Structural Variation in "Galileo" Space:

Effects of Varying the Criterion Pair in

Metric Multidimensional Scaling

by

Thomas F. Gordon

and

Henry C. De Leo

October 1975

Temple University
School of Communication & Theatre
Philadelphia, Pa. 19122

As a viable technique for examining cognitive/communication related processes, multidimensional scaling (MDS) has been receiving increasing attention (Gregson, 1975; Gillham and Woelfel, 1975; Shepard, Romney, and Nerlove, 1972; Woelfel, 1974ab). The objective of the present study was to empirically assess the effects of varying the criterion pair provided to subjects in a ratio judgment metric MDS task. Specifically, the study questioned whether the spatial representation of structure provided by version three of the Galileo metric MDS program would vary significantly as the criterion pair is varied.

The Galileo program, as adapted from Torgerson's (1958, chap. 11) classical metric MDS procedure by Woelfel and Serota, possesses three major characteristics (Woelfel, 1974b, p. 8). First, the methodology takes large arrays of data and facilitates interpretation by plotting the n stimuli or concepts in k orthogonal dimensions where k < n. Second, no information is lost in mapping of dissimilarity judgments onto a multidimensional space since the mapping is one-to-one. And third, "the function which maps discrepancies . . . can be seen to conform in essential respects to the spatial coordinate system of classical (and modern) mechanics." (Woelfel, 1974b, p. 8.)

The dissimilarity judgments made by the subject for entry into this program take the form of paired comparisons. To establish a base from which judgments are made, a criterion pair is defined as being a specified number of units apart. Woelfel (1974a, pp. 15-17) recommends the following to begin the judgment process: "If x and y are u units apart, how far apart are a and b?" The key advantages of this technique are expressed by

Woelfel as: (a) the resultant scale is continuous over its entire range and is unbounded at its high end; (b) the scale is a ratio measure and as such permits the social scientist the use of all standard arithmetic operations; (c) the experimenter provides the basis of measurement and may maintain the use of that basis across samples and across time.

From the above, it is apparent that the initial criterion pair provided to begin the judgment process is an important operation. Barnettt (1972), Danes and Woelfel (1975), and Gillham (1972) have demonstrated the reliability of applying a single criterion pair to a set of concepts over time, however, different studies have used different criterion pairs. Woelfel (1974a, p. 16) suggests certain criteria for selection of a comparative standard:

> First, the standard should be relatively stable. Changes in the standard over time can confound time series measurements and prevent meaningful comparisons of measurements made at different times. Secondly, the standard should be the same for all observers regardless of reference point, i.e., two independent observers must both agree on the length, for example, of a meter or a kilometer. Less important, but nonetheless worthy of consideration, good practice for minimum error suggests using a standard approximately midway between the largest and smallest measurement likely to be encountered, (measurement of astronomical distances in miles, for example, is cumbersome, as would be measurement of terrestrial distances in fractions of light-years).

Berlin and Kay (1969, p. 104) have suggested that red and white are fairly invariant concepts in a cross cultural analysis. Several ratio data MDS studies have used red and white as a criterion pair and have specified that they were either 10 "units" apart or 10 "Galileos" apart (Barnett, 1972; Barnett, 1974; Barnett, 1975; Woelfel and Barnett, 1974). Other studies have used a criterion pair from within the concept domain, for example, Barnett, Serota and Taylor (1974) applied a 50 unit seperation to the evaluation of 16 political concepts.

In the present study, four major manipulations of the criterion pair

were applied to the same set of concepts and the resulting structures, as spatial representations in three-dimensional orthogonal space, were compared. In each case, a 10 unit base was applied to the criterion pair. It appeared to us that the Red-White criterion pair used in the studies cited above was a criterion that would be relevant only to the judgment of colors. As such, we did not expect this distance assessment to be functional in concept domains unrelated to color and the following was offered as the major hypothesis for testing:

H_1: The No Criterion treatment will produce a structure equivalent to the Red-White treatment.

To further explore the effects of criterion variation but without forming specific hypotheses, treatment pairs were selected which represented (a) the extremes from the concept domain and (b) two concepts close together in the concept domain. These treatment conditions are detailed in the methodology section.

Methodology

To examine the effects of varying the standard criterion pair given to subjects for making ratio scale distance judgments in a metric MDS routine, the FORTRAN IV program utilized was version 3.0 of Galileo developed by Joseph Woelfel and Kim Serota at Michigan State University.

Concepts. Since the measurement criterion was the primary concern in this study, the set of concepts to be judged were secondary. The concepts chosen for judgment were radio stations in the Philadelphia market area, more particularly, the music formats of the stations. It was thought that on the basis of general familiarity, subjects would be able to judge the degree of similarity or difference among the major stations. The stations were selected on the basis of high American

Research Bureau May 1975 ratings for the Philadelphia metropolitan area. In addition to the stations, the basic types of music formats became concepts to be judged, and the concept of "me" (or self) was included. On the data collection instrument, the total set of 15 items were presented in all possible pairs (105) and subjects were asked to judge the distance between the concepts on the basis of music formats only. The music formats of the stations (identified only by call letters) were not defined for the subjects. The total set of concepts were:

I1 Beautiful Music	6. WCAU-FM	11. WIP-AM
2. Oldies	7. WDVR-FM	12. WMMR-FM
3. Top 40	8. WFIL-AM	13. WPEN-AM-FM
4. Middle of the Road	9. WIBG-AM	14. WWSH-FM
5. Rock	10. WIOQ-FM	15. Me.

Treatment Conditions. The four criterion treatment conditions were:

1. Red-White

2. Oldies-Top 40

3. Beautiful Music-Rock

4. No Criterion.

In each treatment independent samples of subjects judged the concepts. In the first three cases Ss were told that the criterion pair was 10 units apart. In the last condition no criterion pair was given, but to keep the general scale base comparable, Ss were told, "As you judge the distances, keep a 10 point scale in mind -- some concepts may be less than ten units apart and other may be more."

The Red-White criterion was included because of its previous use, as noted earlier. The criterion pairs from within the concept domain were selected with two major considerations in mind. First, the music formats seemed to represent a crude continuum ranging from Rock to Top 40, Oldies, Middle of the Road, and Beautiful Music. Thus, the Beautiful

Music-Rock pair represented the extremes. Oldies-Top 40 represented a closer pair which would permit judgments greater than and less than the criterion distance. It was expected that the No Criterion pair would produce a structure equivalent to the Red-White pair. The rationale, here, was that red and white as concepts would be seen as irrelevant to the music format domain. Therefore, designating these concepts as a certain number of units apart would be essentially the same as providing no concept anchors while telling the Ss to keep this scale base in mind.

Subjects. Two hundred and fifty-four undergraduate students attending the first summer session at Temple University were the subjects. The students, from a variety of departments, were randomly assigned (by classes) to treatment conditions. The departments sampled included Anthropology, Business, Education, Journalism, Psychology, Radio-TV-Film, Speech, and Theatre. Since it was summer session class enrollments were relatively small, ranging from six to twenty-five students. Consequently, in order to obtain the 60-plus subjects per treatment several classes were needed for each treatment. The random assignment of classes to treatments, therefore, produced highly equivalent groupings of subjects in each treatment relative to the demographics of (a) year in school, (b) age, (c) sex, (d) race, (e) income, and (f) hours of radio listening per day (see Appendix A).

Procedure. The data were collected during the month of June 1975. Subjects in classrooms were handed one of four questionnaires and the following instructions were read with them:

> This questionnaire askes you to tell us how different (or in other words, how "far apart") concepts are from each other. Difference between concepts can be measured in units, so that the more different two concepts are, the more units apart they are from each other. To help you know how big a unit is, _____ and _____ are 10 units apart.

You are supposed to tell us how may units apart the concepts on the next few pages are from each other. Remember, the more different the two concepts are from each other, the bigger the number of units apart they are. If you think any of the two concepts are more different than _____ and _____, write a number bigger than 10. If you think they are not so different, use a smaller number. Remember, the more different the concepts are from each other, the higher the number you would write.

(Blanks above were filled by the criterion pair used in a particular treatment.) After reading of the instructions, the Ss were told: First, that the concept "me" on the questionnaire should be taken to mean themselves and judgments involving "me" should indicate how close they feel to that type of music or that particular station. Second, pairs involving stations with which they are not familiar should be left blank. Third, that zero is a legitimate response; and if they see two things as identical then those two items would be zero distance apart. Fourth, they were asked to work as quickly as possible, doing the items as pairs rather than attempting to ponder the interrelationships of the total set. On the average, the items were completed in 15-20 minutes. If, during the data collection respondents asked for a definition of one of the basic music formats, they were told to respond with their own conception of what the music format was.

Of the 105 distance estimates possible, few of the Ss were familiar enough with all station call letters to provide judgments on the total set. The subget of concepts which included only the basic music formats and "me" received the highest response rate since nearly all subjects were able to judge these. The numbers of Ss judging any particular pair ranged from 14 to 54. On the basic music formats the range was from 49 to 62.

Results

The mean distance between concepts was computed through the Galileo program which sums every subject's judgment for each concept pair (i.e., every cell in a 15 x 15 symmetric matrix) and divides by the total number of judgments for that pair. In addition, a three dimensional plot of the concepts (which are actually located in an n dimensional space), and the percentage of variance explained by the three largest dimensions are provided for each of the four treatments. By a least squares rotation of the axes of treatment two, three and four to treatment one's axes (arbitrarily defined as the mainspace) each of the four spaces were located around a common origin (defined by the centroid of the treatment one space) and thus the same XYZ axes (see Figure I). This rotation greatly facilitates comparison and interpretation of multiple plots and possesses great potential for use in time series studies.

Figure I A & B about here

The analysis of the structures produced by the four treatment conditions involved two steps. First, the overall structures of the four treatments were examined for significant differences. And second, the four treatments were examined for the variability or stability of the structures. Further, each dimension of each structure was analyzed for variability.

To compare the four treatment structures, the mean distance matrices were used. There were 105 mean distances for each of the four treatments, these cell values being invariant in rotation. Using the values in these cells as "scores" in a one way analysis of variance for all four treatment groups, a significant difference was evident ($F = 64.45$; $df = 3,416$; $p < .001$, see Table I). Inspection of the seperate treatment means indicated that the Oldies-Top 40 (OT) treatment produced the deviant structure

while the Red-White (RW), Beautiful Music-Rock (BR), and No Criterion (NC) treatments were very similar. To confirm this, an analysis of variance using these three latter treatments was computed, showing no significant difference in overall structures (F = 2.23; df = 2,312; p <.05, see Table II). On the basis of structure then, the treatment group given Oldies-Top 40 as the criterion pair produced a structure which was significantly different from each of the other treatment groups while these latter three did not differ from each other.

Tables I and II about here

To examine variability or stability, the variances of the cells with the largest ns (i.e., for each treatment the fifteen pairs of concepts involving the basic types of music formats and "me") were examined as the most accurate indicators of variability associated with the treatment conditions. Using these variances as "scores" in the one way analysis of variance across the four treatment groups produced a significant difference (F = 6.98; df = 3,56; p <.001, see Table III). As with the structural analysis of means, it appeared that the OT treatment was the source of the difference. Again, to verify this the OT treatment was removed and the recomputed analysis of variance for RW, BR, and NC showed no significant difference (F = 1.72; df = 2,40; p <.05, see Table IV). It can be concluded from these analyses (using the concepts with the largest ns) that treatments RW, BR, and NC had the same variability, that is, changing the criterion pair did not significantly increase the variability of judgments.

Tables III and IV about here

In order to further isolate the variability evident in the OT treatment, the individual orthogonal factor dimensions were examined

across the four treatments. To do this the values given for each concept on the first dimension of treatments one thru four were compared through Bartlett's test for homogeniety of variance. Then, the values on dimension two were compared across the four treatments -- as a seperate analysis -- followed by the same analysis for the third dimension. The comparison of the four treatments on dimension 1 revealed no significant difference ($X^2 = 6.60$, df = 3). Similarly, dimension 3 was found to be homogeneous across the four treatments ($X^2 = 6.48$, df = 3). However, dimension 2 was found to be significantly different across the four treatments ($X^2 = 9.35$; df = 3; p < .05). Again, the OT treatment was excluded from the Bartlett's test verifying that there were no significant differences among RW, BR, and NC on the second dimension ($X^2 = 0.12$, df = 2). The significant variability, then, was contributed entirely by the concepts most closely associated with the second dimension of the OT treatment group. This variability is clearly evident in the Y-Z Plane of Figure IB relative to the concepts of "rock" music, "WMMR", and the collective "me." It appears that because the Oldies-Top 40 concepts are relatively close together, a greater number of judgements longer than that 10 unit criterion were generated and, subsequently, variability was increased.

Table V about here

Finally, a note on the face or content validity of the three dimensional space in Figure I. Over the four treatments, the three dimensions

accounted for 74.91 to 80.36 percent of the variance explained. As
well, empirically, we know that WDVR-FM and WWSH-FM are beautiful music
stations, WCAU-FM and WPEN-AM-FM are oldies stations, WIBG-AM and WFIL-
AM are top 40, WIP-AM is middle of the road, WMMR-FM is rock, and WIOQ-
FM is a mixture of rock and contemporary top 40. Thus, the station call
letters, whose music formats were not defined for the Ss, are oriented
in the space close to their actual formats. Similarly, aside from the
pairing, Ss were asked to specify their favorite radio station. For this
sample WMMR was the most frequently mentioned, receiving 82 mentions as
most prefered with the next most prefered (WIOQ) receiving only 17.
Thus, on face validity, the stations are oriented around the appropriate
music formats and the self or "me" is closest to the most prefered station.

Discussion

The results of this study confirm the hypothesis that a no criterion
treatment will produce a spatial representation equivalent to a red-
white criterion (for concepts other than color). However, more interestingly,
they demonstrate that a criterion pair involving the extremes will produce
a structure statistically identical to both of these. This is probably
the case because in the no criterion treatment the subject was simply told
to keep a 10 point scale in mind while making the dissimilarity judgments,
that some concepts may be less than that and some may be more. Giving the
subject a conceptually irrelevant pair like red and white apparently
achieved the same effect in that the only useful information for the
judgment problem was the 10 unit base. With the extreme pair, the concepts
were meaningful but the fact that they represented the extremes produced
the same effect by essentially limiting the judgments to that 10 unit base.

The criterion pair involving concepts that were close together in
the concept domain (Oldies-Top 40) produced the significantly different

structure. The mean distances indicate that for that treatment more estimates had to be greater than 10 units and this contributed to the difference in structural distance, and to the greater variability of the judgments as compared to the judgments produced by the other criterion pairs.

To return to Woelfel's suggestions for the selection of a criterion pair (1974a), these results shed empirical light on the proposition that ". . . good practice for minimum error suggests using a standard approximately between the largest and smallest measurement likely to be encountered. . . " (p. 16). This certainly holds when the concepts are too close together, forcing most of the judgments outside the criterion distance. However, it seems that the extreme or near extreme pair would produce the least variable judgments if we assume a sufficiently large distance between the pair such that the concepts are not forced into a restricted space and also assume a relatively homogeneous set of concepts. In the present study, the structure involving the extreme criterion pair allowed the greatest variability to be accounted for by the three orthogonal dimensions (80.4%), while the No Criterion condition had the least variance explained (74.9%).

The problem, of course, is that if the extremes are routinely used, the scale becomes essentially interval rather than ratio, violating an important assumption of the Galileo system. It may be that functionally the subject is actually operating on "social science" concepts in an interval fashion and the ratio assumption, though theoretically important, is not a crucial assumption to the practical application of this methodology. This may be especially true in light of the averaging process involved in the structuring.

If the criterion pair involving the extremes proves to consistently

produce the judgments with the least variability, it is still theoretically possible to use these as the criterion concepts and yet maintain the open ended "ratio" assumption. Obviously, further studies exploring these questions, replicating the present findings, and further testing the criterion assumption are called for. For example, will structures produced by subjects responding relative to 10, 50, or 100 unit standards be equivalent in shape and stability? So far, the methodology shows exceptional promise and answering these basic questions should become priority items for those interested in its application.

References

Barnett, G. A. Reliability and metric multidimensional scaling. Unpublished research report, East Lansing, Michigan State University, 1972.

Barnett, G. A. Social system homophily as a function of communication. Paper presented at the annual meeting of the International Communication Association, New Orleans, 1974.

Barnett, G. A. The diffusion of complex innovations. Unpublished research report, East Lansing, Michigan State University, 1975.

Barnett, G. A., Serota, K. B., & Taylor, J. A. A method for political communication research. Paper presented at the annual meeting of the International Communication Association, Chicago, 1975.

Berlin, B., & Kay, P. Basic color terms: their universality and evolution. Berkeley, Ca.: University of California Press, 1969.

Danes, J. E.,& Woelfel, J. An alternative to the "traditional" scaling paradigm in mass communication research: multidimensional reduction of ratio judgment of seperation. Paper presented at the annual meeting of the International Communication Association, Chicago, 1975.

Gillham, J. & Woelfel, J. The Galileo system of measurement: preliminary evidence for precision, stability and equivalence to traditional measures. Unpublished research report, Akron, Ohio, University of Akron, 1975.

Gregson, R. A. M. Psychometrics of similarity. New York: Academic Press, 1975.

Shepard, R., Romney, A., & Nerlove, S. (Eds.) Multidimensional scaling: theory and applications in the behavioral sciences. New York: Seminar Press, 1972.

Torgerson, W. S. Theory and methods of scaling. New York: John Wiley and Sons, Inc., 1958.

Woelfel, J. Procedures for the precise measurement of cultural processes. Unpublished manuscript, East Lansing, Michigan State University, 1974. (a)

Woelfel, J. Metric measurement of cultural processes. Paper presented at the annual meeting of the Speech Communication Association, Chicago, 1974. (b)

Woelfel, J., & Barnett, G. A. A paradigm for mass communication research. Paper presented at the annual convention of the International Communication Association, New Orleans, 1974.

Table I

One Way ANOVA of Mean Intercell Distances

for Treatments RW, OT, DR, NC and Concepts 1-15

Source	df	SS	MS	f
Total	419	7894.642		
Between	3	2504.998	834.999	64.449*
Within	416	. 5389.644	12.956	

*F ratio significant at p< .001

Table II

One Way ANOVA of Mean Intercell Distances

for Treatments RW, BR, NC and Concepts 1-15

Source	df	SS	MS	F
Total	314	2685.794		
Between	2	37.901	18.951	2.233*
Within	312	2647.892	8.487	

*F ratio not significant at $p < .05$

Table III

One Way ANOVA of Intercell Variances

for Treatments RW, OT, BR, NC and Concepts 10-15

Source	df	SS	MS	F
Total	59	1755535.830		
Between	3	477533.429	159177.810	6.975 *
Within	56	1278002.401	22821.471	

*F ratio significant at p< .001

Table IV

One Way ANOVA of Intercell Variances

for Treatments RW, BR, NC and Concepts 10-15

SOURCE	df	SS	MS	F
Total	44	582741.302		
Between	2	44126.093	22063.046	1.720*
Within	42	538615.209	12824.172	

*F ratio not significant

Table V

Bartlett's Test for Homogeniety of Variance

Dimension 1	Dimension 2	Dimension 3
Treatments	Treatments	Treatments
RW, OT, BR, NC	RW, OT, BR, NC	RW, OT, BR, NC
$x^2 = 6.6017*$	$x^2 = 9.3488**$	$x^2 = 6.4838*$
df = 3	df = 3	df = 3
	Treatments	
	RW, BR, NC	
	$x^2 = 0.1174*$	
	df = 2	

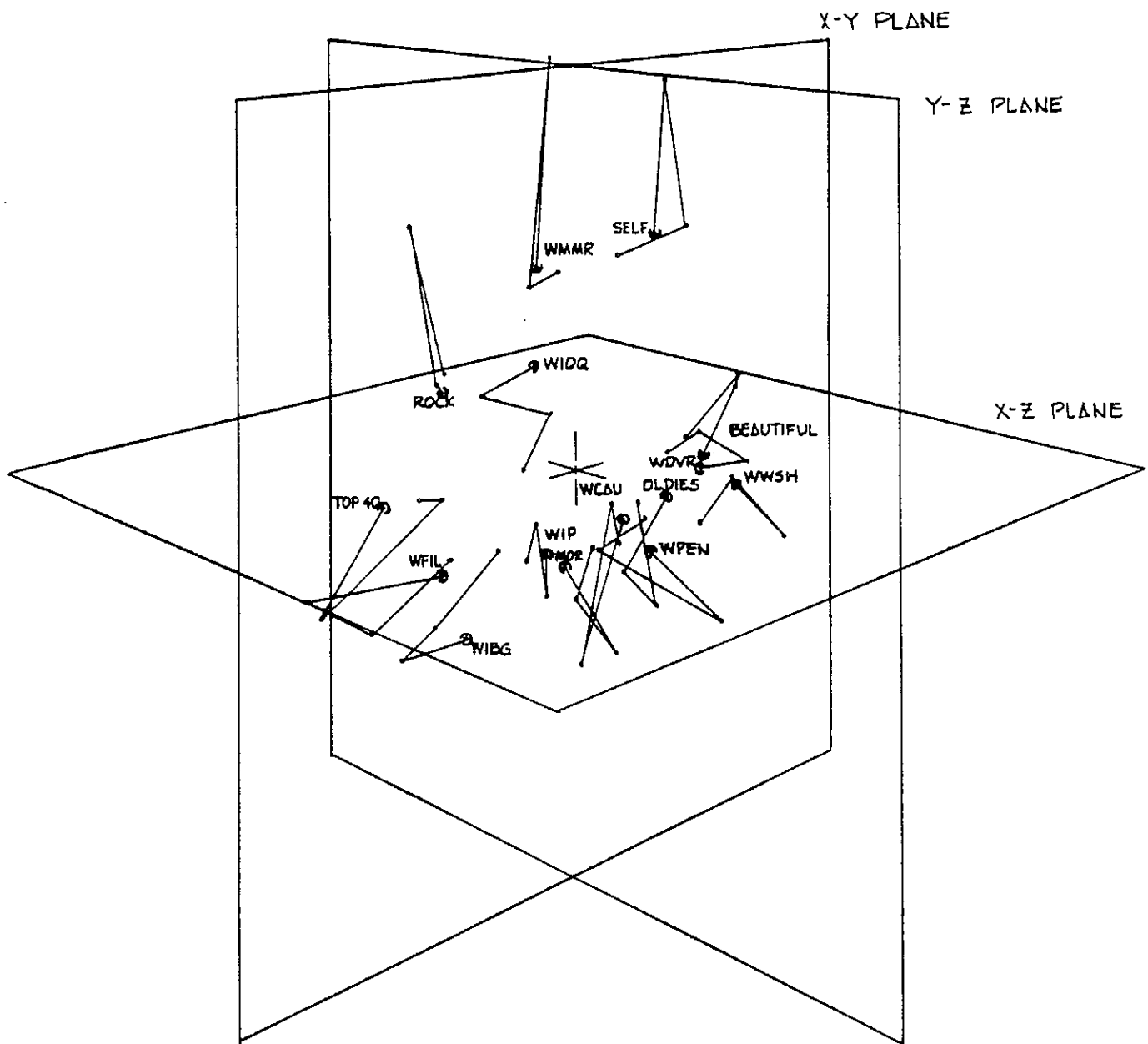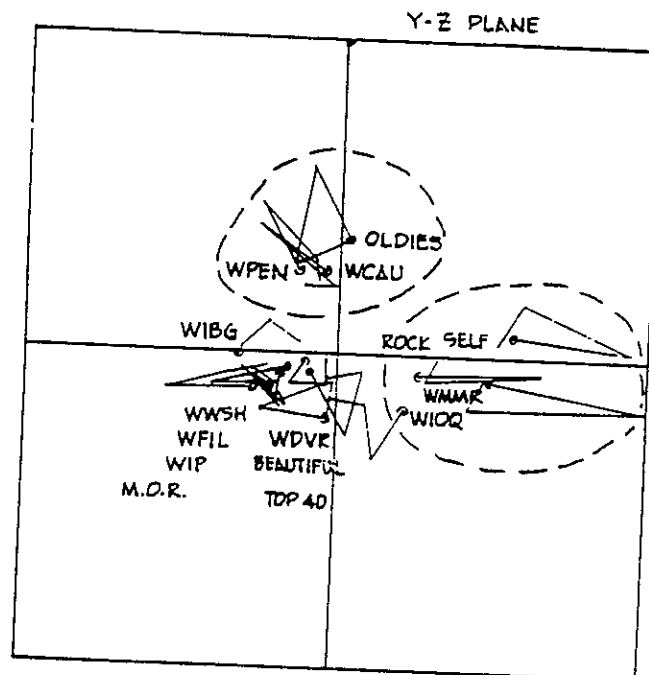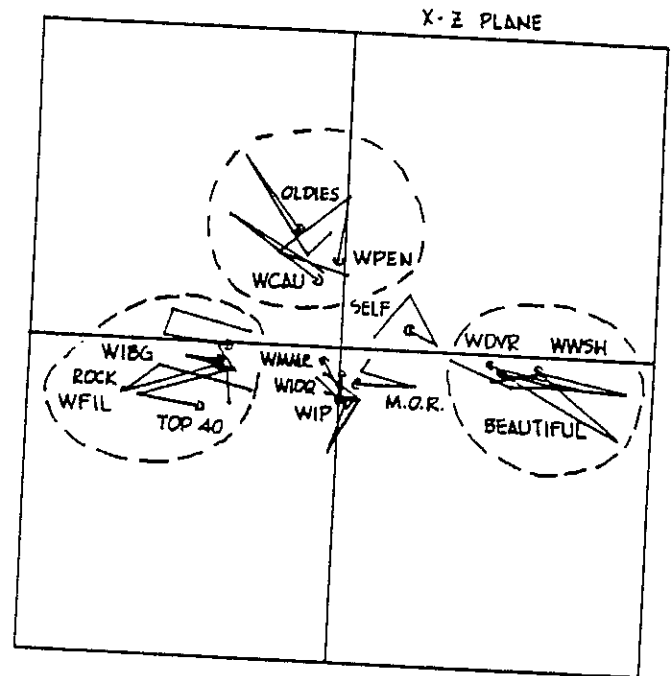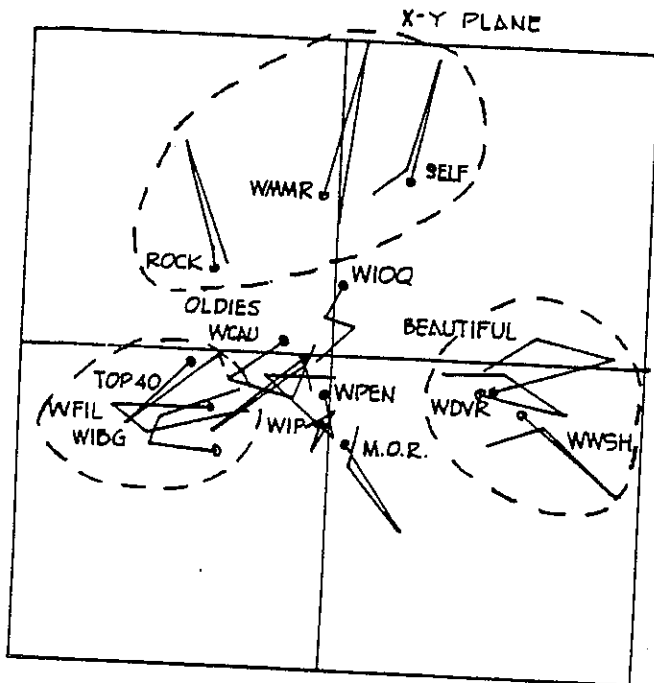$*x^2$ not significant

$**x^2$ significant at $p < .05$

Figure IA

Plot of First Three of n-Dimensions

Figure IB

Individual Planes from Three-Dimensional Plot

TREATMENTS	YEAR				AGE			SEX		RACE		
	Fr.	Soph.	Jr.	Sr.	18-25	26-35	36+	Male	Female	Black	White	Other
Beautiful - Rock	4	8	17	37	55	11	0	33	33	9	50	6
Top 40 - Oldies	0	13	15	33	46	13	4	27	35	8	46	1
Red - White	3	14	22	26	44	16	2	35	30	13	36	1
None	3	5	19	35	46	13	4	37	24	12	40	2

TREATMENTS	AVERAGE FAMILY INCOME/YEAR					
	0-4,999	4,000-7,999	8,000-9,999	10,000-14,999	15,000-20,000	20,000+
Beautiful - Rock	6	9	5	15	8	18
Top 40 - Oldies	7	3	3	8	14	21
Red - White	5	7	2	21	12	14
None	9	7	2	14	10	14

TREATMENTS	AVE. HRS. OF RADIO LISTENING/DAY				
	1.9	2-2.9	3-3.9	4-4.9	5+
Beautiful - Rock	24	15	12	4	10
Top 40 - Oldies	21	15	9	8	9
Red - White	14	22	6	3	11
None	24	13	10	4	8

* Cross tabs of subjects relative to treatments and demographic characteristics.