Chapter 7

Free Classification of Dialects

7.1 Introduction

While it is certainly useful to study explicit labeling of dialects in an *n*-alternative forced-choice task, we are also interested in what other distinctions people might make when perceiving dialectal variation. Certainly, most *n*-alternative forced-choice tasks carefully select stimuli and present listeners with selection options based on known dialect regions. But speech and dialects represent social acts, and, as such, are wrapped up in the social landscape of the regions where they take place. When people think about dialects, they think about more than just language. Even on a task that is explicitly about where dialects are spoken, many responses on the draw-a-map task presented in Chapter 4 were about characteristics of the talkers. Commenting on some of the earliest results using this method, Preston (1986) notes that "it is clear that the informants took this geographical task to be an evaluative rather than descriptive one" (238). That is to say, whatever it is people are doing when they hear speech, they are evaluating the talker at the same time as they are processing the speech sounds.

A technique that gives listeners more freedom to categorize talkers without being constrained by a fixed number of prelabeled categories is called *free classification* (Imai, 1966). Asking a similar question as asked with the draw-a-map task, Tamasi (2003) gave participants 50 index cards with the names of the states of the United States written on them and asked participants to sort the cards according to where people speak differently from one another. While this task limits the granularity of variation to the state level, it allows people to ignore the geographic proximity of individual states. Thus, it was possible for a group of Texas and Oklahoma—neighboring states—to include Wyoming. The connection between the states appears to be "cowboy" speech, and, indeed, major universities in each state have relevant mascots: Texas Longhorns, Oklahoma State Cowboys, and Wyoming Cowboys. Participants sorted the states into groups of 5 to 35 with a mean of 14. About 6 or 7 consistent dialect regions were formed. This is slightly fewer than the 8–12 dialect regions participants in Preston (1986) drew on their maps.

While Preston's draw-a-map task and Tamasi's card sorting task provide people with the freedom to classify dialects according to their own conceptions of the makeup of the dialect space, all of the judgments are purely in the mind of the respondents; none of the judgments are based on classifying actual speech. Clopper and Pisoni (2007) used a method similar to Tamasi (2003) but replaced the state names with recordings of people from various dialect regions. Participants were instructed to listen to the recordings and group the talkers according to where they believe they are from. With 66 stimuli from 6 dialect regions, participants made between 3 and 30 groups with an average of 10 and a median of 7. While participants were sensitive to geographic characteristics, especially a north/south dimension, they also evaluated the dialects based on a standard/marked dimension. As in their force-choice task with similar stimuli (Clopper, 2004; Clopper & Pisoni, 2006a), participants perceived three distinct dialect groups: Northeast, South, and Midwest/West. The geographic provenance and mobility of the listeners affected their classification patterns, though. Mobile listeners, those who lived in at least two dialect regions before the age of 18, responded very similarly whether they were from the North or Midland regions. However, the non-mobile listeners grouped their own dialect region closer to the North talkers if they were from the North or closer to the South talkers if they were from the Midland. This suggests that dialect representations vary based on the prior linguistic experience of the individual, and that these patterns of individual perceptions are expressed at the group level. Such a finding echos Preston's observation that people draw more detailed regions for areas closer to their home area on perceptual dialect maps (Preston, 1986). Not only do people have more finely graded mental perceptions of variation closer to home, who they consider to be part of their home region shows gradation, too.

The following experiment builds on the earlier work exploring the effects of linguistic experience on the perceptual classification of spoken examples of dialectal variation (Clopper, 2004; Clopper & Pisoni, 2006a). In this experiment, all of the listeners are non-mobile and lifelong residents of their home region. In addition, the recorded tokens are only from nearby dialect regions, dialects that participants have likely come in contact with at some point due to their relatively close geographic proximity.

7.2 Methods

7.2.1 Listeners

Participants in this part of the study are drawn from the participants in the state-wide study (see Chapter 5). One participant from southern Indiana failed to complete the task and was excluded. The total number of participants was 108. No participants were excluded based on other performance measures.

7.2.2 Talkers

Twenty-four female talkers from the Indiana Speech Project Corpus (Clopper et al., 2002) were selected, four from each of the six regions in the corpus (near Chicago, Fort Wayne, Indianapolis, Bloomington, near Louisville, and Evansville). All talkers were in their early 20s at the time of the recording. The talkers in this experiment are identical to those from the forced-choice categorization task.

7.2.3 Stimulus Materials

The stimulus materials consisted of eight sentences repeated by each of the six talker regions for a total of forty-eight unique tokens. The sentences are a subset of the materials used in the 4AFC task. Half of the sentences contained phonemes that are stereotypically associated with the North (/ae/) or South (/ay/) dialect regions while the other half did not contain any of these stereotyped phonemes in content words. A complete list of the stimulus materials used in this experiment are in Appendix V. Sentences were randomly assigned to talkers, and talkers read one sentence each from the stereotype and non-stereotype categories. All of the stimuli were processed in the same way as in the forced-choice task, cropped to include only speech material and leveled to the same average intensity.

7.2.4 Procedure

Participants were seated in front of a computer equipped with a mouse, a trackpad, and circumaural headphones. The experiment was constructed in Microsoft Powerpoint based on Clopper (2008). At the beginning of the experiment, participants saw four columns of numbered black boxes on the right side of the screen and a 16x16 grid on the left. An example of the screen in the initial state is shown in Figure 7.1. Stimuli could be heard by clicking on the black boxes. Participants were able to listen to the stimuli as many times as they wanted. The boxes associated with the recordings could be moved using the mouse.

Participants were told the boxes represented talkers from different parts of Indiana. They were instructed to listen to the stimuli and put them into groups which contained talkers of the same dialect. Listeners were allowed to make as many groups as they wanted with however many talkers they believed belonged in each group, but groups needed to have at least two members. No strict time limit was imposed, however, participants were told that the task can be completed in 10–15 minutes. In pilot testing most participants

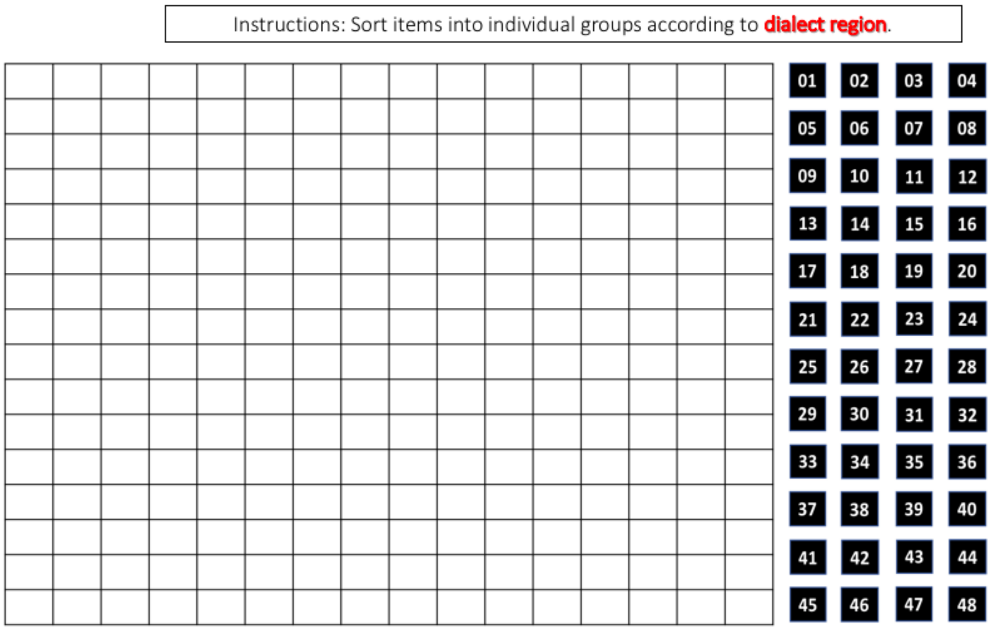

Figure 7.1: Free classification task in the initial state

finished within this time range; some participants, though, might still be working on the task if they had not been encouraged to cease. The explicit mention of the typical time for completion served as an anchor for participants, and most completed the task within 20 minutes. An example of the screen after the task was completed is shown in Figure 7.2.

7.2.5 Analysis Methods

The data were processed[1] by first transferring token groupings for each participant into an Excel spreadsheet. Each group of tokens was recorded on a unique row with the grouped tokens recorded in one cell according to the number associated with the token and each number separated by a comma. An example of the coding procedure is given in Table 7.1.

The data were transferred to the spreadsheet by a single coder, me. In order to ensure that all and only 48 tokens were counted per participant, a script was written to check this condition. Of the 167 participants whose data were coded, 30 failed the condition (data from

[1]Thanks to Ryan Lidster, Franziska Krüger, and Danielle Daidone for sharing their analysis procedures with me and a special thanks to Aaron Albin for writing the R script used to create the similarity matrix.

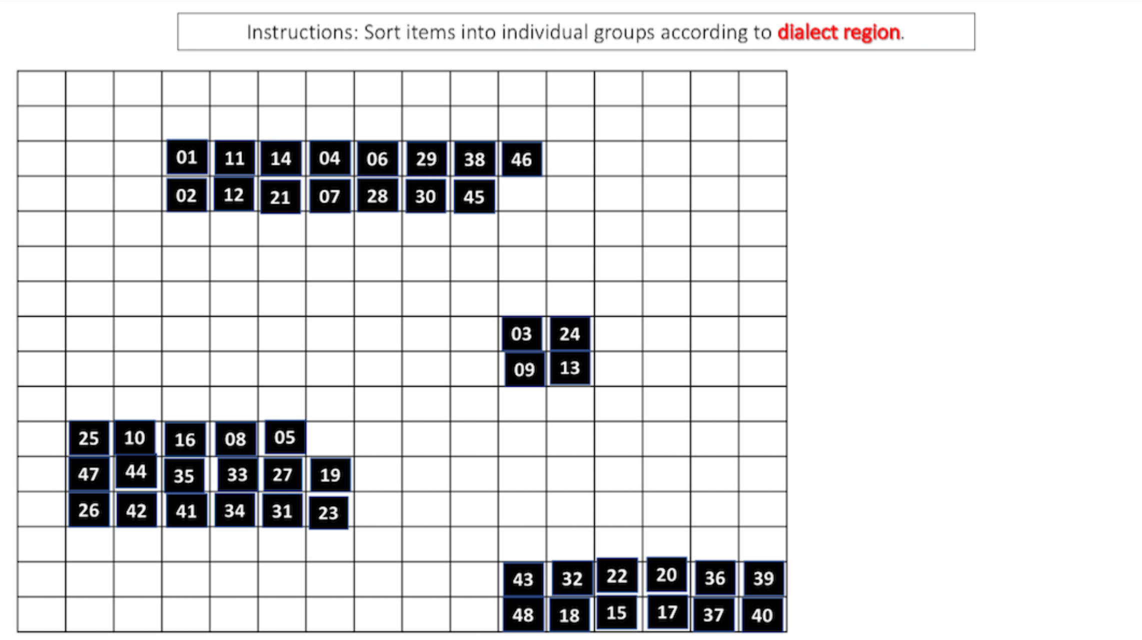

Figure 7.2: Free classification task in a completed state

all participants were coded even though this chapter only presents data from the subset that met the age and residential history requirements). Three types of errors were found in the data coding. The first type included typographical errors that caused the test condition to fail, such as typing two commas between tokens, and affected three participants. The second type of error was the most common, affecting 24 participants, in which the coder misread or mistyped the token numbers (e.g., coding 23 instead of 32). The third type of

Table 7.1: Example of spreadsheet structure used for coding groupings by participants

Subject	Group	Tokens
Inperc-32	1	1, 20, 30, 32, 44, 46
Inperc-32	2	29, 10, 12, 16, 21, 45, 38, 39
Inperc-32	3	34, 2, 6, 11, 18, 17, 24
Inperc-32	4	40, 9, 23, 25, 7, 5, 19
Inperc-32	5	42, 14, 47, 28, 13, 26
Inperc-32	6	36, 3, 4, 31, 35, 27, 22, 8, 43, 48
Inperc-32	7	33, 37, 41
Inperc-33	1	1, 3, 18, 19, 26, 41 ,43, 11
Inperc-33	2	2, 4, 8, 24, 35, 34, 28, 23, 22
...	...	...

error occurred in three participants' data. In these cases, the participant either duplicated a token and assigned the two identical tokens to different groups or failed to include a token among a group[2]. The first two types of errors which were the result of coder error were corrected. The third type of error which originated with the participant would have resulted in the exclusion of those participants' data from analysis in this particular task; however, none of these participants met the age or residency criteria for inclusion.

After the tokens were coded according to groups and participants, a script written in R by Aaron Albin, which I modified as appropriate, was used to construct similarity matrices. The script created a matrix by counting how many times each individual token occurred in a group with every other token. For example, Table 7.1 shows that for participant Inperc-32, Token 1 occurred with Token 20. However, participant Inperc-33 did not place Token 1 into a group with Token 20, so for the data show, Token 1 and Token 20 only co-occurred once. Across the entire data set, Tokens 1 and 20, spoken by talkers from Bloomington and from near Louisville, were included in the same group 66 times. Since the maximum number of times any two tokens were included in the same group is 89, the minimum is 4, and the median is 30, Tokens 1 and 20 are considered quite similar by the participants.

The analysis methods I use require distances (difference) rather than similarities, so I convert the similarities to relative distances using the following formula:

$$1 - (similarity\ count/max\ possible\ similarity)$$

where the *similarity count* is the number of times two tokens were included in the same group, *max possible similarity* is the maximum number of opportunities to tokens had to be

[2] As this experiment was constructed using Microsoft PowerPoint, all of the normal functions of the program were available to participants during the task. Participants were, therefore, able to copy and paste the tokens. Fortunately, this error only occurred rarely. In the cases where tokens were not assigned to a group, it was due to the participant having moved the token icon beyond the limits of the visible window. The token was hidden from view, forgotten by the participant, and not apparent to the experimenter following the completion of the task.

Table 7.2: Summary of participants' free classification behavior

	Mean	Minimum	Maximum	Median
Number of Groups	6.33	3	24	6
Number of Talkers Per Group	7.5	2	35	6

be grouped together (i.e., the total number of participants included in the analysis), and 1 represents the value of maximum dissimilarity. The resulting value is a decimal between 0 and 1 where 0 represents complete similarity and 1 represents complete dissimilarity.

I chose to base the maximum possible similarity on the total number of participants included in the analysis instead of the maximum number of times any two tokens co-occurred. The former gives a more accurate measure of the actual distance between two tokens whereas the latter gives a normalized distance. The normalized value would imply that the most similar pair of tokens are identical, having a distance of 0, whereas the non-normalized value shows that the most similar pair have a relative distance of 1 – (44/108) = 0.41 or that 41% of participants did not find the pair to be similar enough to place them into the same group.

7.3 Results

A summary of the participants' free classification behavior is shown in Table 7.2. On average participants made 6.3 groups of talkers with a range of 3 to 24 groups and a median of 6. The mean number of talkers per group was 7.5 with a range of 2 to 35 and a median of 6.

A summary of the number of groups created by participants in each listener group is shown in Table 7.3. The number of groups of talkers is fairly consistent across listener groups with averages ranging from 5.8 to 6.7 groups. Most of the groups had maximum numbers of groups between 10 and 16 (one participant from the Northwest group grouped

Table 7.3: Descriptive statistics on the number of talker groups produced by each listener group

	Mean	Minimum	Maximum	Median
Northwest (all)	6.72	3	24	5
Northwest (B)	6.20	3	11	5
Northwest (W)	6.37	3	16	5.5
Northeast	5.80	4	10	6
Central	6.60	3	12	5.5
South	5.87	3	13	6

Table 7.4: Descriptive statistics on the number of talkers per group produced by each listener group

	Mean	Minimum	Maximum	Median
Northwest (all)	7.14	2	30	6
Northwest (B)	7.74	2	19	6
Northwest (W)	7.54	2	30	6
Northeast	8.26	2	23	8
Central	7.27	2	35	5
South	8.18	2	32	7

pairs of tokens resulting in 24 groups). A one-way ANOVA on the number of talker groups created by each listener group was not significant.

A summary of the number of talkers in each group created by participants in each listener group is shown in Table 7.4. The number of talkers in each group is fairly consistent across listener groups with averages ranging from 7.1 to 8.2 talkers. The listener groups created groups of talkers as large as 19 to 35. A one-way ANOVA on the number of talkers per group created by each listener group was not significant.

7.3.1 Perceptual Dialect Similarity

The perceptual similarity structure of the six talker regions was obtained from the free classification data using an additive clustering analysis (Hornik, 2005; Sattath & Tversky, 1977). A 6 x 6 matrix of dialect similarity was constructed based on the free classification data. The matrix represents the number of times participants placed tokens from each talker region into groups with tokens from each talker region. The value of each cell in the diagonal represents the number of times tokens from the same talker region were grouped together.

To obtain the clustering solution, the 6 x 6 similarity matrix was converted to a distance matrix (as described in the previous session) and submitted to an ADDTREE implementation in the CLUE package for R (Hornik, 2005). The clustering solution is shown in Figure 7.3. The vertical branches represent the distance between any two dialect nodes while the length of the horizontal branches is not meaningful. The clustering analysis shows that the talker regions are quite similar. There are two general perceptual clusters: a North and a South.

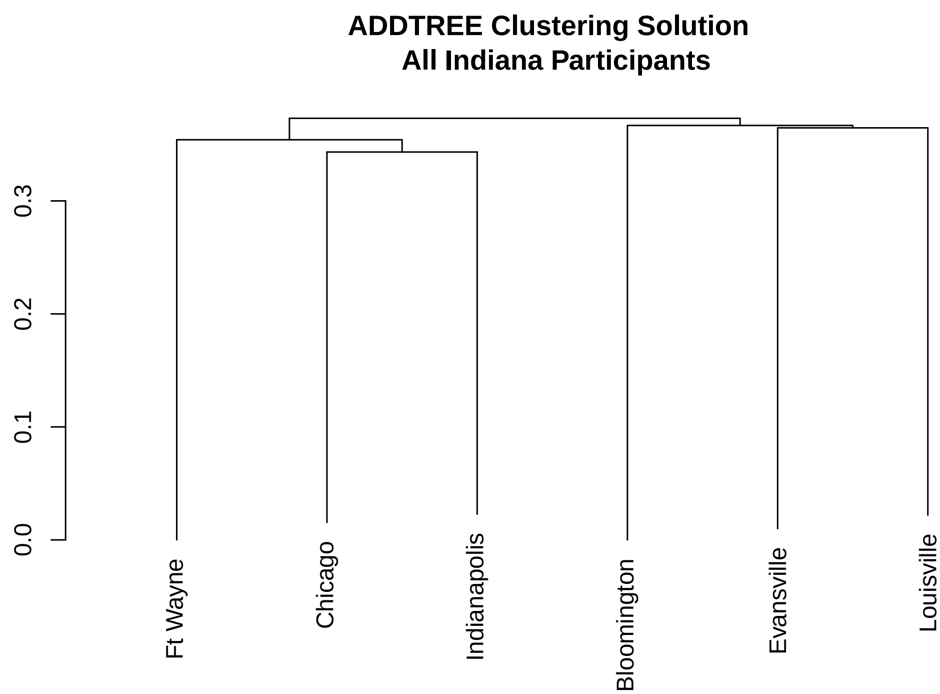

Figure 7.3: ADDTREE clustering solution for all listeners

Clustering solutions by each listener group are presented in Figure 7.4. The solutions for the all northwest listeners, the white northwest listeners, and the central listeners are basically identical to the solution for the aggregate of all listeners in that they have clusters for north and south talker groups. The white northwest listeners perceive talkers from the Chicago area to be more similar to talkers from Indianapolis than Fort Wayne, while the central listeners perceive talkers from the Chicago area and Fort Wayne to be more similar than talkers from Indianapolis. The black northwest listeners include Indianapolis talkers in the south cluster. The northeast listeners include Bloomington talkers in the north group. The south listeners' data produces a clustering solution with a binary-branching hierarchical structure in which the northern talkers form a cluster on the lower levels and the southern talkers are added at higher levels.

Taken as a whole, all of the listener groups distinguish North and South dialect clusters; however, differences arise in which talker groups are included in the clusters. The most northern talker groups, Chicago and Fort Wayne, as well as the most southern talker groups, Evansville and Louisville, are always clustered together early. The central talker groups, Indianapolis and Bloomington, show more variability in their clustering.

7.3.2 Perceptual Talker Similarity

In order to assess the perceptual similarity between talkers and their respective dialect regions, the pair-wise perceptual distances between the talkers was calculated. A 24 x 24 similarity matrix was constructed representing the number of times each of the 24 talkers was paired with the other 23 talkers. The diagonal of the matrix, representing the self-similarity of the talkers, was omitted. The similarity matrix was converted into a distance matrix by dividing the number of grouping occurrences by the number of possible opportunities for grouping subtracted from one.

The distance matrix was submitted to the multidimensional scaling function `cmdscale` in R. Multidimensional scaling seeks to recover the underlying dimensions that participants use to evaluate the similarity or distance between objects and arrange the objects on a scale of those dimensions. A common example of multidimensional scaling involves the actual

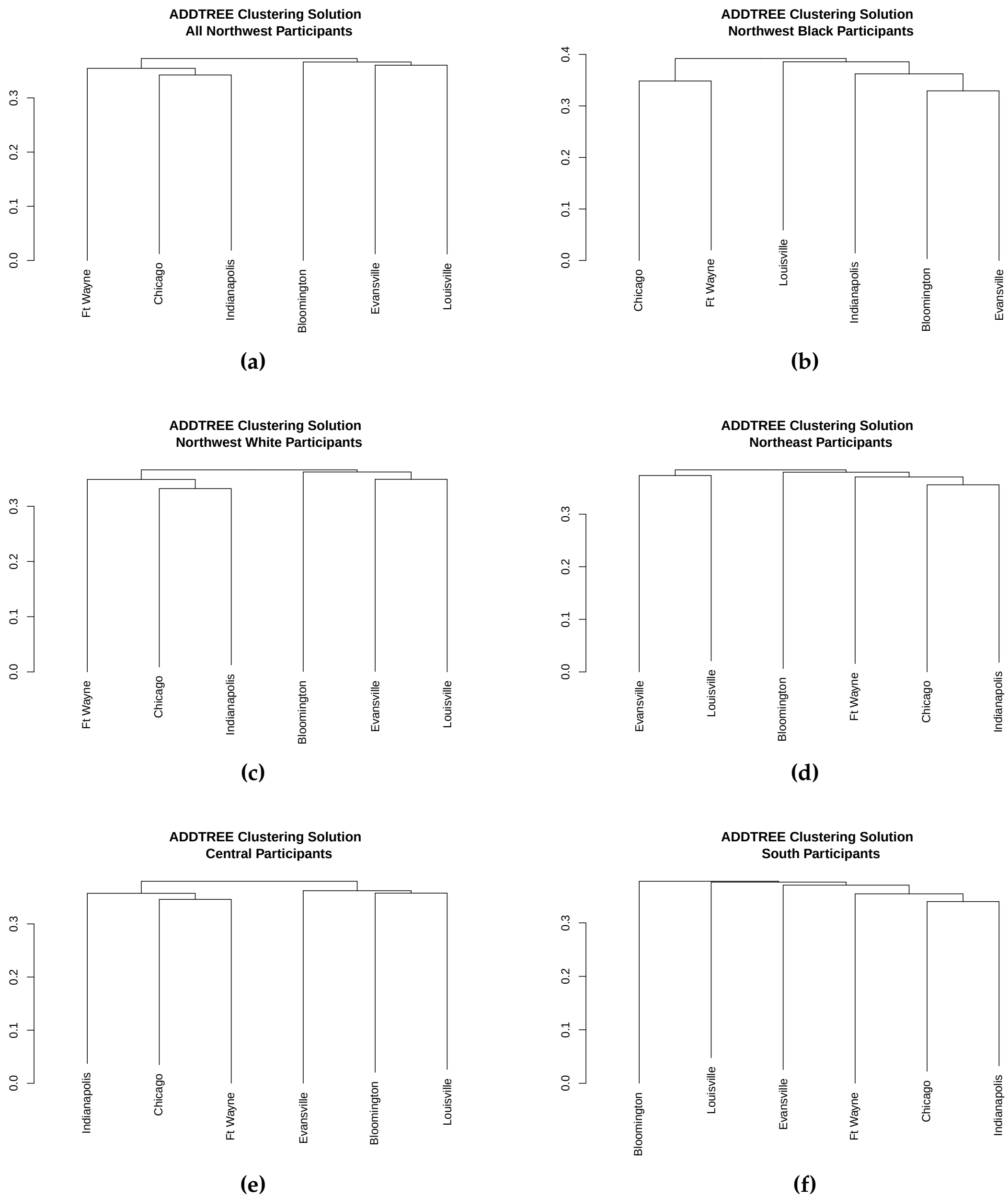

Figure 7.4: ADDTREE clustering solutions by listener groups

physical distances between cities in the United States. If we were to construct a matrix of the distances between major cities in the United States and submit only the distance matrix to multidimensional scaling, we would get a solution that gives the accurate position of the cities relative to each other. The solution would be best fit with two dimensions representing north/south and east/west scales. A third dimension might further improve the fit depending on how the distances between cities was measured (over the surface of the earth or the shortest point through the spheroid).

Multidimensional scaling on perceptual distances could contain many more meaningful dimensions. For example, physical objects may be evaluated on color, shape, texture, and size. A red ball and a pink cube would be close together on the color scale but rather distant on the shape scale. Objects that vary on these four salient dimensions would likely be evaluated in such a way that the solution would improve in fit up to four dimensions and further dimensions would be fitting only noise.

Perceptual objects such as voices do not have a set number of dimensions that they might be evaluated on, so some interpretation of the results is necessary. Clopper (2004), using a very similar experimental design to the one described here, found that listeners evaluated talkers from different dialect regions in the United States on the geographical location of the dialect, the markedness of the dialect, and the gender of the talker. It is possible that talkers were evaluated on other dimensions such as pitch, speaking rate, or timbre, but these features were less salient to listeners and may have, therefore, been evaluated inconsistently. Including additional dimensions to account for these features with lower salience may not significantly improve the fit, or these features might be overwhelmed by noise in the data.

A common way to decide how many dimensions to include involves plotting the stress or eigenvalues against the number of dimensions evaluated by the multidimensional scaling algorithm (Kruskal & Wish, 1978). Often called a *scree plot*, this plot is used to visually identify an "elbow" in the plot beyond which inclusion of additional dimensions does not improve the fit. Another criteria in determining the number of dimensions to use is the interpretability of the dimensions. As mentioned above, a dimension identified by the algorithm might just be noise, and attempting to interpret the scale would be detrimental to the analysis. On the other hand, a dimension might represent a real feature being evaluated by the listeners, but the data needed to confirm the scale is not available. For example, Clopper's gender dimension was determined because she knew the gender of the talkers in her experiment; however, if there were a pitch dimension, data on the pitch of each talker would need to be collected to validate this dimension. Even if data on a feature such as pitch were collected, the measurement taken might not correspond to the way pitch is perceived and used for evaluation by the listeners. Do they use average pitch or maximum pitch or pitch range or rate of pitch modulation or some combination? It may be impossible to recover a correlate to the scale on which a real perceptual dimension is evaluated. Therefore, only an interpretable number of dimensions are retained.

Figure 7.5 shows a scree plot of the eigenvalues (or stress) for multidimensional scaling solutions for up to ten dimensions based on data from all retained participants. The first elbow occurs at four dimensions and no further improvements occur after eight dimensions.

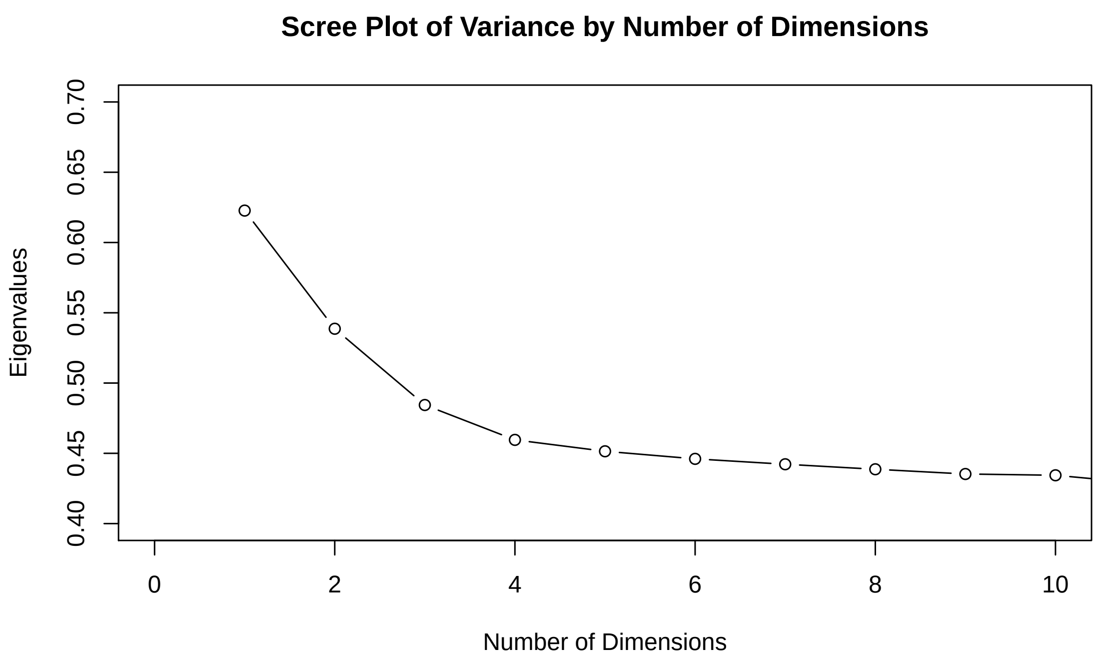

Figure 7.5: Scree plot based on multidimensional scaling solutions for all listeners

Because the protocol used in the present experiment is very similar to that used by Clopper (2004), I may expect to also find geography and markedness dimensions. In debriefing interviews following the free classification task, participants mentioned that talkers sounded northern, southern, Chicago, country, normal, "like me", and "I can't place them", which I would expect to translate to geography and markedness dimensions. The present experiment included only female talkers, so I would not expect a gender dimension. Because sentences were repeated by talkers in each talker group, a sentence dimension might be present. Other descriptions mentioned in the debriefing interviews include pitch (high and low), voice quality (harsh, soft, nasal), and rate (fast and slow). In the following section, I will present plots of a four-dimension solution.

Figure 7.6 shows a plot of the first and second dimensions of the MDS solution with the text of each point giving the talker identification code (the letter represents the talker's region) and the color representing the talker dialect region. The general structure of the data is dispersed with a greater concentration of points on the far right and vacant space in the central part of the plot. Considering the dialect regions encoded by the color of each point, the points are also distributed without any particularly evident clusters. All of the talkers from the Chicago region are on the right side while all of the talkers from Fort Wayne are on the top of the plot. There are no contrasting patterns to suggest that either of these orientations represent meaningful scales.

Perhaps the listeners are not sensitive to the particular variation in these regions or the talkers are not uniformly representative of the dialect regions they are purported to represent. We can consider how these same talkers were categorized by these same listeners according to the labels provided during the 4-alternative forced choice (4AFC) task. Figure 7.7 shows the same solution on the first two dimensions as Figure 7.6 but now the colors

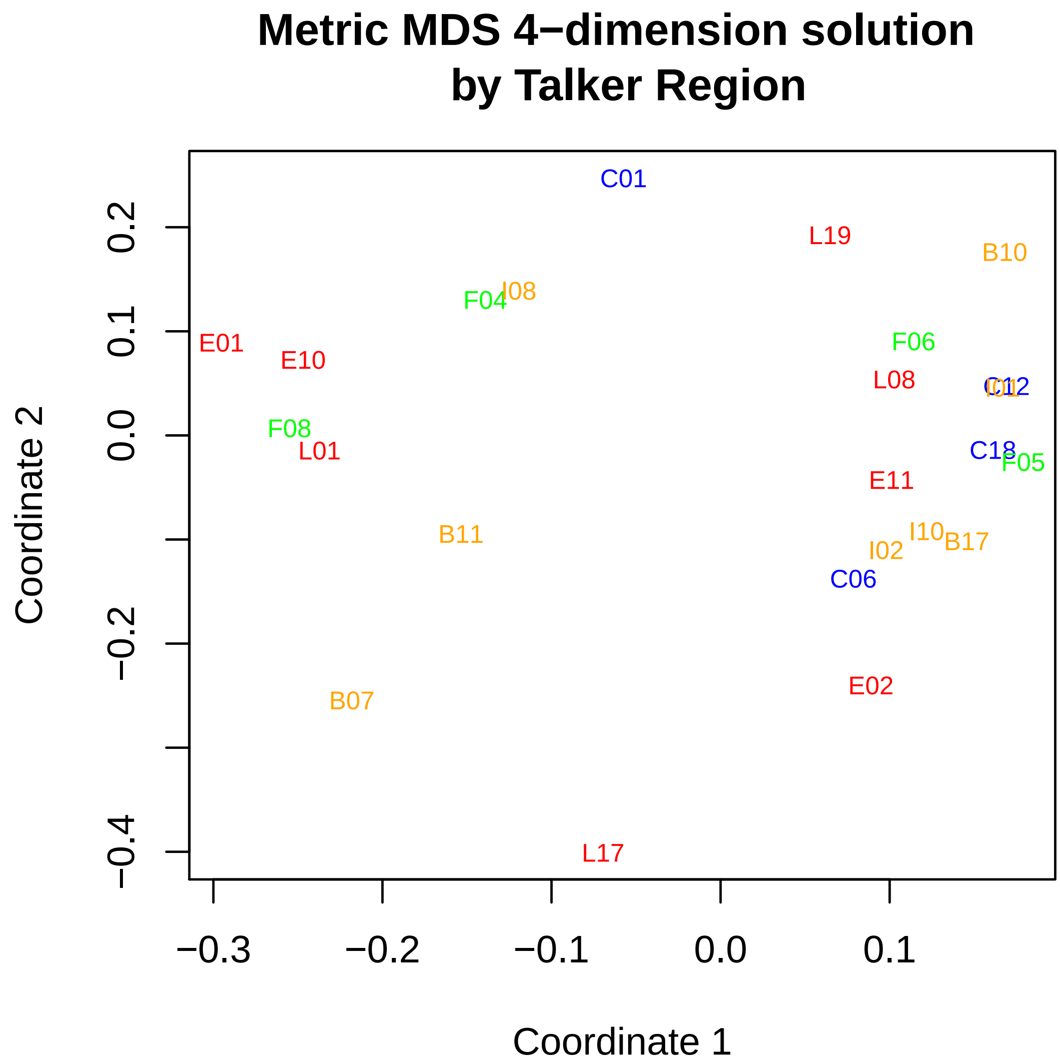

Figure 7.6: Multidimensional scaling solution plot of first and second dimensions for all listeners with points colored by the talker's dialect region (blue = Inland North, green = North, orange = Midland, red = South

correspond to the region each talker was categorized as most frequently during the 4AFC task.

Now we can see a clear pattern in the free classification data. Although the plot could be rotated about 45°, the first dimension corresponds to the talkers categorized as Southern on the left and Midlands on the right. The two talkers on the far right who were categorized as Northern suggest this dimension is a North/South scale rather than Midlands/South.

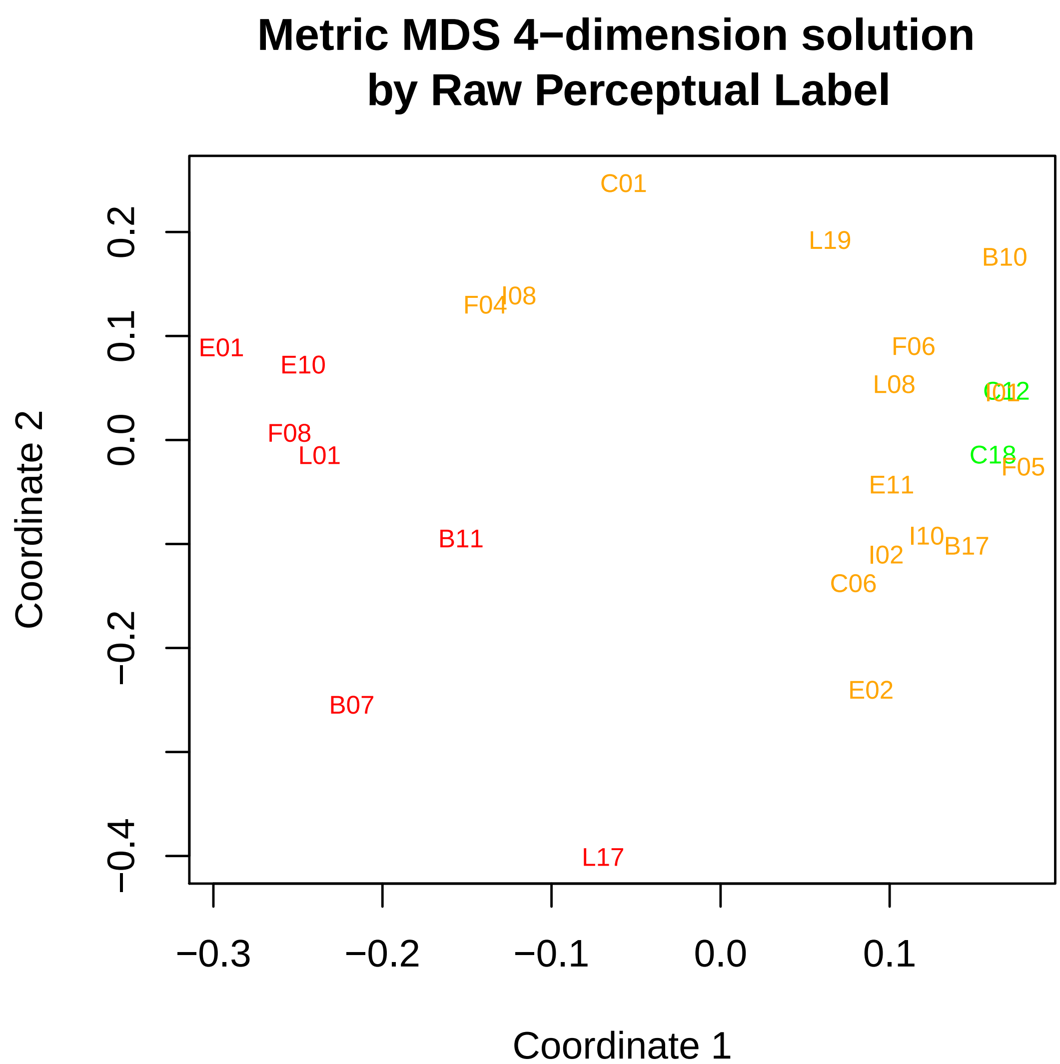

Figure 7.7: Multidimensional scaling solution plot of first and second dimensions for all listeners with points colored according to the region most associated with each talker in the 4-alternative forced choice task (blue = Inland North, green = North, orange = Midland, red = South

Since the 4AFC data was biased toward Midlands and South, using the maximum value of the regional classification frequencies is perhaps too coarse-grained—it only gives two North talkers and no Inland North talkers. One way to control for the response biases and pick out which region each talkers are associated with is to calculate the z-scores for each talker by each dialect category and then using those values to select the normalized most frequent region.

Figure 7.9 shows the same MDS solution as the previous two figures with the colors reflecting the region with the highest z-score for each talker. In this figure, all seven of the talkers that were classified as South in the previous figure have retained the same classification. Both of the talkers that were classified as North in the previous figure are, likewise, still classified as North, but they are joined by an additional three talkers that were classified as Midlands based on the non-normalized values. A total of seven talkers are now classified as Inland North. Five talkers have remained classified as Midlands. The groupings roughly correspond to the regional provenance of the talkers, although there are some notable confusions. For example, a talker from Fort Wayne is categorized among the South group and two talkers from the Louisville area are categorized as Inland North and North.

With the dialect classifications based on the normalized categorization values, Figure 7.8 still shows that dimension 1 distinguishes the talkers perceived as southern from the other talkers; however, the other end of the scale seems to correspond to "nonsouthern" rather than to "north". The interpretation of the second dimension is less clear. The groups categorized as North and Midlands form a relatively tight cluster compared to the group labeled as Inland North. Drawing on a dimension found by Clopper (2004), we could posit this as the "markedness" dimension. In this case, the marked end of the scale is in the upper-left corner and the unmarked is the lower-right corner. The North and Midlands groups are more so toward the unmarked end and the Inland North talkers are distributed across the scale. Applying this scale to the Southern group, the talkers from the prototypical southern areas, Louisville and Evansville, are toward the marked end while the Bloomington talkers are toward the unmarked end.

Figure 7.9 presents the solution with the first and third dimensions. The argument for a markedness dimension is easier to make with reference to this third dimension. Again, a

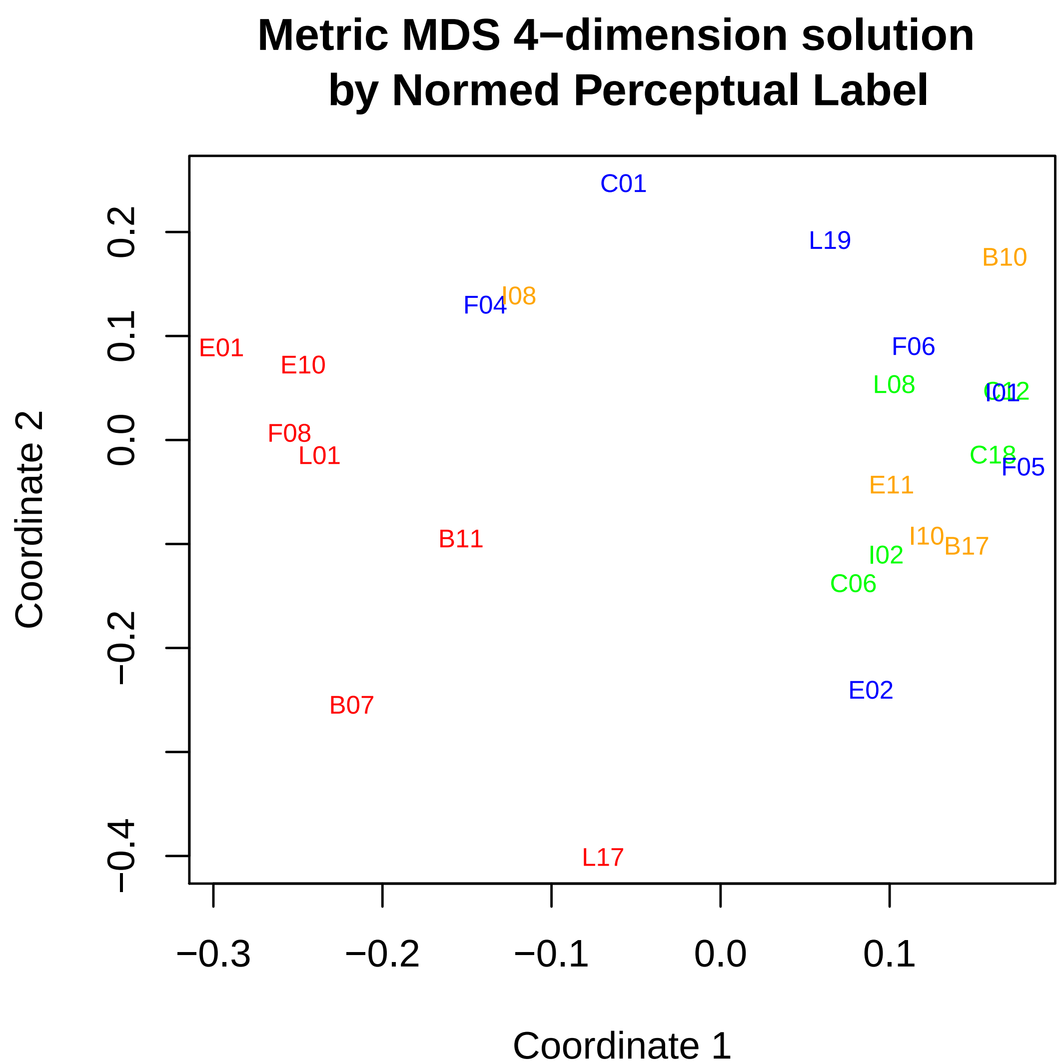

Figure 7.8: Multidimensional scaling solution plot of first and second dimensions for all listeners with points colored according to the region with the highest normalized value for each talker in the 4-alternative forced choice task (blue = Inland North, green = North, orange = Midland, red = South

rotation of about 45° is needed. In the orientation shown in the figure, the talkers below the diagonal from the upper-left to the lower-right corner are marked. The perpendicular diagonal separates the North from the South. The triangle formed above the intersection of these two axes is empty as there are not talkers that are considered both unmarked and southern.

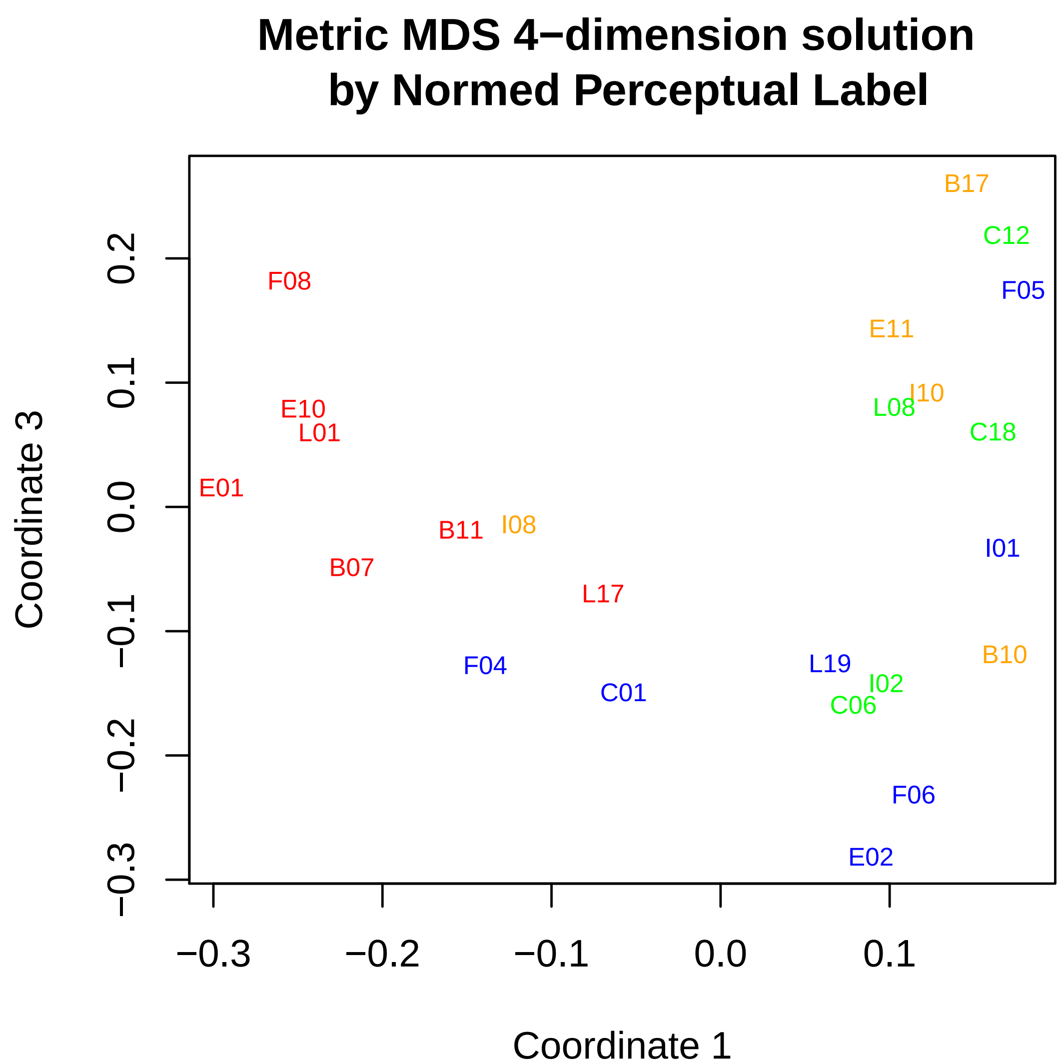

Figure 7.9: Multidimensional scaling solution plot of first and third dimensions for all listeners

Figure 7.10 shows the MDS solution plotting the first and fourth dimensions. On the fourth dimension, all of the talkers from each group are evenly distributed along the scale. Thus, whatever this dimension measures, it does not have to do with regional dialect. Perhaps this dimension measures one of features described by participants in the debriefing interviews mentioned above.

Further measurements not yet available would be needed to see if this dimension represents a feature related to some aspect of voice quality, speech rate, or pitch. These

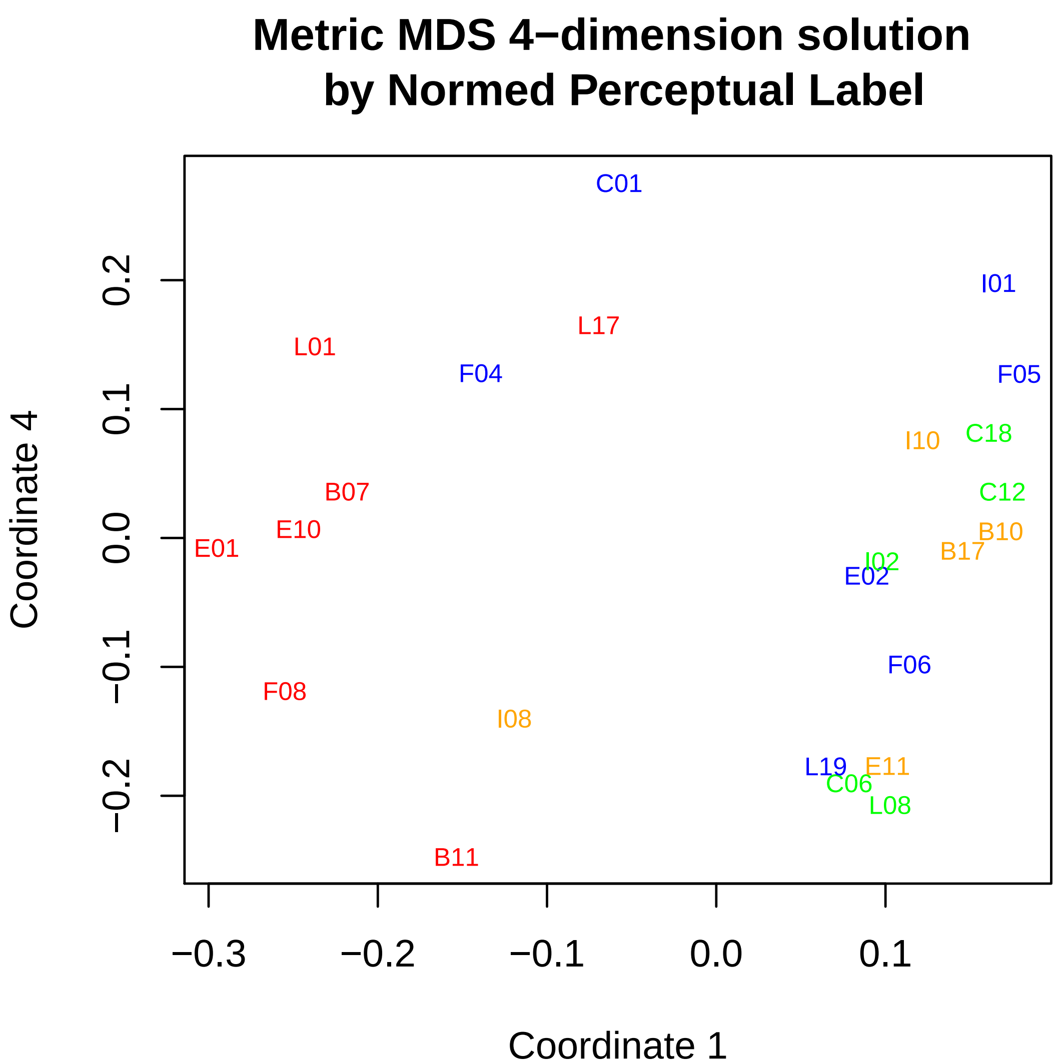

Figure 7.10: Multidimensional scaling solution plot of first and fourth dimensions for all listeners with points colored according to the region with the highest normalized value for each talker in the 4-alternative forced choice task (blue = Inland North, green = North, orange = Midland, red = South

measurements could be either acoustic or subjective based on participants' evaluation of these same talkers.

The data are evenly distributed across the second and fourth dimensions when grouped by geographic perceptual categorization. Two constants that may have guided listeners' groupings of tokens are the talker and the sentence. Each talker spoke two tokens and each sentence was repeated by one talker from each of the six regions. In order to assess the relationship of the data to these dimensions a new MDS solution is calculated based on the groupings of individual tokens rather the sum of the token groupings by talker. The solution was calculated for four dimensions as before.

Figure 7.11 shows a plot of the MDS solution by token for the second and fourth dimensions. The pairs of tokens spoken by each talker are connected by a line. If either of these dimensions describe some unique aspect of each talker, we would expect to see the talkers' tokens clustering along one of these dimensions. A perfectly vertical or horizontal line between pairs would indicate a close clustering on the tokens on second or fourth dimension, respectively.

The points plotted by token pairs in Figure 7.11 are generally more dispersed across the fourth dimension and less dispersed across the second dimension. While the data are fairly noisy, we can consider the second dimension to represent an evaluation related to the individual talker.

Figure 7.12 shows a plot of the same MDS solution as in 7.11, but now the lines connect the unique sentences. Here we see a fairly narrow dispersion of sentence sets along the fourth dimension. We can consider the fourth dimension to represent an evaluation of the sentence.

The spaces presented in Figures 7.6 through 7.12 are the result of combining all of the responses from participants in all groups. We can assess variation in perceptual similarity between groups based on region of residence. Similarity matrices were calculated for each of the five groups. An Individual Differences Scaling (INDSCAL) analysis was conducted on these five matrices (Carroll & Chang, 1970). INDSCAL assumes that all individuals—or in this case groups—attend to the same dimensions and calculates the degree to which individuals attend to each dimension. For example, if two groups, architects and nonarchitects, were to assess the similarity of buildings, the architects would likely attend to architectural

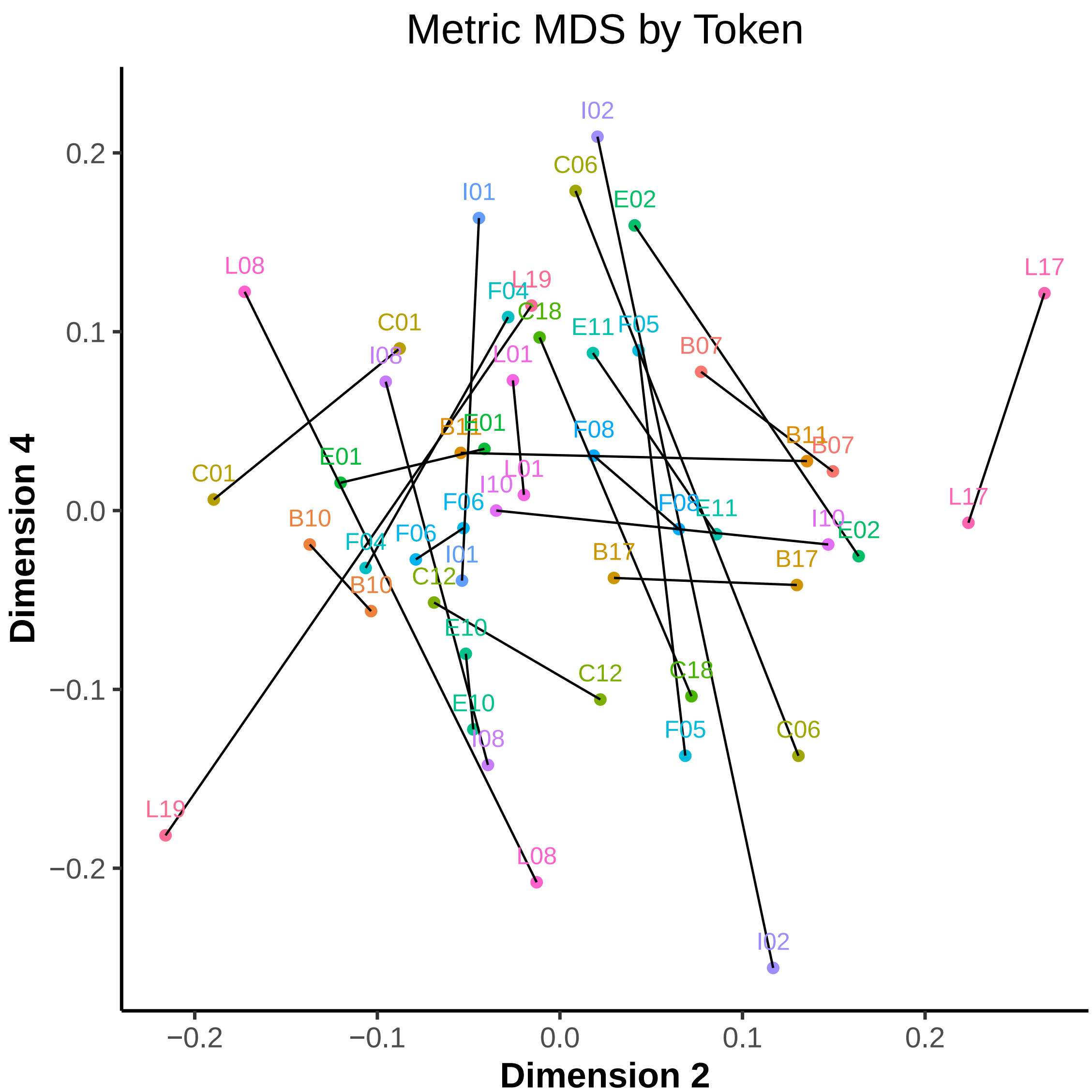

Figure 7.11: Multidimensional scaling solution plot of second and fourth dimensions for each token with lines connecting pairs of tokens spoken by the same talker

style more than the nonarchitects even though both might attend to building material to an equivalent degree.

Table 7.5 gives the normalized group weights for the four dimensions described above. The weights are normalized for each group of listeners so that the sum of the weights on the four dimensions is one. Most groups weighed Geography (dimension 1) most heavily, followed by Talker (dimension 2) or Markedness (dimension 3), and finally Sentence

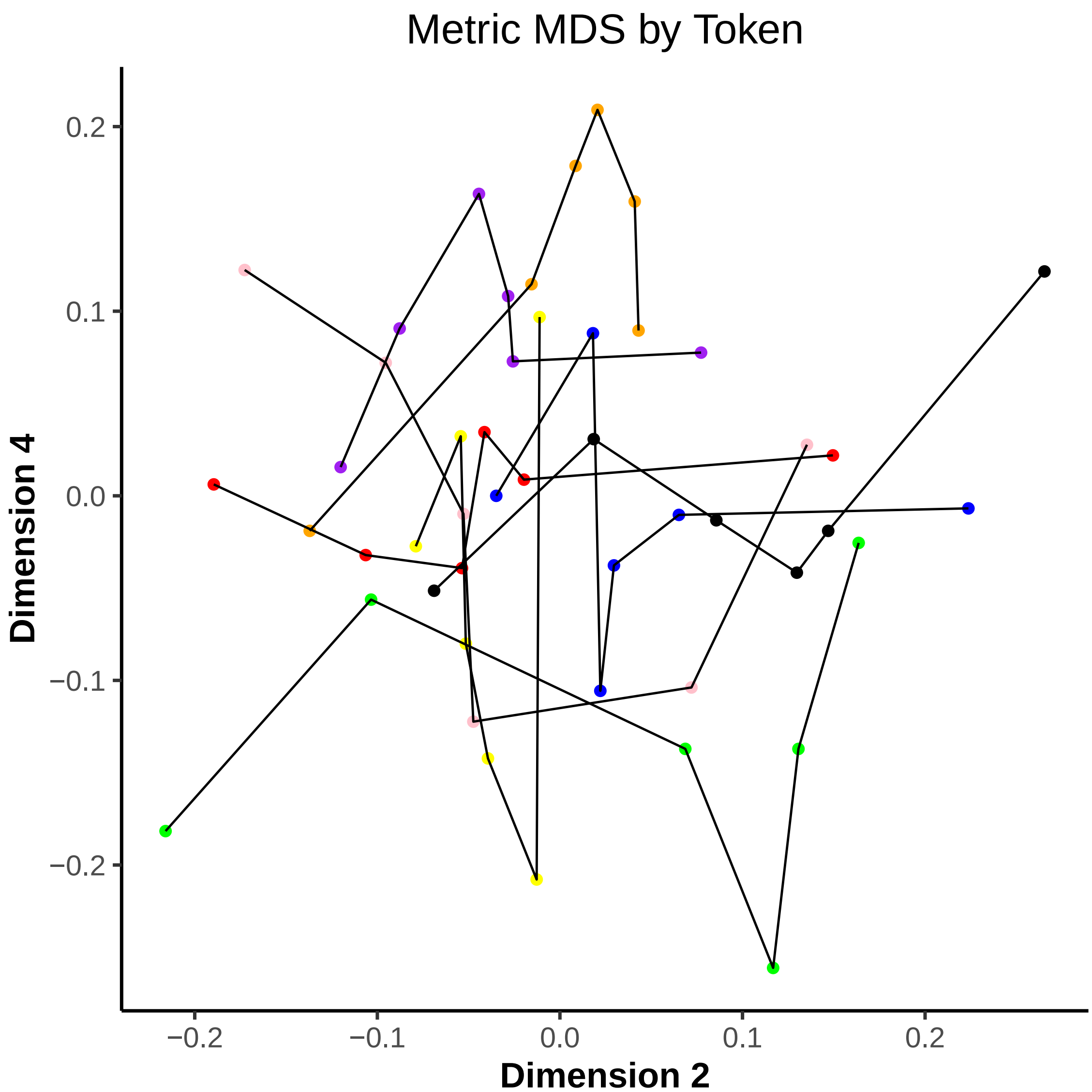

Figure 7.12: Multidimensional scaling solution plot of second and fourth dimensions for each token with lines connecting set of tokens of the same sentence

(dimension 4). The Central and white Northwest listeners weighted Geography the most while the South and black Northwest listeners weighted Markedness the most. There is a roughly inverse relationship between the weighting on the Geography and Markedness dimensions. Overall, though, the group differences are quite close to the expected average for four dimension of 0.25, suggesting that the groups used similar criteria to determine the similarity of the talkers.

Table 7.5: Group weights on each dimension from INDSCAL with dimension number in parentheses

	Geography (1)	Talker (2)	Markedness (3)	Sentence (4)
Northwest(B)	0.24	0.23	0.27	0.25
Northwest(W)	0.30	0.26	0.21	0.23
Northeast	0.28	0.25	0.25	0.22
Central	0.30	0.23	0.25	0.22
South	0.26	0.26	0.26	0.22

The MDS solutions for each listener group showing the dimensions related to Geography and Markedness are shown in Figure 7.13. The coordinates encoding these dimensions are not the same for each group because the MDS algorithm sorts the dimensions according to each dimensions' impact on the degree of fit. The color of the points are based on the same normalized maximum classification values on the four-alternative forced-choice task for all participants as depicted in Figures 7.6 through 7.10. This color encoding is retained in order to facilitate comparison of between group variation. Plots with colors encoding 4AFC classifications by each listener group can be found in Appendix VI. An important note regarding the nature of multidimensional scaling is that solutions only encode relative distances between points. Therefore, the relationship between the points in the graph space is invariant in regards to rotation, mirroring, translation, and scaling. An effort has been made to retain the general configuration of the plots in Figure 7.13 by flipping the axes as appropriate. That is, the South end of the X dimension is on the left while the Marked end of the Y dimension is on the bottom.

The general patterns observed in the solution for all participants are present in the solutions by listener group. The most prominent pattern is the consistency of the South/Non-South dimension. This pattern is shown by the clustering of talkers perceived as Southern shown in red on the left of the plots and the other talkers dispersed on the right side of the plots. The Markedness dimension shows less consistency. The talkers perceived as being from the Chicago region shown in blue do tend to be clustered with themselves while the talkers perceived as Northern and Midland in green and orange are dispersed across similar spaces as each other within the Non-South area.

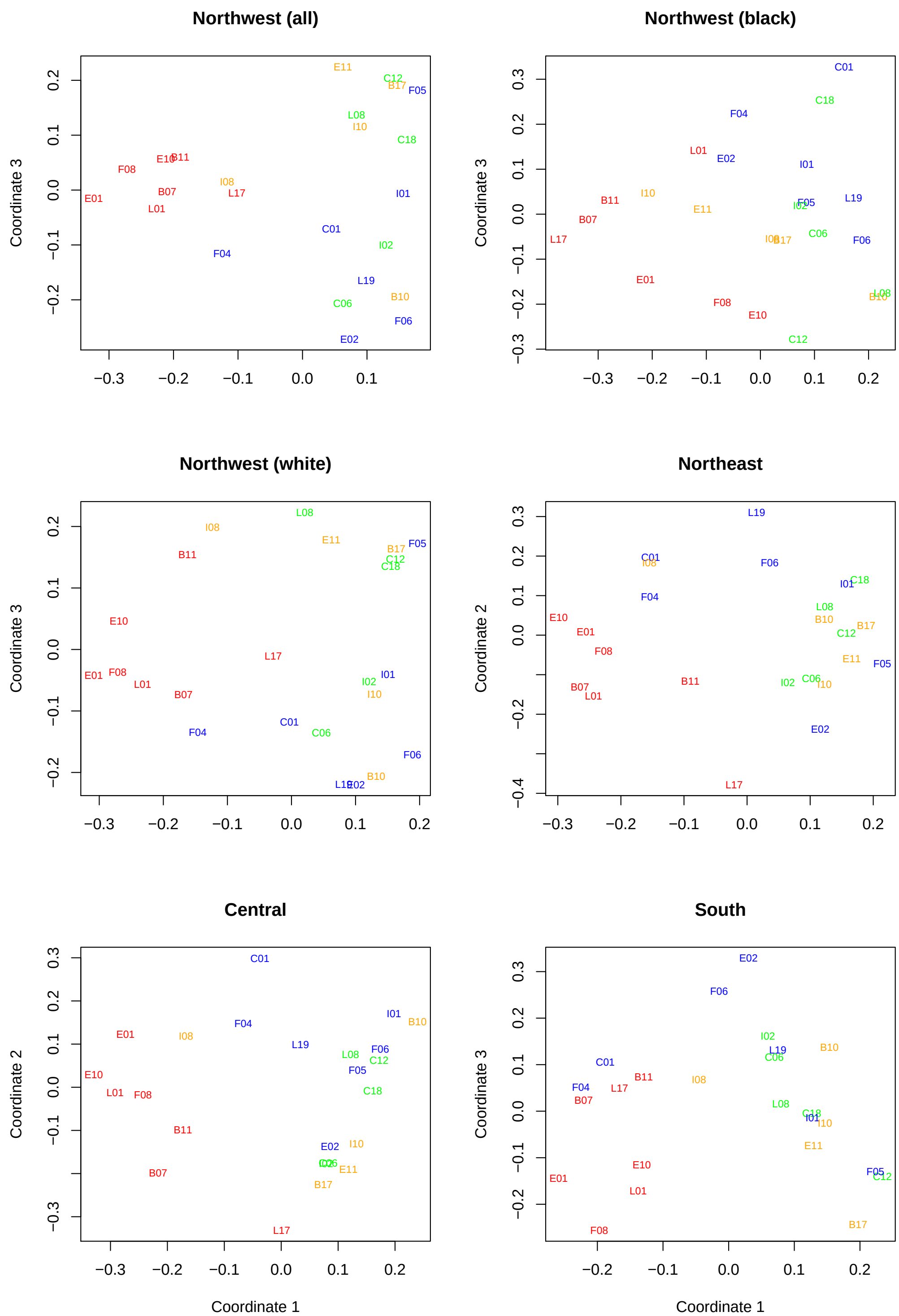

Figure 7.13: Multidimensional scaling solution plots of dimensions corresponding to Geography and Markedness for listener groups with points colored according to the region with the highest normalized value for each talker in the 4-alternative forced choice task for all listeners (blue = Inland North, green = North, orange = Midland, red = South)

7.4 Discussion

The results of this experiment show that listeners in Indiana can categorize other talkers from Indiana into generally consistent groups. The groups are not as clearly defined as in an earlier study of American dialects that focus on the entire United States (Clopper & Pisoni, 2007) suggesting an effect of either less extreme differences between the dialects included in the study or a reduced sensitivity to local dialect diversity.

Two major dialect categories emerged across listeners in all four research sites: a North and a South. Talkers from the most northern and southern regions were always included in their respective groups, but there was some variation in which of the two groups included the talkers from the Central region of the state (Indianapolis and Bloomington).

Multidimensional scaling analysis revealed that listeners sorted talkers along a South–Nonsouth dimension as well as a Marked–Unmarked dimension. Participants were also influenced by the talker identity and particular sentence in their categorizations. Listeners may have been influenced by voice quality features as well, but further work would be needed to determine which features were used and to what extent.

Individual differences between listener groups were slight, though there was a tendency for groups to inversely weight attention to Geography and Markedness. There was no immediately evident interpretation of the minor variations in attentional weighting, so the individual differences can be considered negligible in this experiment.

It is notable that the analysis of these dimensions does not work well based on the actual provenance of the talkers, but it does work well based on the perceived identities of the talkers according to the listeners.

One issue to consider further based on this analysis is the linguistic situation in Fort Wayne. Across most listener groups, the perceived North dialect includes talkers from near Chicago and Indianapolis before talkers from Fort Wayne are included. Thus, Fort Wayne is not a core member of the North group. Talker F08, from Fort Wayne, was perceived as Southern on the 4AFC task and appears firmly among the Southern talkers in the MDS solution. Talker F04 appears close to the Southern group, too, although she is identified as Inland. These stimuli were recorded about 15 years before this experiment took place, so these observations could be due to changes in progress in Fort Wayne, or they could reflect some idiosyncrasies in the dialect situation in Fort Wayne. Ongoing work on Canadian Raising in Fort Wayne (Berkson, Davis, & Strickler, 2017; Davis, Berkson, & Strickler, To Appear) could shed light on some aspects of the surprising perceptual categorization of talkers from this area.

7.5 Appendix V: Simulus Materials for Free-Classification Task

Talker	SPIN Number	Sentence Type	Sentence Text
B07	101	nonStereotype	A round hole won't take a square peg.
B07	164	aeay	The flashlight casts a bright beam.
B10	137	aeay	They tracked the lion to his den.
B10	179	nonStereotype	Banks keep their money in a vault.
B11	129	aeOnly	Instead of a fence, plant a hedge.
B11	189	nonStereotype	Break the dry bread into crumbs.
B17	124	ayOnly	Her entry should win first prize.
B17	173	nonStereotype	Cut the meat into small chunks.
C01	101	nonStereotype	A round hole won't take a square peg.
C01	164	aeay	The flashlight casts a bright beam.
C06	137	aeay	They tracked the lion to his den.
C06	179	nonStereotype	Banks keep their money in a vault.
C12	124	ayOnly	Her entry should win first prize.
C12	173	nonStereotype	Cut the meat into small chunks.
C18	129	aeOnly	Instead of a fence, plant a hedge.
C18	189	nonStereotype	Break the dry bread into crumbs.
E01	101	nonStereotype	A round hole won't take a square peg.
E01	164	aeay	The flashlight casts a bright beam.
E02	137	aeay	They tracked the lion to his den.
E02	179	nonStereotype	Banks keep their money in a vault.
E10	129	aeOnly	Instead of a fence, plant a hedge.
E10	189	nonStereotype	Break the dry bread into crumbs.
E11	124	ayOnly	Her entry should win first prize.
E11	173	nonStereotype	Cut the meat into small chunks.
F04	101	nonStereotype	A round hole won't take a square peg.
F04	164	aeay	The flashlight casts a bright beam.
F05	137	aeay	They tracked the lion to his den.
F05	179	nonStereotype	Banks keep their money in a vault.
F06	129	aeOnly	Instead of a fence, plant a hedge.

F06	189	nonStereotype	Break the dry bread into crumbs.
F08	124	ayOnly	Her entry should win first prize.
F08	173	nonStereotype	Cut the meat into small chunks.
I01	101	nonStereotype	A round hole won't take a square peg.
I01	164	aeay	The flashlight casts a bright beam.
I02	137	aeay	They tracked the lion to his den.
I02	179	nonStereotype	Banks keep their money in a vault.
I08	129	aeOnly	Instead of a fence, plant a hedge.
I08	189	nonStereotype	Break the dry bread into crumbs.
I10	124	ayOnly	Her entry should win first prize.
I10	173	nonStereotype	Cut the meat into small chunks.
L01	101	nonStereotype	A round hole won't take a square peg.
L01	164	aeay	The flashlight casts a bright beam.
L08	129	aeOnly	Instead of a fence, plant a hedge.
L08	189	nonStereotype	Break the dry bread into crumbs.
L17	124	ayOnly	Her entry should win first prize.
L17	173	nonStereotype	Cut the meat into small chunks.
L19	137	aeay	They tracked the lion to his den.
L19	179	nonStereotype	Banks keep their money in a vault.

7.6 Appendix VI: Multidimensional Solutions by Listener Group Coded by 4AFC Classifications by Listener Group

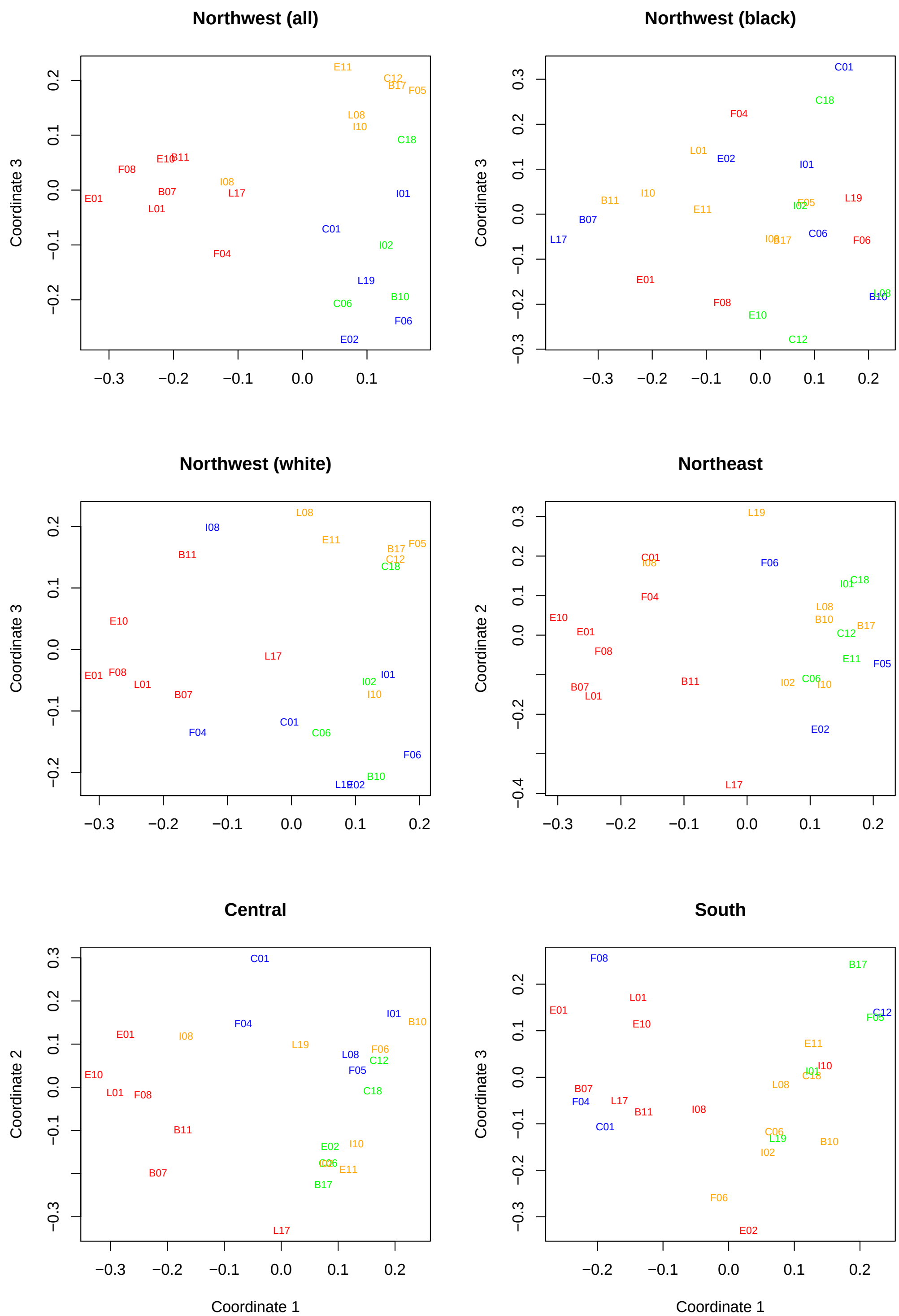

Figure 7.14: Multidimensional scaling solution plots of dimensions corresponding to Geography and Markedness for listener groups with points colored according to the region with the highest normalized value for each talker in the 4-alternative forced choice task by listener group (blue = Inland North, green = North, orange = Midland, red = South)