PART III

Scaling

10

Multidimensional Scaling

Yoshio Takane, Sunho Jung,
and Yuriko Oshima-Takane

INTRODUCTION

The notion of similarity plays a fundamental role in psychology, especially in cognitive psychology. According to Tversky (1977), similarity is an organizing principle by which we categorize, generalize, and classify objects. These activities are crucial for the survival of species. Multidimensional scaling (MDS) is a collection of data analysis techniques for analysis of proximity data. The word 'proximity' here refers to the degree of similarity or dissimilarity among stimuli (objects) of interest. (We use the word 'proximity' as a superordinate term that includes both similarity and dissimilarity.) More specifically, MDS is a class of data analysis techniques that represents a set of stimuli as points in a multidimensional space in such a way that the distances between them best represent the observed proximity data between the stimuli.

To illustrate, let us look at Figure 10.1A. This is the Greek letter ψ. Ten points on this letter were selected arbitrarily and Euclidean distances between them were measured. The measured distances are presented in Table 10.1. Measuring the inter-point distances is straightforward using a rules. But what about the reverse operation? Is it as easy to recover relative locations of the ten points based on the measured inter-point distances? We may use some geometric devices (e.g., a pair of compasses). This, however, is generally a much more difficult task than measuring the inter-point distances. The role of MDS is, roughly speaking, to perform this reverse operation. That is, it recovers the relative locations of points based on a set of inter-point distances. Figure 10.1B presents the ten points on the letter ψ recovered by one of the most basic algorithms for MDS. It can be seen that the relative locations of the ten points are almost perfectly recovered. Note that the recovered configuration is 'flipped' and rotated relative to the original one. This is because MDS uses only the interpoint distance (usually Euclidean) information, which has no information regarding the 'right' orientation of the coordinate axes. (The remaining parts of Figure 10.1 will be discussed later.)

What is the main purpose of MDS? In essence, MDS obtains a graphical display of stimuli (like the one given in Figure 10.1B) based on their proximities (like those given in Table 10.1). The pictorial representation of the stimuli facilitates our understanding

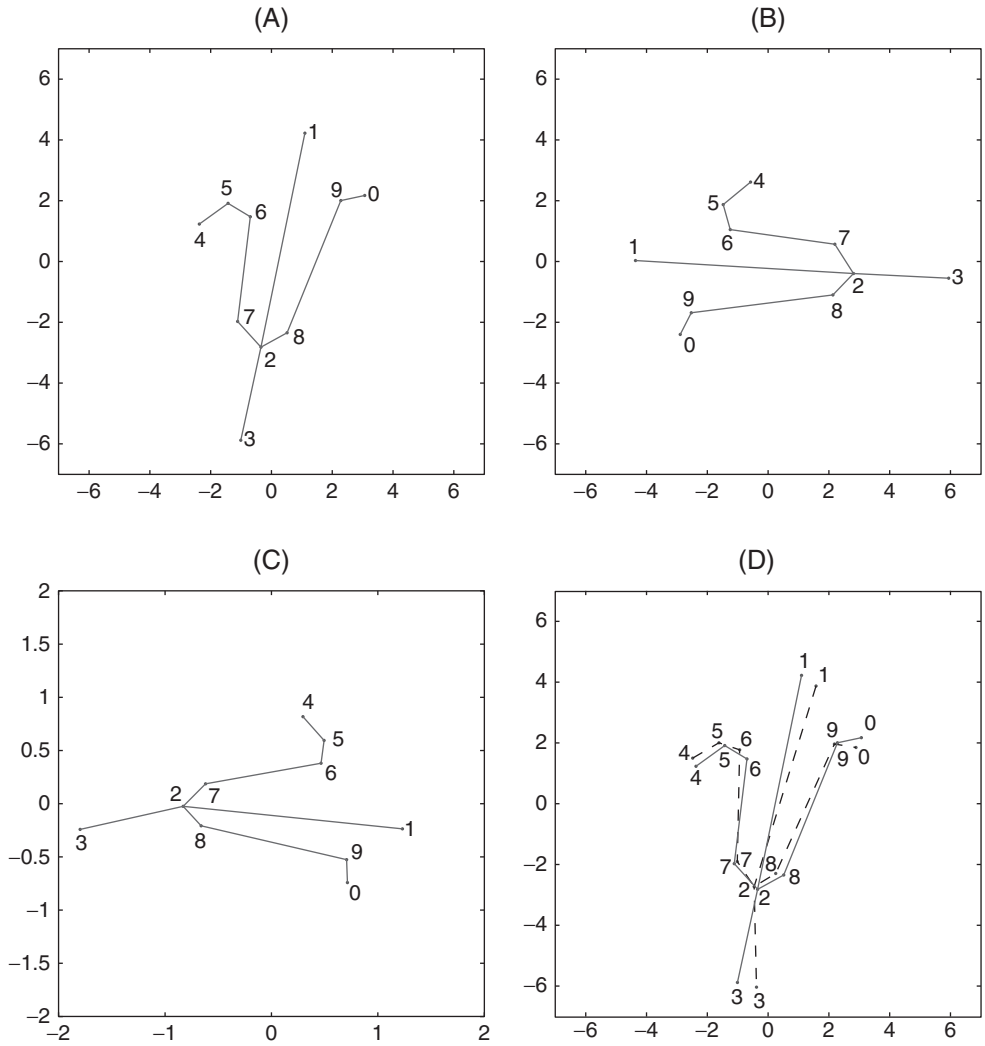

Figure 10.1 Recoveries of ten points sampled on letter ψ. (A) The original configuration. (B) The configuration derived by the exact reverse operation (the Young–Householder transformation followed by the eigenvalue and vector decomposition). (C) The configuration derived from ranked distances by non-metric MDS. (D) Configurations in (B) and (C) rotated into the best agreement with (A).

of the proximity relations between the stimuli. By identifying meaningful directions and/or regions in the space, we may be able to discover organizing principles governing the proximity relations between the stimuli. While this point may not be so clear from the artificial example given above, it will be made clearer in more realistic applications of MDS to be given later.

Does MDS make sense? Or, to be more exact, does it make sense to represent the proximity data by a distance model? In MDS, stimuli are represented by points in a multidimensional space in such a way that the proximity relations between the stimuli are best represented by the distances between the points. This implies that, in order to apply MDS to proximity data, the latter should in some sense behave like distances.

Table 10.1 Distances between 10 sampled points on letter ψ

Label	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
2	7.18								
3	10.32	3.13							
4	4.58	4.53	2.74						
5	3.42	4.85	7.80	1.17					
6	3.28	4.30	7.36	1.70	.81				
7	6.58	1.14	3.91	3.45	3.90	3.47			
8	6.59	.98	3.85	4.60	4.68	4.00	1.68		
9	2.51	5.49	8.54	4.72	3.71	3.02	5.23	4.70	
0	2.84	6.04	9.02	5.52	4.50	3.82	5.89	5.19	.81

The distance is formally defined as any function of two points satisfying three metric axioms: minimality, symmetry, and the triangular inequality. Minimality states that the distance is always non-negative, and is minimal (zero) when two points coincide. Symmetry means that the distance from point A to B is the same as the distance from B to A. The triangular inequality means that the distance from one point to another by way of a third point is never smaller than the straight distance (the shortest path) between the two points. Do proximity data have similar properties? We often observe that: (1) a stimulus is most similar to itself (minimality); (2) if stimulus A is similar to B, stimulus B is also similar to A (symmetry); and (3) if stimuli A and B are similar, and stimuli B and C are similar, then stimuli A and C are also reasonably similar to each other (triangular inequality). That is, in a majority of situations empirical similarity data possess distance-like properties. [however, see Tversky (1977), who presented a number of counter examples]. This means that MDS is a sensible method to apply for analyzing similarity data at least as a first approximation. (In the above, we only referred to similarity data, but essentially the same argument holds for dissimilarity data as well.)

The rest of this chapter is organized as follows. We first discuss several elements that need to be addressed before applying MDS. Specifically, we answer the following questions in the next few sections: Which distance models do we use? How do we collect proximity data? How are proximity data and distance models functionally related? How do we measure the goodness of representation? These methodological sections are followed by examples of the application of MDS: simple MDS, individual differences MDS, and unfolding analysis. Throughout this chapter, technical matters (e.g., optimization algorithms) are kept to a minimum. The reader is referred to Borg and Groenen (2005) for discussions on more technical details.

DISTANCE MODELS

Many distance functions satisfy the three metric axioms. Which distance functions do we use in MDS? In this chapter, we largely limit our attention to the Euclidean distance model and its variants. [see Arabie (1981), and Hubert, Arabie, and Hesson-Mcinnis (1994) for applications of the city-block distance model, the distance model next most often discussed after Euclidian distance.] The Euclidean distance is the most familiar distance function in our everyday life, and consequently allows relatively easy interpretation of stimulus configurations derived by MDS. It is also relatively easy to fit this model compared to other distance functions.

The Euclidean distance model can easily be parameterized in term of the Cartesian coordinate system, which is another attractive feature of this model. Let x_{ir} denote the coordinate of point i on dimension r. Then d_{ij}, the Euclidean distance between the points i and j, is calculated by:

$$d_{ij} = \left\{ \sum_{r=1}^{R} (x_{ir} - x_{jr})^2 \right\}^{1/2} \qquad (1)$$

where R is the dimensionality of the representation space. In MDS, the set of coordinate values $\{x_{ir}\}$ for $i = 1, \cdots, n$ (where n is the number of points) and $r = 1, \cdots, R$ are determined in such a way that the set of d_{ij}s calculated from the x_{ir}s are as close as possible to the observed proximity data.

The Euclidean distance is invariant over rotation of the coordinate axes and over

translation (shift) of the origin. These indeterminacies are often handled by putting the origin at the centroid of the stimulus configuration and by placing the axes in the principal axis directions (a set of orthogonal directions in the space in which the variability in coordinate values is successively largest). These conventions are, however, essentially arbitrary. The stimulus configuration may also be rotated in such a way that the coordinate axes have easier interpretations.

A set of distances may be arranged in matrix form D with d_{ij} as the ijth element (the element in the ith row and the jth column). Matrix D is symmetric, and hollow (the diagonal elements are zero). The stimulus coordinates x_{ir} may also be collected in matrix form denoted by X. This matrix is n by R with x_{ir} representing its irth element.

In many applications of MDS, proximity data are collected from a group of subjects. How are those proximity matrices related? If no systematic individual differences are suspected, a single common Euclidean distance model may be fitted to all of them simultaneously. However, in many situations the assumption of no systematic individual differences is unrealistic. In such a case, each proximity matrix may be analyzed separately, yielding as many stimulus configurations as there are proximity matrices. A natural question is how they are related. In most cases, there are both common and unique aspects in proximity judgments obtained from different individuals. If so, couldn't there be a better way of analyzing the data?

The individual differences (ID) MDS model we discuss in this chapter is designed to partially answer the above question. It captures both commonality and individual differences in a unified framework (Carroll and Chang, 1970). More specifically, it postulates a common stimulus configuration, but that dimensions in the common configuration are differentially weighted by different individuals to give rise to differences in proximity data by different individuals. To illustrate, let us look at Figure 10.2, where the same letter ψ as in Figure 10.1 is displayed. The letter ψ depicted in Figure 10.2A may be perceived differently by different individuals. For example, subject 1 may perceive it as the dashed ψ in Figure 10.2B, subject 2 as the solid ψ in Figure 10.2B, subject 3 as depicted in Figure 10.2C, and subject 4 as depicted in Figure 10.2D. These configurations are all related by differential weighting of dimensions, uniform contraction or dilatation (10.2B), vertical elongation (10.2C), and horizontal elongation (10.2D). The particular ID MDS technique we discuss in this chapter assumes these kinds of relationships among the stimulus configurations obtained from different individuals.

The idea of differential weighting of dimensions in a common stimulus configuration can be captured by the weighted Euclidean distance model written as:

$$d_{ijk} = \left\{ \sum_{r=1}^{R} w_{kr}(x_{ir} - x_{jr})^2 \right\}^{1/2} \qquad (2)$$

where d_{ijk} is the distance between stimuli i and j for individual k, x_{ir} is the coordinate of stimulus i on dimension r in the common stimulus configuration, and w_{kr} is the weight attached to dimension r by subject k. Uniform contraction can be captured by weights smaller than one across all dimensions, uniform stretching by weights uniformly larger than one, vertical elongation by a weight for the vertical dimension larger than one, horizontal elongation by a weight for the horizontal dimension larger than one, and so on. The individual difference weights may be arranged in a K by R matrix W, where K is the total number of individuals.

To eliminate the size indeterminacy between the stimulus configuration and the individual difference weights, the former is typically constrained to satisfy $\sum_{i=1}^{n} x_{ir}^2/n = 1$ for $r = 1, \cdots, R$. In contrast to the simple Euclidean distance model, the orientation of the coordinate axes is uniquely determined (except for reflection and permutation) in the weighted Euclidean model.

Individual differences are much more prevalent in preference judgments. Preference data are often analyzed by a variant of MDS called unfolding analysis (Coombs, 1964).

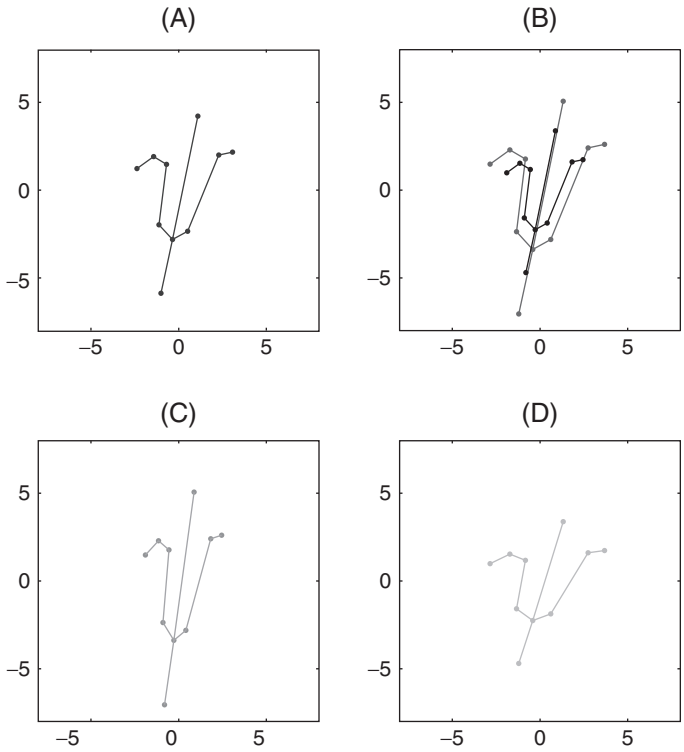

Figure 10.2 Letter ψ dimensionwise differentially weighted. (A) The original configuration. (B) Uniformly dialated (dashed) and contracted (solid) configurations. (C) A vertically elongated configuration. (D) A horizontally elongated configuration.

In unfolding analysis, each subject is assumed to have an ideal stimulus represented as the subject's ideal point in the multidimensional space in which actual stimuli are also represented as points. The distances between the ideal point and the stimulus points are assumed inversely related to the subject's preferences on the stimuli. Let x_{ir} denote the coordinate of stimulus i on dimension r, and y_{jr} the coordinate of subject j's ideal point on dimension r. The Euclidean distance between stimulus point i and ideal point j is calculated by:

$$d_{ij} = \left\{ \sum_{r=1}^{R} (x_{ir} - y_{jr})^2 \right\}^{1/2} \qquad (3)$$

The coordinates of the ideal and stimulus points are determined in such a way that the preference values of stimuli for a particular individual are a decreasing function of the distances between the stimulus points and his ideal point. This implies that the closer a stimulus point is to his ideal, the more it is preferred by that subject. The preference relations are thus regarded as representing proximity relations between the subjects' ideal stimuli and actual stimuli. 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 stimulus points (an n by R matrix X with x_{ir} as the irth element), and the other for subjects' ideal points (an N by R matrix Y with y_{jr} as the jrth element).

DATA COLLECTION METHODS

There are a number of different ways of collecting proximity data. In this section, we

discuss some that are often used in MDS. The methods can roughly be classified into two groups. One involves direct judgments of (dis)similarity, and the other involves indirect judgments. In the latter, the investigator typically asks the subjects to do a certain task (e.g., discrimination between stimuli), and how well (or badly) the subjects do in the task is taken as a (dis)similarity measure between stimuli.

Direct judgments

A majority of data analyzed by MDS use direct judgments of (dis) similarity. There are several variants in this category.

Rating

The most straightforward method is to ask the subject to rate the degree of (dis)similarity between two stimuli at a time on a rating scale. It is preferable to have as many categories as possible in the rating scale, considering the statistical efficiency of judgments. Consider, as an example, a brand manager who wants to collect dissimilarity data about 10 different brands of chocolate using this method. He may ask the subjects to record their overall impression of the degree to which each possible pair of chocolates differ on 10-point scale (i.e., 1 = extremely similar, and 10 = extremely dissimilar). Most of the example data sets analyzed in this chapter were collected by this method.

Multiple ratio judgments

In addition to a set of experimental stimuli (i.e., stimuli of direct interest to the investigator), a reference stimulus, which is not among the experimental stimuli, is prepared. The investigator chooses a stimulus as a standard stimulus from the set of experimental stimuli, and indicates the dissimilarity between the standard and the reference stimuli by a physical distance. (He places the two stimuli at a certain distance apart and tells the subject that the physical distance between them represents the dissimilarity between them.) The subject is asked to judge dissimilarities between the standard and all other experimental stimuli in terms of the distance between the standard and the reference stimuli. Once all dissimilarity judgments are obtained for a fixed standard, the standard stimulus is replaced, and the whole process is repeated until all stimuli in the experimental set have served as a standard. Inukai (1981) used this method to collect dissimilarity judgments on facial expressions constructed by varying the curvature of lips and eyes systematically [see Example 9 in Takane (2007)].

Rank order

The subject is asked to rank the dissimilarities between stimuli. The most representative method in this class is the method of conditional rank orders. In this method one of the stimuli serves as a standard stimulus at a time. The subject is asked to pick a stimulus among the remaining stimuli that is most (dis)similar to the standard, and after this stimulus is excluded from the comparison set, to pick the next most (dis)similar stimulus, and so on until a complete rank ordering of (dis)similarity is obtained for a fixed standard. Then, the standard stimulus is switched and the same procedure is followed until all stimuli have served as a standard. The body parts data analyzed by Takane, Young, and de Leeuw (1977) were collected by this method [see also Example 10 in Takane (2007)].

Sorting

A group of subjects is given a set of stimuli and asked to sort them into as many categories as they want in terms of (dis)similarity among them, so that the stimuli within the same group are more similar to each other than those classified into different groups. In a fairly standard format, the investigator prepares a deck of 3-inch by 5-inch index cards with stimuli printed on the cards. The subjects are asked to sort them into several piles. The frequency of two stimuli bring classified into the same category is most commonly used as a similarity measure between the stimuli. The sorting method is very easy to use, particularly when the number of stimuli involved is very large. Because of its simplicity, it is a very

popular method in social sciences. Dunn-Rankin and Leton (1975) used this method to collect similarity data on 46 Japanese Kana characters (phonetic symbols). Results of MDS are reported as Example 2 in Takane (2007). Takane (1980) also developed a special MDS technique that is specifically designed to analyze sorting data. He applied the method to 29 have words (e.g., 'belong', 'lose', etc.) sorted according to the similarity of their meaning by 10 university students [see also Example 3 in Takane (2007)]. Later in this chapter (see the section 'Example 2. animals', p.000), we present an example of MDS analysis of sorting data.

Indirect methods

Confusion data

Clearly, the more similar two stimuli are, the more confusable they are (the higher the probability that the two stimuli will be confused). Stimuli are presented in pairs and the subject is asked to judge whether the two stimuli are the *same* or *different*. (This is called a *same–different* judgment.) The proportion of the *same* judgments when a pair of different stimuli are presented is taken as a measure of similarity. Typically, equal numbers of same and different pairs are presented not to bias subjects' responses in one way or the other. Rothkopf (1957) obtained a confusion matrix between 36 Morse code signals [see Shepard (1963) for an analysis of Rothkopf's data by MDS]. Schneider (1972) collected confusion data from pigeons, who were trained to discriminate between two simultaneously presented colors (to peck the left lever when two colors were the same, and the right lever when they were different), and analyzed the data by MDS [see Examples 1 and 4 in Takane (2007)].

Another form of confusion data is called *stimulus identification* (or recognition) data. The subject is presented with one stimulus at a time out of n possible stimuli, and is asked to tell which of the n stimuli is presented. The number (or the proportion) of times the presented stimulus is misjudged as another stimulus is used as a similarity measure between the two [see Takane and Shibayama (1986, 1992) for examples]. The stimulus identification data are typically asymmetric. There are special MDS methods specifically designed to analyze this type of data, incorporating bias parameters to account for the asymmetry. When a general-purpose MDS program is used, the data are usually symmetrized by taking averages of the corresponding elements.

Frequency of co-occurrences

The sorting data described above may be viewed as a special case of co-occurrence frequency data. An example is the frequency with which two personality traits are used to describe the same person. The more frequently two traits are used to describe the same person, the more similar they are. In the same way, two individuals who share more personality traits in common are more similar to each other than those with fewer traits in common.

Response latency (reaction time)

When two stimuli are similar, it takes more time to discriminate between them. Thus, the time required to tell the difference between them may be used as a similarity measure. Reaction time is usually measured in the context of 'same–different' judgments described above. Two stimuli are presented at a time, and subjects are instructed to judge whether the stimuli presented are the 'same' or 'different' as quickly as possible. Reaction time data are usually very variable, and quite a large number of replicated observations are necessary to obtain a reliable stimulus configuration by MDS. Takane and Sergent (1983) used reaction time data for MDS of line drawings of faces. Takane (1994) also used this type of data for MDS of digits [see Examples 7 and 8 in Takane (2007)].

Social interaction

The frequency of social interactions (e.g., the number of times two persons have dinner together) may be used as an indication of the degree of intimacy in the relationship. The degree of intimacy may be analyzed by MDS

to derive an intimacy map for a group of people.

Profile dissimilarity

Sometimes it happens that stimuli are rated on a number of attributes (multivariate or profile data). Then profile dissimilarity, defined by

$$o_{ij} = \left\{ \sum_{p=1}^{J} (z_{ip} - z_{jp})^2 \right\}^{1/2} \quad (4)$$

where z_{ip} is the value of stimulus i on attribute p, may be used as a dissimilarity measure between stimuli i and j. The idea is that if the profiles on various attributes are similar, the stimuli must be similar overall. Alternatively, a correlation coefficient between two stimuli over the set of attributes may be calculated and used as a similarity index between the stimuli.

Different (dis)similarity measures may represent different aspects of similarity relations among stimuli. Thus, it is possible to obtain somewhat different representations of the same set of stimuli if different data collection methods are used.

The observed proximity (similarity or dissimilarity) between stimuli i and j is denoted by o_{ij}. The o_{ij}s may be placed in matrix O. This matrix is n by n, and has o_{ij} as its ijth element. It is usually symmetric, and hollow as the matrix of distances D. If it is initially asymmetric, it is often symmetrized by $(O + O')/2$, where O' indicates the transpose of O. When there are more than one replicated observation for each pair of stimuli, we add a third subscript k to o_{ij}. Thus, o_{ijk} denotes the observed proximity between stimuli i and j in replication k. The matrix of o_{ijk} is denoted by O_k.

We discuss a couple more kinds of proximity data. So far proximity data have been defined between stimuli within one set. That is, a set of stimuli of interest is specified, and proximity data are obtained for pairs of stimuli drawn from this set. Proximity data may be defined between 'stimuli' drawn from two distinct sets. For example, suppose a group of people are responding to a set of questions in an opinion survey. In this case, there are a set of respondents and a set of question items. One element each is drawn at a time from these two sets, and a proximity relation (the degree of agreeableness to an item) is observed between them. This kind of proximity data may be analyzed by unfolding analysis as described in the previous section.

Preference data

A group of people give preference judgments on a set of stimuli. Preference data are viewed as indicating similarities between respondents' ideal stimuli and actual stimuli.

Contingency tables

Entries in a contingency table indicate frequencies of joint occurrences of row and column categories, which may be considered as representing similarities between them. This type of proximity data have been traditionally analyzed by a technique called correspondence analysis (Greenacre, 1984; Hwang, Tomiuk, and Takane, Chapter 11; Nishisato, 1980). Takane (1987) and collaborators (Takane, Bozdogan, and Shibayama, 1987; van der Heijden, Mooijaart, and Takane, 1994) developed a maximum likelihood MDS technique called ideal point discriminant analysis (IPDA) specifically designed to analyze this kind of proximity data.

SCALE LEVELS OF MEASUREMENT

As noted above, a variety of proximity measures can potentially be used in MDS. These measures differ not only in their appearance, but also in the type of functional relationships they have with underlying distances. The method of multiple-ratio judgments is intended to collect dissimilarity data that are linearly related to the distances, although whether the subjects can meet the demand is an empirical question. In some cases, an explicit analytic function can be postulated, which relates distances to observed proximity data. These cases are rather rare, however, and in most cases we may assume that the proximity data are only approximately monotonically related to the underlying distances. In some

cases, the data measure similarity rather than dissimilarity. In such cases, the data have to be transformed to make them more directly (linearly) related to the underlying distances. The transformation of the data may be performed either before MDS is conducted if an appropriate transformation is known in advance, or may be done within the MDS algorithm.

Approximate functional relationships between observed data and models are called scale levels of measurement. Five scale levels are traditionally distinguished in psychological literature: ratio, interval, log-interval, ordinal, and nominal, of which only the first four types are relevant in MDS. When the dissimilarity data are roughly proportional to underlying distances, i.e., $o_{ij} \approx ad_{ij}$, where a is a positive constant, we say that the observed data are measured on a ratio scale. This type of relationship between distances (d_{ij}) and dissimilarities (o_{ij}) is depicted in Figure 10.3A. It is linear and passes through the origin. (This is called the "similarity' transformation in mathematics.) In the ratio-scaled measurement, there is an intrinsic origin (the 0 point), so that the ratio of two numbers is meaningful. As a can be absorbed by the size of stimulus configuration, we may assume without loss of generality that it is unity, and we can directly fit the distances to observed dissimilarity data in this case. However, it is rare to find dissimilarity data measured on a ratio scale.

When the dissimilarity data are approximately linear but the zero distance does not correspond to zero dissimilarity, i.e., $o_{ij} \approx ad_{ij} + b$ for nonzero b, we say that the data are measured on an interval scale. This is similar to the ratio scale above, but the function that relates distances to dissimilarity does not pass through the origin, as depicted in Figure 10.3B, where b is assumed positive. (This type of transformation is called an affine transformation.) In the interval-scaled measurement the ratio of two numbers cannot be meaningfully interpreted due to an arbitrary origin, although the ratio of the differences between two numbers is meaningful. The difference effectively cancels out the effect of an arbitrary origin. When $b = 0$, this case reduces to the ratio-scaled measurement. Some classical methods of collecting dissimilarity data (e.g., the method of tetrads, not mentioned in the previous section) are believed to provide interval-scaled dissimilarity data after appropriate scaling of pair comparison judgments. However, these methods are often very time consuming, and have rarely been used in practice.

When the observed dissimilarity data and the underlying distances are approximately related by a power transformation, i.e., $o_{ij} \approx ad_{ij}^b$, we say that that the data are measured on an log-interval scale. This type of functional relationship is depicted for $b < 1$ in Figure 10.3C, which is a negatively accelerated monotonic function. (If $b > 1$, the power transformation is positively accelerated.) A power transformation reduces to an affine transformation, if the log is taken of both sides of the equation (i.e., $\ln o_{ij} \approx b \ln d_{ij} + \ln a$), thus the name log-interval scale. When $b = 1$, this transformation reduces to a similarity transformation. Rating data often satisfy this level of measurement scale.

When the observed data are only monotonically related to underlying distances, i.e., $o_{ij} \geq o_{i'j'}$ implies $d_{ij} \geq d_{i'j'}$, or the observed data are inversely monotonically related to distances, i.e., $o_{ij} \geq o_{i'j'}$ implies $d_{ij} \leq d_{i'j'}$, we say that the data are measured on an ordinal scale. (The monotonic relationships between o_{ij} and d_{ij} cannot be expressed by an explicit analytic function.) In the former case, we have dissimilarity data, while in the latter we have similarity data. The case of ordinal dissimilarity data is depicted in Figure 10.3D, and the case of ordinal similarity data in Figure 10.3E. Monotonic functions are sometimes called order-preserving transformations. When the observed data are measured on an ordinal scale, an MDS algorithm has to be able to find the best monotonic transformation of the data as well as the stimulus configuration that best fits to the monotonically transformed data.

The four types of scale are hierarchically organized. The ratio scale is a special case of both interval ($b = 0$) and log-interval

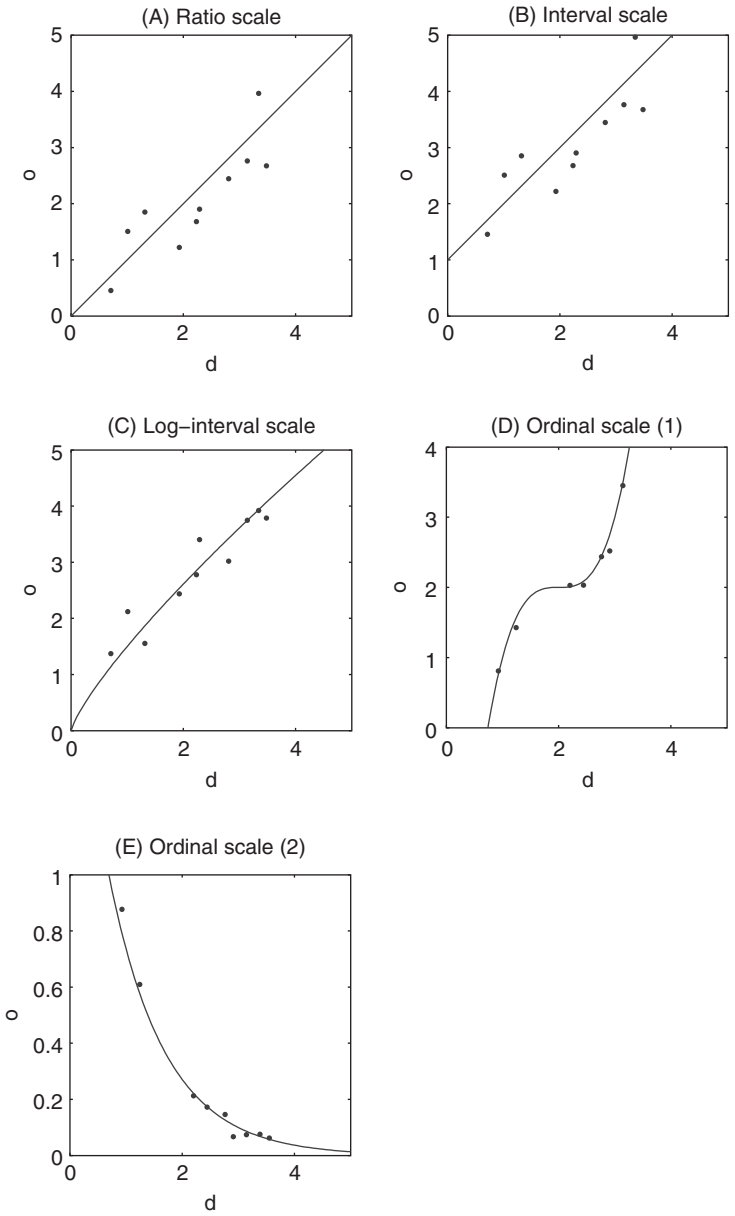

Figure 10.3 Four scale levels of measurement. (A) Ratio scale ($b = 0$). (B) Interval scale ($b = 1$). (C) Log-interval scale ($a = .8$, $b = 1.2$). (D) Ordinal scale (monotonically increasing), indicating that *o* is dissimilarity. (E) Ordinal scale (monotonically decreasing), indicating that *o* is similarity.

($b = 1$) scales. All the transformations discussed are monotonic including the ratio, interval, and log-interval scales. The ratio scale is the most stringent, while the ordinal scale is the least stringent and most flexible. There is a trade-off between a more stringent and a less stringent scale level assumption in MDS. While the less stringent assumption tends to produce estimates of parameters that are less biased, but with larger variances, the more stringent assumption tends to produce just the opposite. It is recommended that

one starts with a less stringent assumption, but once an explicit functional relationship is found, one may switch to a more stringent assumption by incorporating the explicit analytic form.

The distinctions among the different scales is important as certain MDS procedures are appropriate only for data collected on some scales. MDS procedures which assume either a ratio, interval, or log-interval scale are called metric MDS. Others that assume only an ordinal scale are called nonmetric MDS (Kruskal, 1964a, 1964b; Shepard, 1962). In nonmetric MDS, a stimulus configuration is determined in such a way that the rank order of distances between stimulus points best agrees with the rank order of observed dissimilarities. Nonmetric MDS is found to be useful in psychology and related fields, where proximity data based on ordinal measures are prevalent.

To give an indication of how it is possible to derive a stimulus configuration based on the ordinal information about distances alone, let us look at Table 10.2, which displays ranked distances from Table 10.1. The tabled numbers are thus only monotonically related to the underlying distances, and are considered dissimilarity data measured on an ordinal scale. A nonmetric MDS procedure was applied to this data set. The derived stimulus configuration is depicted in Figure 10.1C. The derived configuration was then rotated to match the original configuration as much as possible in Figure 10.1D. Although there are some distortions in the derived configuration due to the loss of information, it can be observed that the original configuration is recovered remarkably well from rank-ordered distances. This indicates that ordinal information is often sufficient to recover a stimulus configuration.

Table 10.2 Rank-ordered interpoint distances between 10 sampled points on letter *ψ*

No.	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
2	39								
3	45	11							
4	25	24	40						
5	13	34	42	5					
6	12	22	41	7	1				
7	37	4	20	14	19	15			
8	38	3	18	26	27	21	6		
9	8	33	43	29	16	10	32	28	
0	9	36	44	34	23	17	35	31	1

FITTING CRITERIA

Observed proximity data typically contain a sizable amount of measurement errors, which is why we described approximate relationships between the observed data and models in the previous section. In such cases, we are not seeking an exact representation of the input data, but rather an approximate solution that 'best' represents the observed proximity data. However, this requires an explicit definition of how to measure the discrepancy between the data and model predictions. Parameters in the distance model (i.e., stimulus coordinates) are then estimated so as to minimize the discrepancy. There are two classes of discrepancy functions traditionally used in MDS: the least squares (LS) criterion, and the maximum likelihood (ML) criterion. In this section, we briefly discuss these criteria. We start with the simplest case (i.e., the ratio scale, no replications), and gradually introduce more complicated cases (weaker measurement scales, replications, individual differences, etc.).

Let us begin with LS estimation. It is assumed for the moment that o_{ij}, the observed dissimilarity between stimuli i and j, is measured on a ratio scale. Let d_{ij} denote the Euclidean distance between points i and j as defined in (1). In the LS estimation, we find stimulus coordinates $X = \{x_{ir}\}$ that minimize the discrepancy defined as:

$$\phi(X) = \sum_{i<j}^{n} (o_{ij} - d_{ij})^2 \qquad (5)$$

(The constant of proportionality a is assumed to be unity without loss of generality.) Finding such a solution presents some challenge. A general strategy is to take the derivatives of the above criterion with respect to the model

parameters (X), which are set equal to zero. This leads to a set of simultaneous equations to be solved by an iterative algorithm in which an initial estimate of X is gradually improved according to the gradients (the derivatives of a fitting criterion with respect to unknown parameters evaluated at the current estimates of parameters) until a sufficiently good approximation to the solution is obtained [see Borg and Groenen (2005) for a more detailed explanation of optimization algorithms used in MDS].

An LS criterion is sometimes defined in terms of inner products derived from squared Euclidean distances. Let:

$$p_{ij} = (\overline{d}_{i.}^2 + \overline{d}_{.j}^2 - \overline{d}_{..}^2 - d_{ij}^2)/2 \qquad (6)$$

where $\overline{d}_{i.}^2$ is the mean of d_{ij}^2 over j, $\overline{d}_{.j}^2$ is the mean of d_{ij}^2 over i, and $\overline{d}_{..}^2$ is the mean of d_{ij}^2 over both i and j, and let:

$$\hat{p}_{ij} = (\overline{o}_{i.}^2 + \overline{o}_{.j}^2 - \overline{o}_{..}^2 - o_{ij}^2)/2 \qquad (7)$$

where $\overline{o}_{i.}^2$, $\overline{o}_{.j}^2$, and $\overline{o}_{..}^2$ are analogously defined. [These transformations are called the Young–Householder (1936) transformations.] Using these quantities, we define a LS criterion:

$$\varphi(X) = \sum_{i>j}^{n} (\hat{p}_{ij} - p_{ij})^2 \qquad (8)$$

which is minimized with respect to X as before. One nice thing about this criterion is that such an X can be obtained in closed form. We simply obtain the eigenvalue and vector decomposition of the matrix of $\hat{p}_{ij}$, and retain only those portions of the matrix of eigenvectors pertaining to the R largest eigenvalues. This procedure is called classical MDS (Torgerson, 1952). The solution is simple and straightforward, but the required scale level assumption is rather stringent. Incidentally, this was the method used to recover the ten points configuration on the letter ψ discussed in the introduction section, where o_{ij} is set equal to d_{ij}.

If the dissimilarity data are measured on an interval scale, (5) is modified to:

$$\phi(X, b) = \sum_{i<j}^{n} (o_{ij} - d_{ij} - b)^2 \qquad (9)$$

which is minimized with respect to both X and the additive constant b. A similar iterative procedure to the above may be used to minimize this criterion. The Young–Householder transformation may be used in this case as well, but the resultant procedure is more complicated, as the effect of the transformation on b must be taken into account. The estimation of b requires an iterative solution in any case, although once b is estimated, the stimulus coordinates can be obtained in closed form as before.

When the dissimilarity data are measured on a log-interval scale, we may take the log of both o_{ij} and d_{ij}, and define:

$$\vartheta(X, a, b) = \sum_{i<j}^{n} (\ln o_{ij} - b \ln d_{ij} - \ln a)^2 \qquad (10)$$

which is minimized with respect to X, a, and b by an iterative method. Although this criterion is rarely used in the context of LS estimation *per se*, essentially the same criterion plays an important role in ML estimation, as will be explained below.

When the (dis)similarity data are measured on an ordinal scale (nonmetric MDS), we simultaneously transform the data monotonically (or inverse monotonically), and fit a distance model to the transformed data. Let $m(o_{ij})$ represent the monotonically transformed data, and define:

$$\phi(X, m) = \sum_{i<j}^{n} (m(o_{ij}) - d_{ij})^2 \qquad (11)$$

This is called the raw stress, and is minimized with respect to both X and m subject to the normalization restriction that $\sum_{i<j}^{n} m(o_{ij})^2 = c$. (The normalization restriction is necessary because there is no intrinsic scale for transformed data, and the raw stress can be made

identically equal to zero by setting $m(o_{ij}) = 0$ for all i and j.) Alternatively, the normalization restriction may be directly incorporated into the stress function. That is, the raw stress can be normalized as:

$$\phi^{(1)}(X, m) = \phi(X, m) / \sum_{i<j}^{n} d_{ij}^2 \qquad (12)$$

or as:

$$\phi^{(2)}(X, m) = \phi(X, m) / \sum_{i<j}^{n} (d_{ij} - \bar{d}..)^2 \quad (13)$$

where $\bar{d}..$ is the mean of d_{ij}. These are called the normalized stress 1 and 2, respectively, and can be minimized without any further normalization restriction. The minimization is done by a rather elaborate minimization strategy, combining a monotonic regression algorithm (Kruskal, 1964a, 1964b) with the iterative optimization procedure described earlier.

When there are replicated observations, each of the above criteria may be modified to include another summation (over replications). Let o_{ijk} denote the dissimilarity between stimuli i and j in replication k. Then (5), for example, may be extended to:

$$\phi(X) = \sum_{k=1}^{K} \sum_{i<j}^{n} (o_{ijk} - d_{ij})^2 \qquad (14)$$

Other criteria mentioned above may also be similarly extended. However, in (9), (10), and (11), a, b, and m may be allowed to vary across different replications. In that case, the normalization restriction should be imposed within each replication separately.

When the individual differences (ID) distance model (2) is fitted, we may simply replace d_{ij} in (14) by d_{ijk}. Other criteria tailored to various scale levels can be similarly redefined. These criteria are minimized with respect to both X and W. In ID MDS, however, it is more popular to define a fitting criterion in terms of inner products as follows: Let $\hat{p}_{ijk}$ and p_{ijk} denote the observed and model inner products, respectively, derived analogously to (6) and (7) for each k. Then:

$$\vartheta(X, W) = \sum_{k=1}^{K} \sum_{i>j}^{n} (\hat{p}_{ijk} - p_{ijk})^2 \qquad (15)$$

This criterion is valid only for ratio-scaled dissimilarity data.

In unfolding analysis (3), we may simply replace the range of summation from $\sum_{i<j}^{n}$ in (5), (9), (10), and (11) to $\sum_{j}^{N} \sum_{i}^{n}$, where i is the index of stimuli and j is the index for subjects. The only difference is that in unfolding analysis these criteria are minimized with respect Y as well as X and other data transformation parameters. Experience has indicated that these criteria often lead to so-called degenerate solutions in unfolding analysis. The degenerate solutions fit the data (or the transformed data) nearly perfectly, but are substantively uninteresting. The most common form of a degenerate solution is one in which stimulus points and ideal points are completely separated in the space. To avoid this type of degenerate solution, Busing, Groenen, and Heiser (2005) proposed to penalize the LS criteria by the coefficient of variation. The resultant computer program, PREFMAP, does a good job in avoiding degenerate solutions.

In the ML estimation, we make a specific distributional assumption on o_{ijk}, based on which we define the likelihood of observing the set of proximity data at hand as a function of X. We then find X that maximizes the likelihood. Let us assume a log-normal distribution on o_{ijk} measured on a log-interval scale, as in Ramsay (1977, 1982). This distribution has several desirable properties as the distribution of observed dissimilarity data. First of all, it is defined only for positive values of o_{ijk}. Second, it is positively skewed, indicating that large errors tend to occur on the positive side. Finally, it has larger variances for larger distances. It is convenient to take the log of o_{ijk}, since the log-normal distribution then reduces to a normal distribution. That is:

$$\ln o_{ijk} \sim \mathcal{N}(b_k \ln d_{ij} + \ln a_k, \sigma^2) \qquad (16)$$

Then, the log likelihood for an entire set of observations can be stated as:

$$\ln L(X, a_k, b_k, \sigma^2) = -\frac{1}{2}\left(\frac{S}{\sigma^2} + M \ln \sigma^2\right) \tag{17}$$

where:

$$S = \sum_{k=1}^{K} \sum_{i>j}^{n} (\ln o_{ijk} - b_k \ln d_{ij} + \ln a_k)^2 \tag{18}$$

and M is the total number of observations. Maximizing (17) with respect to σ^2 leads to:

$$\hat{\sigma}^2 = S/M. \tag{19}$$

Let:

$$\ln L^*(X, a_k, b_k) \stackrel{def}{=} \ln L(X, a_k, b_k, \hat{\sigma}^2) = -\left(\frac{M}{2}\right)(\ln S + 1 - \ln M). \tag{20}$$

Maximizing this criterion with respect to X, a_k, and b_k is equivalent to minimizing $\ln S$, which in turn is equivalent to minimizing S. (Note that S reduces to (10) when there is a single replication per stimulus pair.) The S is minimized with respect to X, a_k, and b_k by a similar iterative optimization technique as those used in the LS estimation.

The likelihood function in ML MDS varies from one type of proximity data to another, since it has to take into account a specific response mechanism that generates a specific type of proximity data. ML MDS procedures have been developed for a variety of proximity data by Takane and his collaborators (Takane, 1978, 1981, 1987; Takane and Carroll, 1981; Takane and Sergent, 1983; Takane and Shibayama, 1986), each requiring a different specification of the likelihood function.

The ML estimation provides asymptotically efficient estimates of parameters, when the fitted model and the distributional assumption are correct. It also provides information regarding how reliably stimulus coordinates are estimated, and some hypothesis testing capabilities. The AIC statistic, defined by:

$$AIC_\pi = -2 \ln L^*_\pi + 2n_\pi \tag{21}$$

may be used to identify the best fitting model, where π indicates a specific model fitted, L^*_π is the maximum likelihood of model π, and n_π is the effective number of parameters. The model associated with the smallest value of AIC indicates the best fitting model. Note that the above remarks should be taken with some caution. The distance model is never exactly correct, and the log-normal assumption is often only approximately true. In addition, in most applications there are not enough observations to rely on the asymptotic properties of ML estimators.

EXAMPLES OF APPLICATION: THE SIMPLE EUCLIDEAN MODEL

In this section, we present two examples of applications of MDS with the simple Euclidean model to real data sets. The first example pertains to dissimilarity judgments made on visual characteristics of ten phonetic symbols in Korean (representing vowels). The second example concerns similarity data for 18 animals collected by the sorting method.

Example 1: similarity of shape among ten Korean phonetic symbols

This study employed simple MDS (MDS with the simple Euclidean model) to represent the visual similarity between ten Korean phonetic symbols in an MDS configuration. The Korean alphabet (Hangul) has ten simple vowels based on two distinctive elements: a long line segment combined with zero, one, or two short line segments. The ten stimuli used are: ㅏ, ㅑ, ㅓ, ㅕ, ㅗ, ㅛ, ㅜ, ㅠ, ㅡ, and ㅣ. The subjects were four university students (one female and three males). All subjects were English speakers (three native and one bilingual (French/English)] with normal vision, who had no previous Korean learning experience. The visual dissimilarity among

the 10 Korean phonetic symbols were rated on a 9-point rating scale. They were allowed to take as much time as they needed to make their judgments. All pairs of stimuli were arranged in random order, and presented to the subjects. All participants completed a questionnaire with 45 pairs of stimuli.

We used MULTISCALE (Ramsay, 1997), a maximum likelihood MDS program, to derive a multidimensional stimulus configuration. The minimum AIC criterion indicated that the two-dimensional solution is the best ($AIC_1 = 322.0$; $AIC_2 = 290.2$; $AIC_3 = 293.1$). The two-dimensional weighted Euclidean model was also fitted, which turned out to be not as good as the two-dimensional simple Euclidean model (AIC = 296.0). Figure 10.4 displays the optimal two-dimensional stimulus configuration. Dimension 1 (the horizontal direction) contrasts symbols with a long vertical line segments on the right and those with a long horizontal line segment on the left. Dimension 2, on the other hand, roughly corresponds to the number of small segments attached to the long segment. Symbols with two short segments are located at the top, those with one short segment in the middle, and those with no short segments toward the bottom. It seems that the similarity relations among the 10 Korean phonetic symbols are organized around two principal attributes: the orientation (horizontal or vertical) of a long line segment, and the number of short segments attached to the long segment. This kind of information may be useful for language acquisition researchers in understanding how people perceive visual relationships among the symbols.

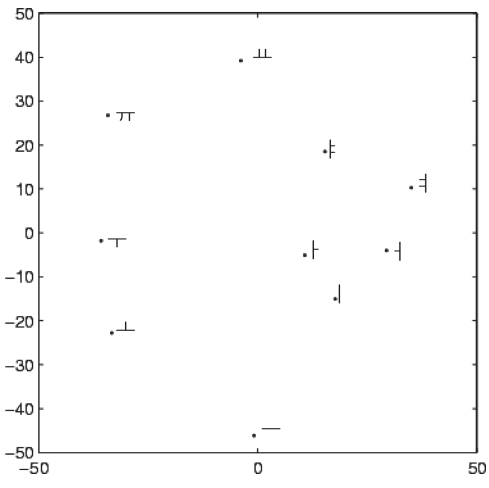

Figure 10.4 Two-dimensional configuration of 10 Hangul symbols (vowels visually presented). See text for further details.

Example 2: animals

The second example in this section involves similarity judgments between 18 animals collected from 20 subjects by the sorting method. The subjects were asked to classify the 18 animals into as many groups as they wanted in terms of their similarity. The number of groups into which stimuli were sorted varied over the subjects. The sorting data can be summarized in the form of a subjects-by-stimuli table, such as in Table 10.3, in which the rows represent the 20 subjects and columns the 18 stimuli. Entries in the table indicate cluster numbers into which stimuli were sorted. Which integers are used to represent which sorting clusters are essentially arbitrary within each subject. From this table, the number of times each pair of animals were sorted into the same group was tabulated and used as a similarity measure between them. Nonmetric MDS with the simple Euclidean model was used to analyze the data.

Figure 10.5 presents the derived three-dimensional stimulus configuration. The three-dimensional solution was chosen primarily for ease of presentation. We drew tick marks (along with animal names) on each dimensional axis, so we could see where the animals are located on each of the three dimensions. In this figure, the 18 animals are labelled as: bear (be), camel (cm), cat, (ct), cow (cw), dog (dg), elephant (el), fox (fx), giraffe (gi), horse (ho), lion (ln), monkey (mk), mouse (ms), pig (pg), rabbit (rb), sheep (sh), squirrel (sq), tiger (tg), and wolf (wf). (Symbols in parentheses are plotting symbols used in Figure 10.5.) The first dimension contrasts farm animals with non-farm animals. Animals such as pig, cow,

Table 10.3 Sorting data for 18 animals

	Stimulus																	
Subject	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*
1	1	2	2	2	2	1	3	1	2	3	1	4	2	4	2	4	3	3
2	1	2	3	4	3	2	1	2	2	1	3	3	4	3	4	3	1	1
3	1	2	3	4	3	2	1	2	4	2	5	6	4	1	4	1	2	1
4	1	2	3	4	5	6	5	2	2	3	6	7	4	8	9	7	3	5
5	1	2	3	4	5	6	5	6	2	3	6	7	4	4	4	7	3	5
6	1	2	3	4	5	6	5	7	4	3	1	8	6	8	4	1	3	5
7	1	2	3	4	3	2	5	2	4	6	7	3	4	1	4	7	6	5
8	1	2	3	4	3	2	1	5	2	6	7	8	4	8	4	8	6	1
9	1	2	3	2	4	5	4	5	2	1	3	3	2	3	2	3	1	4
10	1	2	3	4	5	2	5	2	4	3	1	6	4	7	4	6	3	5
11	1	2	3	3	3	2	1	2	3	1	2	4	5	4	3	4	1	1
12	1	2	3	4	3	5	6	5	4	1	2	5	4	6	5	3	1	6
13	1	2	3	2	4	5	4	2	6	3	7	8	9	8	9	8	3	4
14	1	2	3	4	3	2	5	6	2	1	7	5	4	5	4	5	1	1
15	1	2	3	4	3	5	6	5	4	7	5	8	4	9	4	10	7	6
16	1	2	3	4	3	2	1	5	2	1	3	3	4	3	4	6	1	1
17	1	2	3	4	3	2	1	2	4	2	2	5	4	5	4	5	2	1
18	1	2	3	4	3	2	1	2	4	1	5	6	4	7	8	6	1	1
19	1	2	3	4	5	6	5	7	8	3	9	10	11	12	13	10	3	5
20	1	2	3	4	3	2	1	2	4	2	2	3	4	1	4	1	2	1

The stimuli are: 1. Bear (be), 2. Camel (cm), 3. Cat (ct), 4. Cow (cw), 5. Dog (dg), 6. Elephant (el), 7. Giraffe (gf), 8. Fox (fx), 9. Horse (hs), 10. Lion (li), 11. Monkey (mk), 12. Mouse (ms), 13. Pig (pg), 14. Rabbit (rb), 15. Sheep (sh), 16. Squirrel (sq), 17. Tiger (tg), 18. Wolf (wf).

sheep, and horse are located on the left side, while fox, wolf, tiger, cat, and so on are placed on the opposite side. The second dimension distinguishes two possible habitats of animals, either a wild habitat or a habitat close to people. Animals such as mouse, cow, pig, dog, etc. are placed on the left towards the back, while lion, tiger, elephant, etc. are located toward the front side. The third dimension separates animals at a higher level of the food chain and those at a lower level. On this dimension, animals such as mouse, squirrel, rabbit, etc. are located at the top, with lion, tiger, bear, wolf, and fox at the bottom. It is interesting to find that similarity judgments among these animals are organized around these three dimensions, which could only be uncovered by MDS.

EXAMPLES OF APPLICATION: THE WEIGHTED EUCLIDEAN MODEL

As noted earlier, the particular kind of individual differences MDS we use postulates a stimulus configuration that is common to all individuals, but that dimensions are differentially weighted by different individuals to generate different proximity judgments. In this section, we present two examples of applications of ID MDS. The first example involves a set of artificial toy-like objects. The second example is concerned with dissimilarity judgments on 14 consonant sounds in Korean.

Example 3: the toy-like objects

The first example in this section pertains to a set of dissimilarity judgments between eight artificially created toy-like objects (still pictures displayed in Figure 10.6) obtained from three groups of subjects. These objects were created for a study investigating whether young children would learn a new object name, "blick', based on the function or on the appearance (Nguyen and Oshima-Takane, 2008). 'Blick' was the name of the target objects (A, D, E, G, and H), which had the function of moving the center rod from side to side in the main body of the object. The objects B, C, and F were distracters,

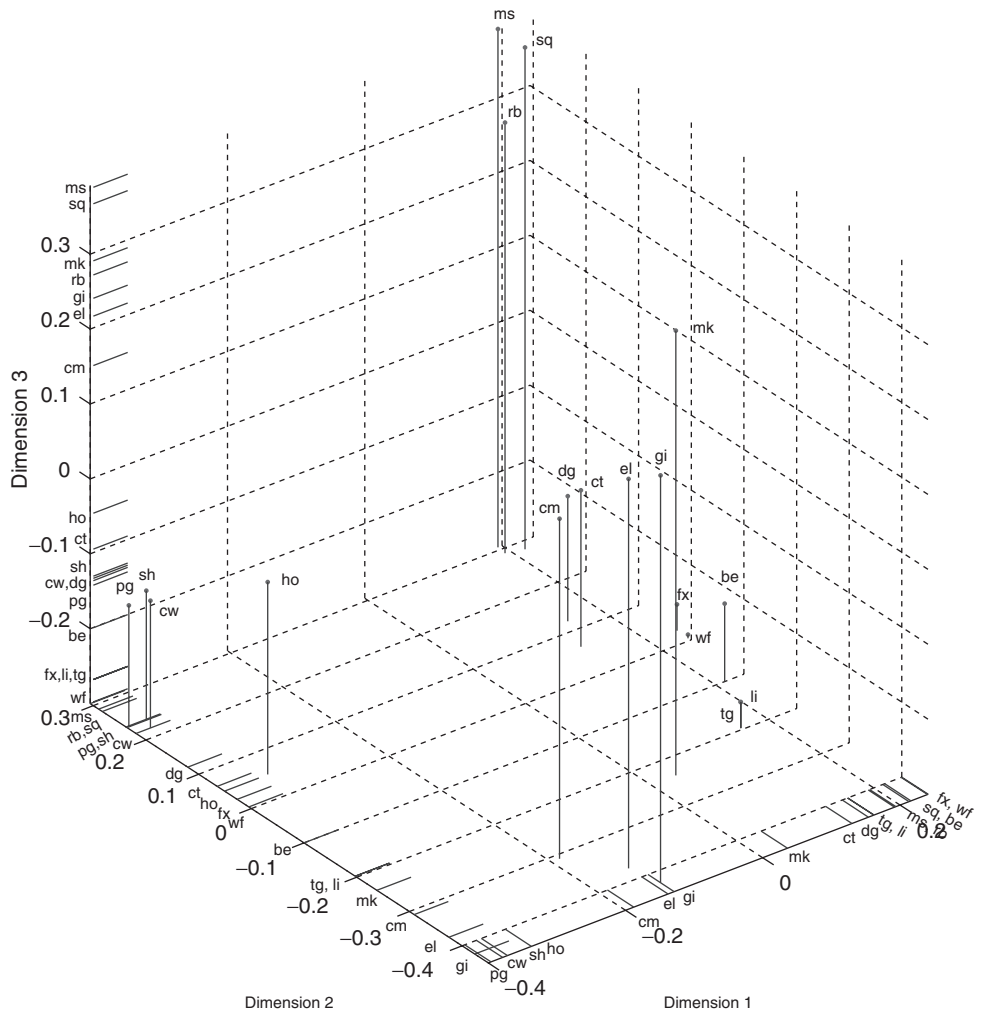

Figure 10.5 Three-dimensional configuration of 18 animals. See text for further details.

which had a function different from the 'blick' objects, although their overall appearance looked similar to the 'blicks'. The function of the non-target objects B and C was to revolve two bolts sticking out of the main body. In F, the rods sticking out of the main body was used as handles to open up the space between the two body parts such that the overall object looked like a mouth opening and closing. Subjects were first shown a movie with all eight objects, one at a time. Then, they saw all the objects on the same screen and were asked to select the most similar and the most dissimilar pairs. They were then presented with a pair of objects side by side and were asked to rate the degree of dissimilarity between the two objects.

The first group of six subjects were asked to judge the dissimilarity between the objects by their appearance, the second group of six subjects by their function (distinguished by the movement of the center rod), and the third group of six subjects according to unspecified criteria. The data were collected using an 11-point rating scale. Figure 10.7 shows the two-dimensional common stimulus configuration derived by a nonmetric ID MDS program. The two dimensions are interpreted as follows: Dimension 1 (the horizontal

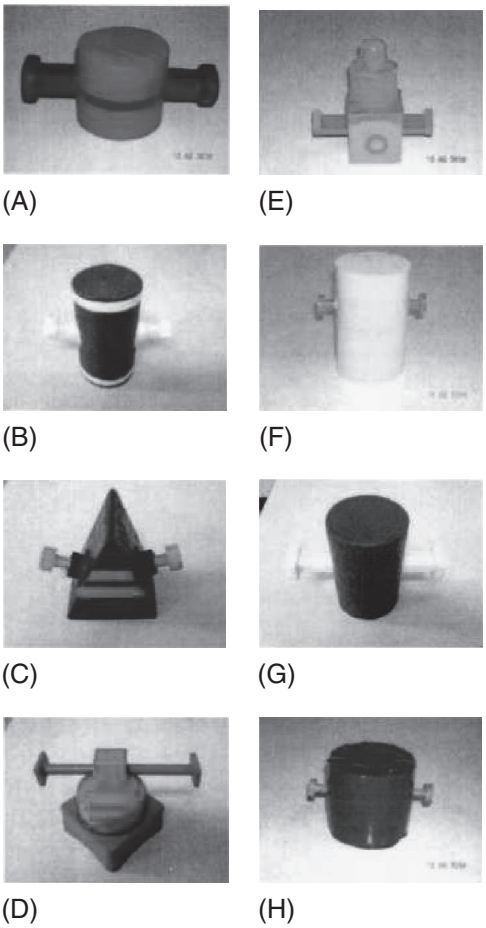

Figure 10.6 Still pictures of eight stimuli used in the 'Blicks' study. There are two groups of objects distinguished by their function (the movement of the center rod). In (A), (D), (E), (G), and (H), the center rod moved from side to side, whereas in (B), (C), and (F), the center rod moved differently.

direction) represents 'function', separating B, C, and F from the rest. Dimension 2 (the vertical axis) represents appearance, contrasting the objects with a slender top (C, D, and E) and those with a non-narrowing top (A, B, F, G, and H). Figure 10.8 presents the weights attached to these two dimensions by the 18 different subjects. The six subjects assigned to the first condition are labeled as 1 to 6, those in the second condition as 7 to 12, and those in the third condition as 13 to 18. Quite naturally, the subjects in the first group tend to put more emphasis on the appearance dimension (dimension 2). (The only exception is subject 2, who put more emphasis on the function dimension.) The six subjects in the second group tend to put more emphasis on the 'function' dimension (dimension 1). The weights are fairly tightly clustered on dimension 1 with a slight exception of subject 10. The subjects in the third group tend to vary between the first two groups, with a majority of them putting similar emphasis on both dimensions. This example shows that the kind of ID MDS we used is working the way it should, and is able to capture the kind of individual differences in (dis)similarity judgments that it is designed to capture.

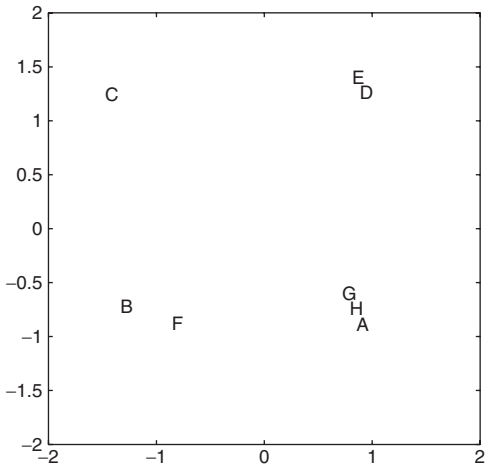

Figure 10.7 Two-dimensional stimulus configuration of toy-like objects.

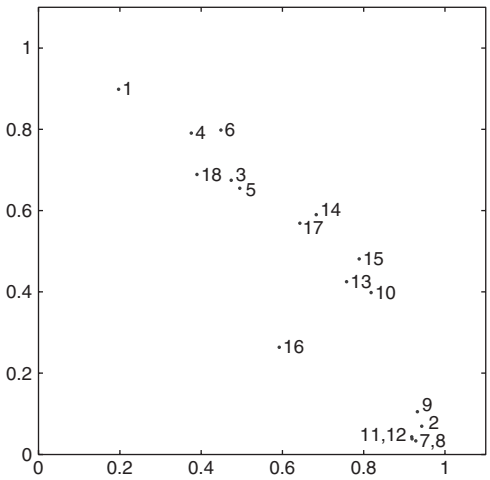

Figure 10.8 Individual differences weights attached to the two dimensions.

Example 4: Korean consonants

The second example in this section concerns dissimilarity judgments between fourteen consonants in Korean as they are pronounced with a particular vowel 'a'. These sounds are: 'ga', 'na', 'da', 'ra', 'ma', 'ba', 'sa', 'a', 'ja', 'cha', 'ka', 'ta', 'pa', and 'ha'. These consonants have been classified by phonologists according to two criteria: points of articulation and methods of articulation. The first criterion classifies them into: 'ka' and 'ga' (palatal); 'na', 'ra', 'ta', and 'da' (lingual, tongue); 'ma', 'ba', and 'pa' (labial, lips); 'sa', 'cha', and 'ja' (dental, teeth); and 'ha' and 'a' (glottal, throat). The second classification scheme, on the other hand, classifies them into: 'ka', 'ga', 'ta', 'da', 'pa' and 'ba' (plosive); 'sa' and 'ha' (fricative); 'cha' and 'ja' (affricative); 'na' and 'ma' (nasal); and 'ra' (trill). The two classification schemes are often combined into a two-way classification table. It is interesting to see how well these classification schemes fare in the subjective judgments of similarities.

Subjects were three undergraduate students at a large Canadian university. All subjects were female and bilingual (French and English) speakers with normal hearing. A 9-point rating scale was used to record dissimilarity judgments. The dissimilarity data were analyzed by individual differences MDS by ML (MULTISCALE).

In the present study, a three-dimensional solution is chosen, partly for ease of presentation, although AIC decreased consistently up to the six-dimensional solution. (Due to the incidental parameters in the weighted Euclidean model, the minimum AIC criterion is not completely reliable.) We also analyzed the data by the simple Euclidean model. However, ID MDS consistently outperformed simple MDS. There seem to be systematic individual differences in the way the three dimensions were evaluated by the three subjects. Interestingly, all three subjects put more emphasis on two of the three dimensions, although the particular two dimensions they put more emphasis on varied among the three subjects. Subject 1 put more emphasis on the second and the third dimensions, subject 2 on the first and the second dimensions, and subject 3 on the first and the third dimensions. Unfortunately, the source of these differential patterns cannot be investigated further without additional information about the subjects.

Figure 10.9 displays a common stimulus configuration of the fourteen Korean consonants. The first dimension separates

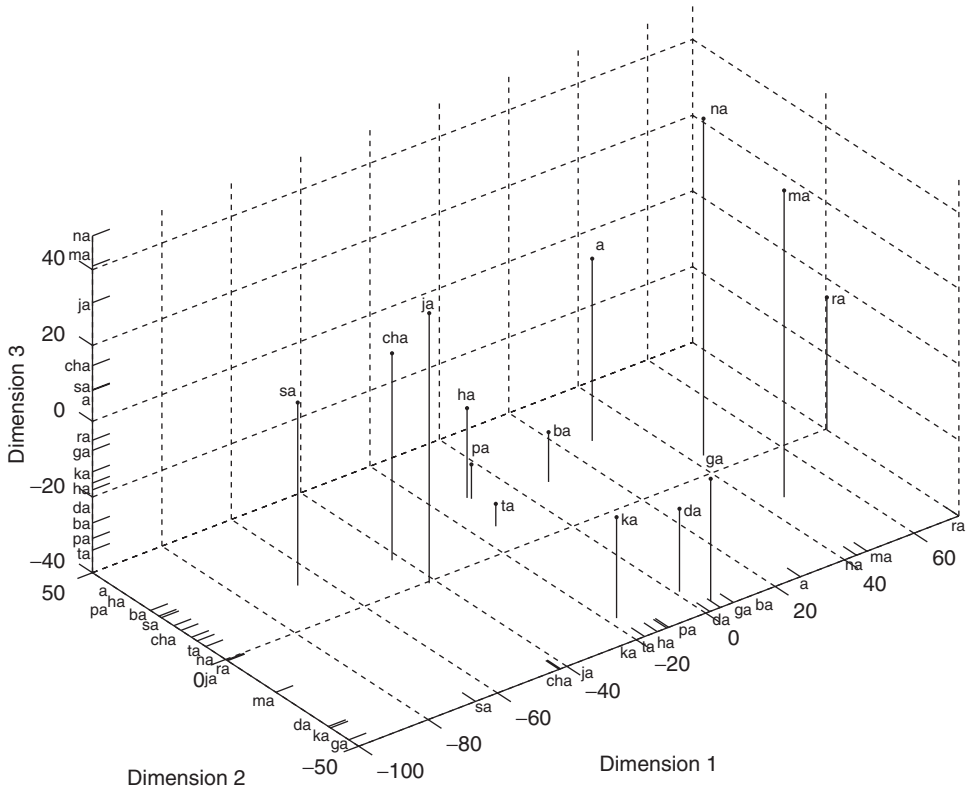

Figure 10.9 Three-dimensional configuration of 14 Hangul consonants as they are pronounced.

dental sounds ('sa', 'cha', and 'ja') on the left hand side from a liquid (trill) sound ('ra') and nasal sounds ('na' and 'ma') on the right. You might also say that the sounds on the left hand side of dimension 1 are fricative ('sa' and 'ha') and affricative ('cha' and 'ja') consonants. The second dimension contrasts palatal consonants ('ka' and 'ga') in the front and glottal (throat) consonants ('a' and 'ha') in the back. (The consonant 'da' is close to the front group, but it should be noted that it is also one of the plosive consonants like 'ka' and 'ga'.) The third dimension separates plosive sounds ('ta', 'da', 'pa', 'ba', 'ka', and 'ga') at the bottom from all other consonants, but most notably from nasal consonants ('na' and 'ma') at the top. Thus, phonologists' classification schemes are useful, although the correspondence is not exact between their schemes and the psychological space.

AN EXAMPLE OF APPLICATION: UNFOLDING ANALYSIS

In this section, we present an example of unfolding analysis of preference data collected on actual commercial products. As has been noted earlier, unfolding analysis is a special kind of MDS for the analysis of preference data, construed as representing proximity relations between subjects' ideal and actual stimuli. It attempts to account for individual differences in preference judgments by mapping subjects' ideal and actual stimuli in a joint multidimensional space in such a way that the closer the stimulus is to one's ideal, the more it is preferred by the subject.

Example 5: MP3 players

This study was designed to investigate the relationship between preferences on various

Table 10.4 MP3 Players and the descriptive variables

Number	*Product*	*Memory (mt)*	*GB (mc)*	*Price (pr)*	*Volume (sz)*	*Time (pt)*
(ip)	iPod	HD	30	$299	70.9	14h
(no)	iPod nano	Flash	4	$229	25.2	24h
(st)	iPod Shuffle	Flash	1	$89	12.1	12h
(zn)	Zune	HD	30	$299	104.1	14h
(mv)	Muvo V100	Flash	2	$79	41.6	18h
(zv)	Zen Vision:M	HD	30	$299	184.4	14h
(tr)	TRIO	Flash	1	$55	36.9	10h
(yk)	YP-K5JZ	Flash	1	$199	85.1	10h
(yz)	YP-Z5	Flash	4	$229	44.1	35h
(wm)	Walkman	Flash	1	$149	40.7	18h

brands of portable MP3 (MPEG-1 Audio Layer 3) players and their features. Stimuli were ten different models of MP3 players characterized by five descriptor variables such as the memory type (either hard drive or Flash drive), memory capacity, price, volume (size), and playback time, as shown in Table 10.4. The ten MP3 players are: (ip) iPod, (no) iPod nano, (st) iPod shuffle, (zn) Zune, (mv) Muvo V100, (zv) Zen Vision:M, (tr) TRIO MP3 player, (yk) YP-K5JZ, (yz) YP-Z5, and (wm) NWS203FB Walkman.

A group of 20 subjects were asked to rank order these products according to their preferences by assigning 1 to the most preferred model and 10 to the least preferred model. When assessing their preferences, subjects were shown pictures of the MP3 players, and listened to detailed descriptions. The preference rankings collected from 20 subjects are shown in Table 10.5. The data were analyzed by unfolding analysis using PREFSCAL (Busing et al., 2005), and a joint configuration of stimulus points and subjects' ideal points was obtained.

Table 10.5 Preference rankings on 10 MP3 players

	*MP3 Players**									
Subject	*ip*	*no*	*st*	*zn*	*mv*	*zv*	*tr*	*yk*	*yz*	*wm*
1	7	5	8	10	1	2	3	6	4	9
2	9	4	5	7	2	8	1	10	6	3
3	4	1	3	5	2	6	10	9	7	8
4	8	4	1	9	2	10	6	7	5	3
5	2	8	5	3	1	4	7	10	9	6
6	1	3	5	2	9	7	10	4	6	8
7	9	7	3	5	2	10	1	6	8	4
8	6	7	3	9	2	8	1	10	5	4
9	9	8	7	10	3	4	2	5	6	1
10	1	5	10	2	8	3	9	6	4	7
11	4	1	10	5	2	6	9	3	7	8
12	6	4	8	10	1	9	3	7	5	2
13	6	1	4	9	5	8	10	3	2	7
14	1	6	4	2	3	5	7	8	9	10
15	7	9	8	10	3	5	4	1	2	6
16	1	2	4	6	8	3	9	7	5	10
17	4	2	5	10	1	3	9	6	7	8
18	4	1	2	8	10	5	9	3	6	7
19	10	4	3	9	1	8	2	7	6	5
20	1	5	9	2	4	3	7	10	8	6

* See Table 10.4 for the definitions of the abbreviations.

Figure 10.10 displays the derived two-dimensional stimulus and ideal point configuration. Stimuli are labeled by two-letter sequences, and subjects are labeled by numbers from 1 to 20. The five descriptor variables are also mapped into the configuration as vectors indicating the directions with which these variables are most highly correlated. These vectors are labeled by boldfaced letter combinations: **mt** (memory type), **mc** (memory capacity), **sz** (size), **pr** (price), and **pt** (playback time). The incorporation of the descriptor information facilitates dimensional interpretations of the derived configuration. Hard drive memory, large memory capacity, size, and price (**mt**, **mc**, **sz**, and **pr**), are most highly correlated with the (upper) right-hand side of the configuration. Products with hard drive (HD) memory and high memory capacity such as (ip) iPod, (zn) Zune, and (zv) Zen Vision tend to be located toward the upper right-hand side of the configuration. Subjects 6, 10, 16, and 20 have strong preferences for this type of product. Products with Flash memory, small memory capacity, relatively

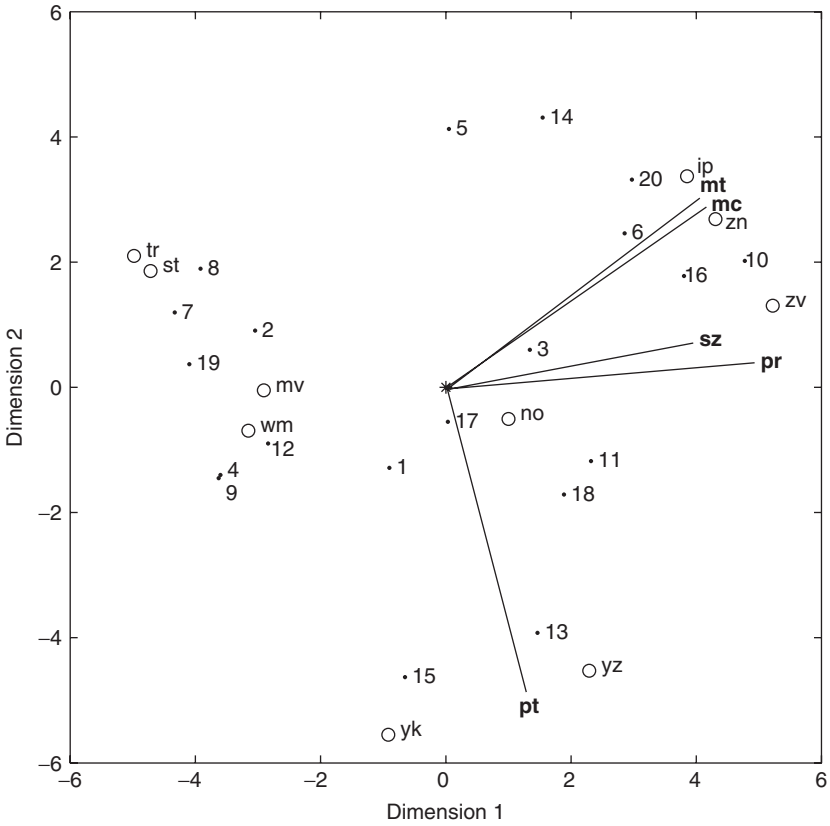

Figure 10.10 Two-dimensional configuration of the ten MP3 players and subjects' ideal points (1–20).

small, and less expensive models such as Trio (tr), iPod Shuffle (st), Muvo V100 (mv), and Walkman (wm), are located on the opposite side. Subjects 2, 7, 8, and 19 are presumed to have strong preferences for the first two of these products, and subjects 4, 9, and 12 for the last two of these products. Playback time (**pt**) is most highly correlated with the bottom side of the configuration. Products with long playback time such as YP-Z5 (yz) and YP-K5JZ (yk) are located toward the bottom of the configuration. Subjects 13 and 15 have strong preferences for this type of product. iPod nano (no) is somewhat unique in that it cannot be well characterized by the two dimensions extracted. (The iPod nano might have loaded highly on the third dimension if the three-dimensional solution had been obtained.) Still, subjects 1, 3, and 17 like this type of of product (fairly expensive, small sized, with small memory capacity but quite a long playback time).

There may be weak relationships between subjects' demographic information and their preferences. There are six female subjects (subjects 7, 8, 9, 11, 13, and 19), none of whom is in the upper right corner. In fact, four of them are in the left-hand, side preferring less expensive models. Two of them (subjects 7 and 8) are also mature subjects (of age above 25). There are five mature male subjects (subjects 2, 9, 5, 6, and 14), two of whom are in the left-hand side of the configuration and the remaining three in the upper right-hand side; none of them is in the bottom portion of the configuration.

Unfolding analysis is a very useful technique in marketing research. It allows us to understand patterns of individual differences in preference judgments, and their

relationships to product features and subjects' background information. This kind of analysis may eventually help marketing analysts to develop practical marketing strategies.

CONCLUDING REMARKS

In this chapter, we have attempted to provide an integrative overview of three representative MDS models: simple MDS, individual differences MDS, and unfolding analysis. A number of empirical examples reflect interesting applications of MDS as a tool for spatial representations of similarity/preference data. In this brief overview of MDS, however, only a few methods and examples of application could be presented. In particular, algorithmic details had to be left out almost entirely. For more detailed explanations of how MDS works, the reader should consult a monograph focused on more technical aspects of MDS. It is expected that MDS will generate further interest with the development of more flexible and reliable algorithms. MDS is expected to remain a powerful and useful methodology in social and behavioral sciences.

A number of popular software programs are making MDS easily accessible to social and behavioral science researchers. MULTISCALE (Ramsay, 1997) is a suitable program for MDS by ML estimation method. KYST (Kruskal, Young, and Seery, 1978) is a good and reliable program for nonmetric MDS. INSCAL (Arabie, Caroll, and DeSarbo, 1987), ALSCAL (Schiffman, Reynolds, and Young, 1981) and PROXSCAL (Busing, Commandeur, and Heiser, 1997) is implemented in SPSS to perform both simple and individual differences MDS. PREFSCAL (Busing et al. 2005) is also available in SPSS for unfolding analysis.

ACKNOWLEDGEMENT

The work reported in this chapter has been supported by research grants, 10630 and 42720, from the Natural Sciences and Engineering Research Council of Canada to the first and third authors, respectively.

REFERENCES

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

Arabie, P., Carroll, J.D. and DeSarbo, W.S. (1987) *Three-way Scaling and Clustering*. Newbury Park, CA: Sage Publications.

Borg, I. and Groenen, P. (2005) *Modern Multidimensional Scaling: Theory and Applications.* (2nd Edn.). New York: Springer.

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 Bandilla, W and Faulbaum, F. (eds.), *Softstat '97: Advances in Statistical Software*. Stuttgart, Germany: Lucius. pp. 237–258

Busing, F.M.T.A., Groenen, P. J.F., Heiser, W. (2005) 'Avoiding degeneracy in multidimensional unfolding by penalizing on the coefficient of variation', *Psychometrika,* 70: 71–98.

Busing, F.M.T.A., Heiser, W.J., Neufeglise, P. and Meulman, J.J. (2005) *PREFSCAL*. Chicago, IL: SPSS, Inc.

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

Coombs, C.H. (1964) *A Theory of Data.* New York: Wiley.

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

Greenacre, M.J. (1984) *Theory and Applications of Correspondence Analysis.* London: Academic Press.

Hubert, L., Arabie, P., 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.

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., Young, F.W., Seery, J.B. (1978) *How to use KYST, a Very Flexible Program to do Multidimensional Scaling and Unfolding* (Technical Report). Murray Hill, NJ: Bell Laboratories.

Nishisato, S. (1980) *Analysis of Categorial Data: Dual Scaling and its Applications.* Toronto: University of Toronto Press.

Nguyen, T.-K. and Oshima-Takane, Y. (2008, March) 'Do 2-year-old children use functional cues to name objects?' Poster presented at the XVIth Conference on Infant Studies. Vancouver, Canada.

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

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

Ramsay, J.O. (1997) *MULTISCALE Manual* (Extended version). Unpublished manuscript, McGill University.

Rothkopf, E.Z. (1957) 'A measure of stimulus similarity and errors in some paired associate learning', *Journal of Experimental Psychology,* 53: 94–131.

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

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.

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

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, 105–114.

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

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

Takane, Y. (1994) 'A review of applications of AIC in psychometrics', in Bozdogan, H. (ed.), *Proceedings of the First US/Japan Conference on the Frontiers of Statistical Modeling: An Informational Approach.* Dortrecht: Kluver Academic Publisher. pp. 379–403.

Takane, Y. (2007) 'Applications of multidimensional scaling in psychometrics', in Rao, C.R. and Sinharay, S. (eds.), *Handbook of Statistics. Volume 26.* Amsterdam: Elsevier. pp. 359–400.

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 Leeuw, J. de., Heiser, W.J., Meulman, J. and Critchley, F. (eds.), *Multidimensional Data Analysis.* Leiden: DSWO Press. pp. 119–138.

Takane, Y. and Shibayama, T. (1992) 'Structures in stimulus identification data', in Ashby, F.G. (ed.), *Probabilistic Multidimensional Models of Perception and Cognition.* Hillsdale, NJ: Lawrence Erlbaum Associates. pp. 335–362.

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.

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

Tversky A. (1977) 'Features of similarity', *Psychological Review,* 84: 327–352.

van der Heijden, P.G.M., Mooijaart, A. and Takane, Y. (1994) 'Correspondence analysis and contingency table models', in M.J. Greenacre, and J. Blasius (eds.), *Correspondence Analysis in the Social Sciences.* Academic Press, New York. pp. 79–111.