Excerpted from *Eliciting, analyzing, and comparing mental models of complex audio recording systems between professional and novice recording engineers.* B. R. Hill

Analyzing card sort data

"Analyzing card sort data is part science, part magic. Analysis can be done in two ways: by looking for broad patterns in the data or by using cluster analysis software" (Maurer & Warfel, 2004, Analyzing the results). With this in mind, it is desirable to analyze the card sort results by reviewing all the resulting piles and looking for apparent patterns in how participants organized the terms; a statistical analysis is also helpful for providing a quantitative perspective. For this study, each participant created their own pile categories, so exact one-to-one comparisons were not possible. However, general organizational structures would most likely be reflected among the resulting sort files, and it was expected that similar categories would be apparent, though perhaps more so with the experts' sorts.

Although the investigator's original intent for statistically analyzing the sort data was to compare multidimensional scaling and Pathfinder network procedures, it was discovered that the current version of PC KNOT software, used for producing PFNets, is an old software application that ran only under DOS-based systems. Discussions with the developer confirmed this, so that option was discarded. Therefore all statistical processing of the sort data was performed with multidimensional scaling procedures using the SPSS software application.

Scoring the card sorts was accomplished by selecting the "Score" function within the software for each sort. This generated three data files that were used for analysis. One of these is a list of all piles for that particular sort along with the pile names given by the participant. The second is a list of all twenty cards along with the names of the piles where each was placed. The third file is a correlation matrix that lists each possible pairing of all cards included in the exercise and which pairs were included in the same pile. A "1" next to a particular pair of cards indicates they were placed in the same pile whereas a "0" signifies they were not.

Of the three output data files from the card sort software, the one used for analysis was the correlation matrix. This binary data for each case was entered manually into a separate Microsoft Excel spreadsheet, creating a 20x20 lower triangular data matrix (see

Appendix B for an example). At this stage the binary configuration was such that a “1” in any cell indicated that those two cards were in the same pile while a “0” indicated they were not. For processing with SPSS, however, it was found that these needed to be inverted so that a “0” indicated a positive pairing. This is known as a dissimilarity matrix. Therefore, a simple formula was employed in Excel so that when the data was entered it was automatically converted (=IF(B2=1,0,1)). These resulting matrices were then imported into SPSS and saved as SPSS data format files.

SPSS is a sophisticated statistics/data measurement software package. The current version (13.0 for Windows) provided two different algorithms for multidimensional scaling (MDS) analysis: ALSCAL and PROXSCAL. Though the latter is the most recently developed approach to producing MDS maps, it was not available on the version licensed at the investigator’s institution. ALSCAL was therefore used for all MDS analysis; ALSCAL has been used extensively for many years and for most applications it works well. Though a detailed mathematical explanation of the MDS process is not appropriate here, a brief overview of the primary issues is presented to instruct the reader in the process of employing this method.

Multidimensional scaling is a mathematical means employed “to detect meaningful underlying dimensions that allow the researcher to explain observed similarities or dissimilarities (distances) between the investigated objects” (Statsoft, Multidimensional scaling, General purpose, n.d.). “Multidimensional scaling (MDS) encompasses a collection of methods which allow to gain insight [sic] in the underlying structure of relations between entities by providing a geometrical representation of these relations” (van Deun & Delbeke, 2000, Introduction). In other words, the procedure generates a spatial map with data points located relative to how they were perceived to be similar or dissimilar; similar items will be grouped together while dissimilar items are spread farther apart (Kruskal & Wish, 1978; Borg & Groenen, 1997). Thus this method is ideal for analyzing proximity measures of structural knowledge patterns, providing results that can be depicted in a visual representation showing relative strengths of connections among concepts in a specified content domain. Of the various methods employed for analyzing card sort data, MDS is one of the most commonly selected methods in the literature (Jonassen et al., 1993).

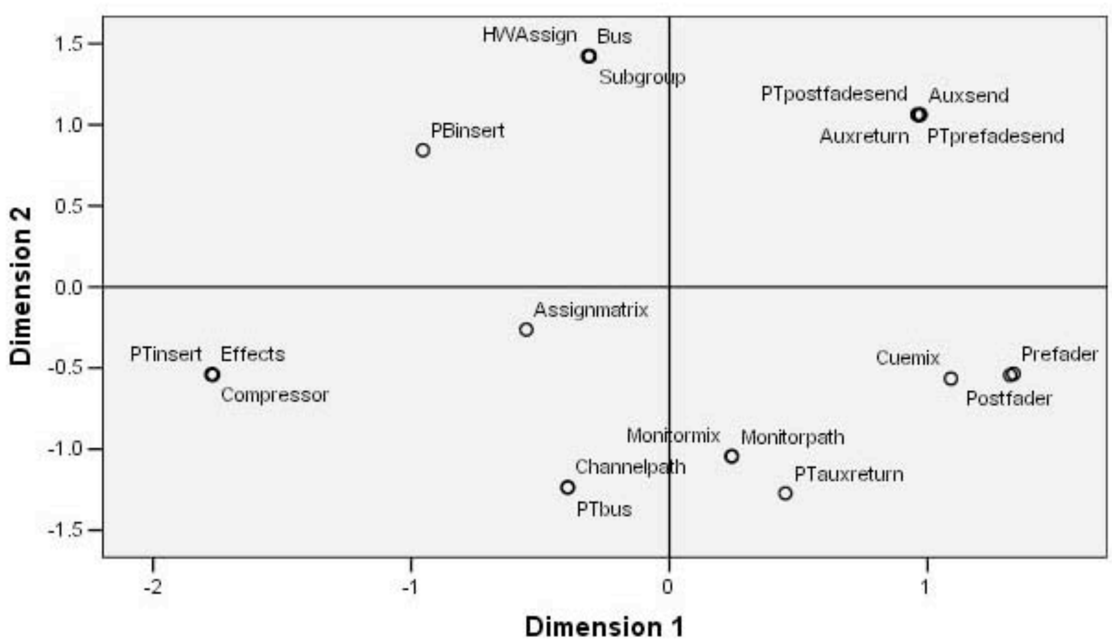

Figure 3.5. Multidimensional scaling plot.

There are two primary components to examine when interpreting MDS maps: clusters and dimensions. When analyzing card sort data, cards (concepts) that were deemed more similar by a majority of the participants are shown clustered together on the plot map, while cards that were never placed in the same pile are located far away from each other on the map. Individual cards that were placed in different piles by the various cases will be “pulled” somewhere between those relevant clusters.

Dimensions tend to show a higher-level categorical division from the data. Cards/clusters along one dimension tend to show reasons why certain cards/piles are considered similar. For example, consider a sort task that included different types of mammals and reptiles. Subjects may divide the mammals into various piles (two-legged, four-legged, etc) and reptiles into different piles; the MDS map might show reptiles leaning toward one dimension (X-axis) while mammals are mostly located toward the Y-axis. This can be a useful clue in interpreting a hierarchical categorization of the results. It is important to note that the actual location among specific dimensions on the map is meaningless. The resulting plot can be rotated among the dimensions and quadrants and not change the interpretation in any way. It is the relative position among the plotted cards/clusters that represents the underlying data patterns.

This representation attempts to reproduce the data patterns as faithfully as possible. However, a perfect match is impossible. Stress and squared correlation values are used to indicate the distortion of the resulting map, otherwise known as "goodness of fit". Higher stress values indicate higher distortion and less accurate representation. Some stress will always be present; the issue is how much is considered unacceptable. Values less than 0.1 are quite ideal; values over 0.15 are considered by some to be not usable as the map output is too distorted to show accurate information (Multidimensional scaling, Stress, n.d.). Conversely, higher squared correlation results are better (as close to 1.0 as possible). Compounding this issue is the large number of variables that affect stress such as data error, level of similarity/dissimilarity in the data, number of dimensions plotted, number of data points, etc.

```
Iteration history for the 2 dimensional solution (in squared
distances)

                Young's S-stress formula 1 is used.

              Iteration     S-stress      Improvement

                  0           .12616
                  1           .11287
                  2           .10412          .00875
                  3           .09926          .00486
                  4           .09632          .00294
                  5           .09429          .00202
                  6           .09276          .00154
                  7           .09152          .00124
                  8           .09048          .00104
                  9           .08957          .00091

                     Iterations stopped because
               S-stress improvement is less than   .001000
```

```
Stress and squared correlation (RSQ) in distances

RSQ values are the proportion of variance of the scaled data
(disparities)
in the partition (row, matrix, or entire data) which is
accounted for by their corresponding distances. Stress values
are Kruskal's stress formula 1.

      Matrix Stress RSQ  Matrix Stress RSQ
        1     .073  .974    2     .118  .938
        3     .168  .903    4     .117  .933
        5     .091  .959    6     .117  .942
        7     .104  .949


        Averaged (rms) over  matrices
    Stress  =   .11591      RSQ =  .94252
```

Figure 3.6. MDS stress and squared correlation output.

There are two primary methods within SPSS to improve stress values. The ALSCAL algorithm is designed to repeat the calculation process (iterations) until stress improvement is less than 0.001. Occasionally it is necessary to increase the number of dimensions in the output, which will lower final stress values. Two dimensions are the most typical and adequate for most situations. Three dimensions are sometimes used, but it is more difficult to read (consider trying to see distances in a 3-dimensional cube on a flat sheet of paper). Four or more dimensions are practically impossible to interpret and are not generally useful.

Stress is somewhat dependent on the number of cases plotted (subjects in a card sort). Larger numbers of cases will typically result in a wider variation of data responses, therefore resulting in a less tidy map plot. This also depends a great deal on the type of data used for the sort task. Cards that represent a very concrete subject content, meaning a group of concepts that can only be sorted in a limited number of “correct” categories, should result in a more consistent map plot (but only if subjects “score” well). Consistent plots with highly clustered cards will have lower stress values. For example, if graduate music students were asked to sort composers into stylistic periods, the map should be fairly tightly clustered as most students should be able to put Bach into a Baroque pile and Mozart into a Classical period pile. Conversely, if 20 terms are presented for sorting that are based on personal preference and are less concrete, the data would probably show a wide variation in response categories. However, though in this situation clearly defined clusters may not be present, data points that lean along one axis may represent a general dimensional category underlying the data.

MDS also produces a weighted subject measurement that can indicate individuals within the plot that appeared to be different for whatever reason (hurrying through the research task, etc). If a particular case lies fairly distant from the others, it might indicate the need to eliminate it from the measurements.

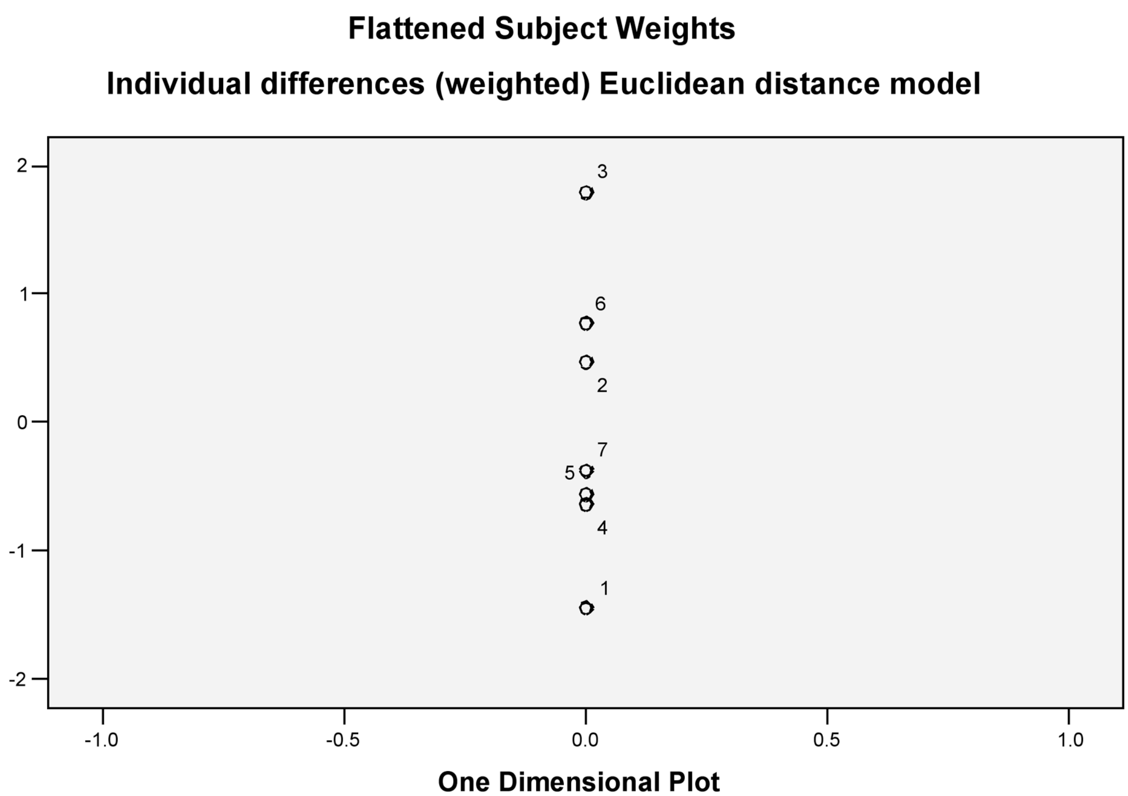

Figure 3.7. MDS weighted subject plot.

In spite of these numerous technical issues with MDS output and interpretation, Borg and Groenen (1997) summarize by stating "All of these criteria are mechanical ones. They are not sufficient for evaluating Stress, nor are they always important. Kruskal (1964a) writes: 'A second criterion lies in the interpretability of the coordinates. If the m-dimensional solution provides a satisfying interpretation…it may be well to use [it].' (p. 45). Thus, while the researcher should keep an eye on stress results and overall goodness of fit, this should not override simply looking at the maps to determine if it makes sense based on the input data and subject matter.

Interpreting MDS maps

Is the MDS map valid?

When interpreting an MDS map one should look for the following to ensure the map is a valid representation of the data:

- Low stress values (closer to 0.0 is best)
- High squared correlation (RSQ) values (close to 1.0 is best)
- Are there clusters on the map? Do they make sense? Clusters indicate categories representing similarity among certain cards. Lack of clusters, or clusters that inaccurately represent the data, may indicate an incorrect MDS data input and/or solution.
- Are there data points that lie along one particular axis or dimension of the map? These can indicate underlying principles for how the data was sorted by participants.

Are all subjects (cases) valid?

- Look for "goodness of fit" among the individual subject (case) data.
- Derived and flattened subject weights show which individuals may lie outside the norm.
- Higher stress values for those individuals also indicate poor fit.

Reading the MDS map

- Clearly-defined sort piles with no overlap among subjects will be clustered in distant areas of the map.
- Cards which may end up in different piles will be "pulled" somewhere in between. Sometimes this means they actually fit in both piles, while other times it may mean the participants' mental models are either inadequate, inaccurate, etc. This can also be caused by an individual who does not focus adequately on the sort exercise.

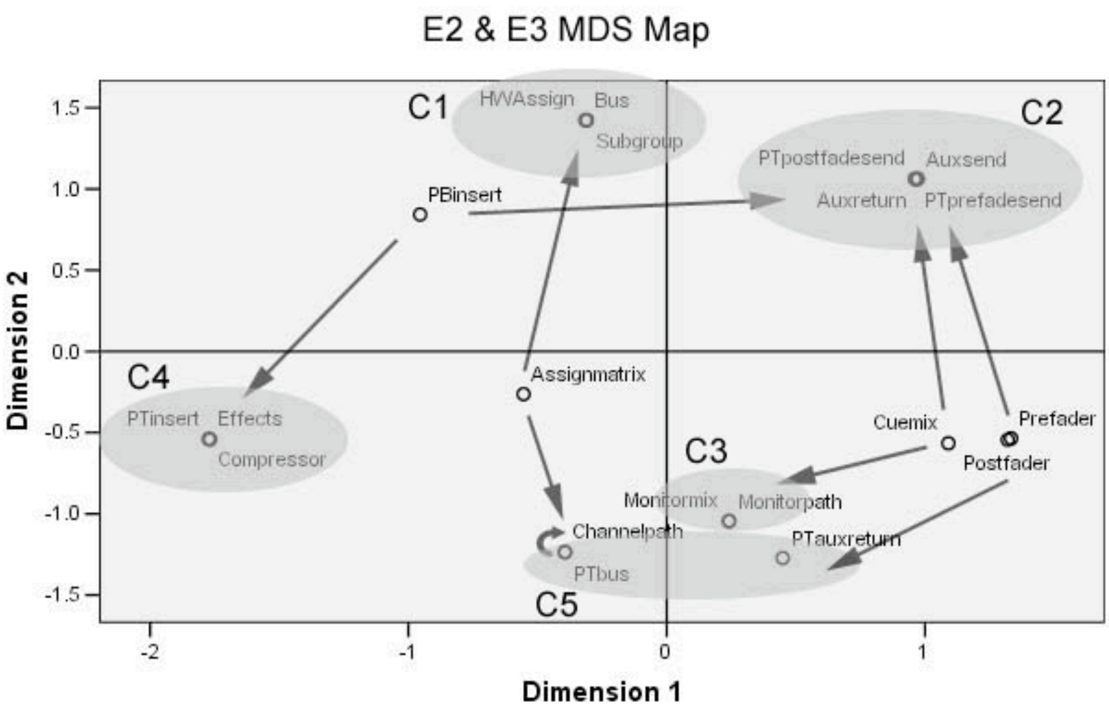

Figure 3.8. MDS map showing outlier cards that were placed in different stacks, shown being "pulled" between both clusters. Note that the shaded clusters were manually added and labeled during analysis and not generated by the MDS process.

This demonstrates the need for background information on the subjects whenever possible in the attempt to explain anomalies. For example, in this study the experts all had different backgrounds and job experiences. This proved valuable in explaining variations in responses on the sort exercise. It was also valuable to visually examine the individual sort results provided by the JPEG exports from the sort software. This becomes overly cumbersome in a study involving a large number of individuals, but for this study it was both simple and very instructive to examine the ten sorts from the experts and students.

MDS procedures for this study

The MDS calculation method employed for this study was ALSCAL. For processing a summed matrix of individual cases (such as for creating an all-student map), an option was selected that invokes the INDSCAL algorithm to provide a weighted, individual subject comparison. Using the individual dissimilarity matrices saved as SPSS files, a new file was created by opening the first of the individual cases. Then each of the other case files (either expert or student) were added using the *Data>Merge Files>Add Cases* command. SPSS sums these while vertically stacking them in the data view window. To run the MDS procedure, *Analyze>Scale>Multidimensional Scaling (ALSCAL)* was selected from the SPSS menu. All variables were selected in the left window and transferred to the "Variables" window. Specific configuration selections were as follows:

- Distances: *Data are distances* / *Square symmetric*
- Model:
    - Level of measurement: *Ordinal / Untie tied observations*
    - Conditionality: *Matrix*
    - Dimensions: 2 x 2
    - Scaling Model: *Individual differences Eucludean distance*
- Options:
    - Display: *Group plots, Individual subject plots, Model and options summary*
    - Default criteria values were used

Output information includes the following:

- Alscal procedure options (summary of the configuration options selected)
    - Stress / squared correlation information
    - Map coordinates for each card

- Subject weight values
- "Derived stimulus configuration" (main MDS map)
- Derived and flattened subject weights plots
- Scatterplots of linear fit (which helps show outliers as well)

Selected examples of these are included in the data analysis discussion in Chapter 4. Separate plots were run for experts and students in various groupings (all students, select students, etc). For processing a single case, i.e. a single expert, only one data file was opened in SPSS with no merging of multiple cases. The only difference in processing was that the "scaling model" was left as default and not set to "individual differences Eucludean distance".