	14	10	11	15	9	5	6	13	8	4	2	12	7	3	1
15	15	14	12	16	13	11	9	10	8	7	5	6	2	3	1	4
16	16	6	2	7	8	1	4	10	3	5	9	11	13	14	12	15
17	16	15	13	10	14	9	8	7	12	5	3	6	11	4	1	2
18	16	8	5	2	15	3	1	10	9	4	6	7	14	11	12	13
19	16	14	12	13	15	10	7	6	11	8	1	3	9	5	4	2
20	15	12	10	14	11	6	2	5	9	1	3	8	13	4	7	16
21	16	15	10	11	13	9	8	5	12	7	1	3	14	6	2	4
22	16	14	11	10	15	9	5	6	13	7	1	3	12	8	4	2
23	16	11	6	12	13	4	1	5	14	2	3	7	15	10	9	8
24	16	15	13	12	14	7	8	10	4	1	5	9	2	3	6	11
25	16	14	10	9	15	8	6	5	12	7	1	2	13	11	4	3
26	16	12	11	10	15	5	6	8	14	7	4	2	13	9	3	1
27	14	15	11	9	16	8	1	5	13	6	2	3	12	7	10	4
28	16	13	12	14	10	1	4	15	5	3	2	8	11	6	7	9
29	16	14	8	7	15	10	6	4	13	9	5	1	12	11	3	2
30	1	2	5	7	4	3	6	10	14	12	8	9	16	15	13	11
31	14	15	10	11	16	6	4	7	12	8	1	3	13	9	5	2
32	16	14	11	9	15	10	2	4	13	3	1	6	12	5	7	8
33	16	4	3	8	6	1	2	10	7	5	9	12	13	11	14	15
34	15	7	6	9	4	1	3	12	5	2	10	14	8	11	13	16
35	16	7	6	5	15	8	3	1	14	9	2	4	13	10	11	12
36	16	2	4	5	11	3	1	6	13	7	8	9	15	14	12	10
37	16	7	8	9	6	3	2	12	4	1	10	14	5	11	13	15
38	16	4	9	10	7	3	2	5	11	6	1	13	12	8	14	15
39	16	12	11	10	15	7	5	8	14	6	1	3	13	9	4	2
40	16	2	3	12	4	1	6	10	5	7	8	9	13	11	14	15
41	8	2	3	10	5	1	4	11	6	7	9	12	13	14	15	16
42	16	9	4	7	12	1	2	8	5	3	6	10	11	13	14	15
43	16	15	13	12	14	8	5	7	10	4	1	3	11	6	2	9
44	16	14	6	5	15	9	7	1	13	10	2	3	12	11	8	4
45	16	10	8	9	12	5	4	2	14	11	1	3	15	13	7	6
46	16	13	1	2	15	3	4	6	11	5	7	8	14	12	9	10
47	1	2	5	9	3	4	7	12	6	8	11	14	10	13	16	15
48	10	3	1	5	4	2	7	11	6	8	12	14	9	13	15	16
49	16	13	4	1	15	6	2	3	14	7	5	8	12	10	9	11
50	16	2	7	9	3	1	5	11	4	6	8	13	12	10	14	15
51	16	15	10	13	14	5	7	8	12	6	2	4	11	9	3	1
52	16	15	12	11	14	10	3	4	13	2	1	6	9	8	5	7
53	16	12	11	9	13	10	4	5	14	6	1	2	15	8	3	7
54	16	12	11	9	15	7	5	6	14	10	1	2	13	8	3	4
55	16	15	13	11	14	4	2	6	12	1	3	8	10	5	7	9
56	16	14	13	11	15	6	2	8	12	4	1	3	10	7	5	9
57	16	15	13	10	14	11	6	8	12	5	3	4	9	7	2	1
58	16	15	13	9	14	11	8	5	12	7	3	2	10	4	1	6
59	16	14	8	7	15	12	6	2	13	10	3	1	11	9	4	5
60	12	16	14	10	15	11	8	6	13	7	4	2	9	5	1	3
61	16	15	13	12	14	9	8	6	11	7	2	4	10	5	3	1
62	16	14	10	13	15	9	7	2	11	8	1	4	12	3	5	6
63	16	14	11	13	15	8	6	5	12	7	4	1	10	9	3	2
64	16	12	5	2	15	11	6	1	14	10	7	3	13	9	8	4
65	16	10	11	13	12	9	7	5	14	8	4	2	15	6	3	1
66	16	10	8	3	15	9	5	2	14	11	6	1	13	12	7	4
67	16	10	12	14	11	9	7	3	13	8	1	2	15	6	4	5
68	15	5	3	6	10	2	1	9	8	7	4	11	12	13	14	16
69	16	15	13	10	14	6	8	9	12	5	1	7	11	3	2	4
70	16	12	8	9	15	10	6	4	14	7	5	2	13	11	3	1
71	6	3	5	9	2	1	8	13	4	7	11	15	10	12	14	16
72	16	12	11	10	15	9	5	7	14	6	4	1	13	8	2	3
73	14	13	8	11	12	2	4	5	15	3	1	9	16	7	6	10
74	16	10	11	14	12	9	1	3	13	5	2	4	15	7	6	8
75	16	14	13	15	12	6	10	8	11	5	1	4	9	7	3	2
76	16	12	10	11	14	4	1	3	13	5	2	6	15	7	8	9
77	16	12	4	3	15	9	5	2	14	10	6	1	13	11	8	7
78	16	10	11	13	15	4	2	6	14	3	1	5	12	7	8	9
79	1	2	3	4	9	5	6	7	14	11	10	8	16	17	13	12
80	16	14	11	10	15	9	7	5	13	8	4	2	12	6	3	1
81	16	15	13	11	14	9	8	6	12	7	4	3	10	5	2	1
82	16	12	8	4	15	11	7	3	14	10	6	2	13	9	5	1

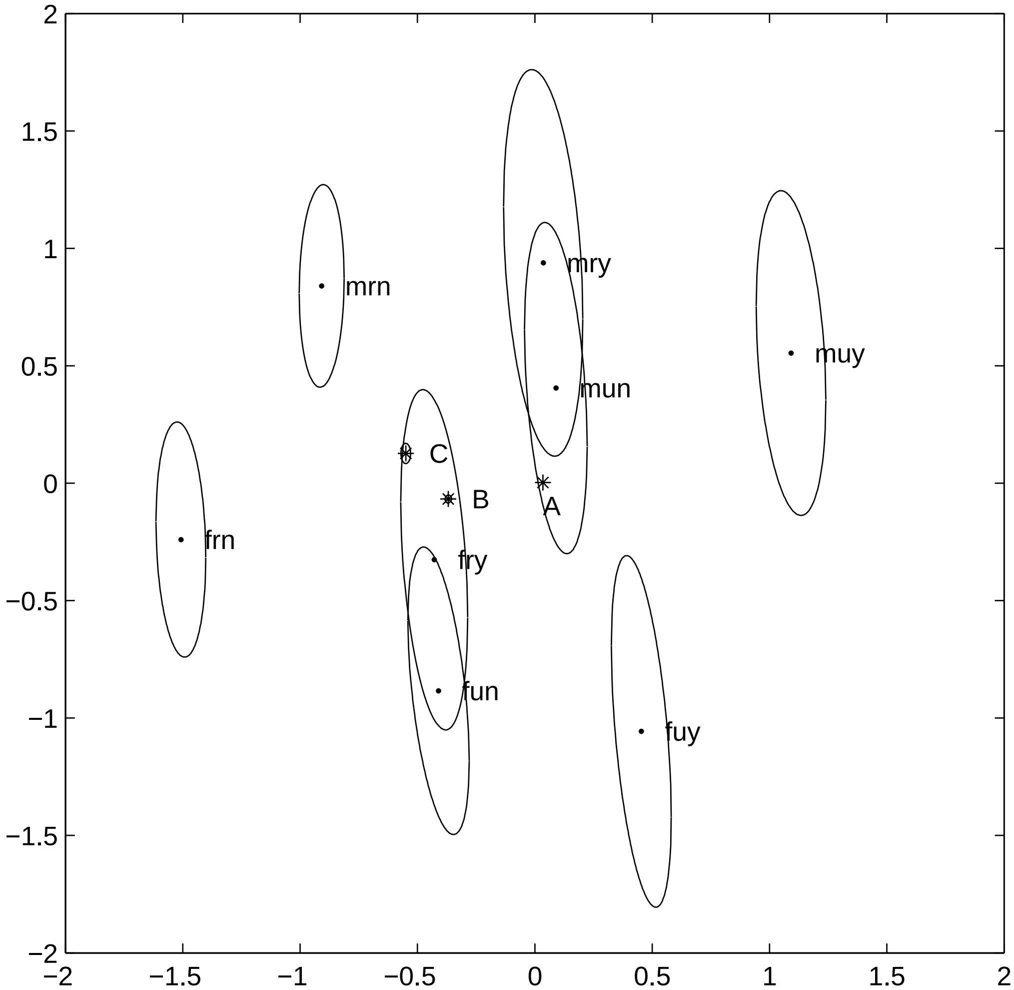

3.0
0.0
3.1
2.0
3.2
3.3
1.0
2.1
2.3
2.2
1.1
1.2
0.1
1.3
0.2
0.3



2.0
1.5
1.0
0.5
0.0
-0.5
-1.0
-0.5
0.0
0.5
1.0
1.5
I
$
L
D
C
&
3
J
E
5
F
K
G
C
a
1
H
4
B
A
b
2



mrn
mry
mun
muy
C
B
A
frn
fry
fun
fuy
2
1.5
1
0.5
0
−0.5
−1
−1.5
−2
−2
−1.5
−1
−0.5
0
0.5
1
1.5
2



2
1.5
1
0.5
0
−0.5
−1
−1.5
−2
−2
−1.5
−1
−0.5
0
0.5
1
1.5
2
mrn
mry
mun
muy
C
B
A
frn
fry
fun
fuy