Behavior Research Methods
2008, 40 (2), 575-581
doi: 10.3758/BRM.40.2.575

Auditory free classification: Methods and analysis

Cynthia G. Clopper
Ohio State University, Columbus, Ohio

The free classification paradigm and its potential application to perceptual acoustic and auditory research are summarized. Specific methods for the application of the auditory free classification paradigm in speech research are then described, including techniques for stimulus preparation, data collection, and statistical analysis, based on a sample auditory free classification study of regional dialect variation. The results suggest that auditory free classification is a promising paradigm for examining the perceptual classification and similarity of speech and nonspeech auditory stimulus materials.

The free classification paradigm (Imai, 1966; Imai & Garner, 1965) has been used in cognitive psychology for decades to explore naive participants' classification strategies in the absence of explicit experimenter-provided category labels. In a typical free classification task, participants are asked to sort a set of visual stimulus items (e.g., ambiguous figures or sets of objects) into two or more groups. The participants' classifications are submitted to clustering or scaling analyses, and the results are interpreted as a reflection of the most salient perceptual dimension(s) of similarity across the stimulus items (Medin, Wattenmaker, & Hampson, 1987).

In the auditory domain, free classification paradigms could be used to examine the perceptual similarity of a range of speech and nonspeech stimulus materials, including linguistic structures such as segments, tones, or intonation contours, and indexical properties such as voices, speaking styles, emotions, dialects, foreign accents, or languages. The critical property of the free classification paradigm that distinguishes it from other experimental techniques is that category labels and dimensions of contrast do not need to be specified in advance by the experimenter. In the domain of speech research, free classification allows the experimenter to explore the complex interaction of perceptual cues to linguistic and indexical categories without requiring any a priori judgments about what the relevant acoustic–phonetic cues or their weightings might be. However, the translation of the free classification paradigm into the auditory domain is not trivial. Marcell, Borella, Greene, Kerr, and Rogers (2000) asked participants to provide verbal category labels for a set of environmental sounds. The task successfully avoided experimenter-imposed categories, but did not ensure that participants produced groups of similar items, because each stimulus could have been assigned a unique category label. That is, the stimulus materials could have been sorted exclusively into groups with one member each. Smith and Baron (1981) and Gattuso, Smith, and Treiman (1991) examined free classification performance for visual and auditory stimulus materials. Both studies used a variant of the oddball-detection task in the auditory domain. Participants were asked to select the two most similar syllables out of syllable triads. The oddball-detection task also avoided experimenter-imposed labels, but it required a series of experimenter-imposed comparisons rather than participant-driven classification.

Several recent studies have developed techniques for presenting an auditory free classification task to participants using a visual display. Granqvist (2003) described a procedure for obtaining perceptual distance ratings of synthetic speech samples using a visual sort task. Esposito (2006) adopted Granqvist's methods to examine perceptual classification of voice qualities. Although these two studies successfully adapted the free classification task for use with auditory stimulus materials, both were fairly restricted versions of the free classification paradigm. Granqvist limited his interpretation of the results to a single dimension of similarity, and Esposito limited her participants to a binary classification by requiring them to produce two groups. The free classification paradigm is more flexible, however, and allows researchers to explore general classification strategies in high-dimensional spaces by specifying any particular number of groups, or by allowing the participants to choose the number of groups that they produce.

McAdams, Vieillard, Houix, and Reynolds (2004) and Clopper and Pisoni (2007) took advantage of one aspect of the flexibility of the free classification paradigm to examine the perceptual classification of contemporary musical themes and regional dialects of American English, respectively. In both studies, participants were permitted to make as many groups of items as they wanted, with as many

C. G. Clopper, clopper.1@osu.edu

items in each group as they wished. The paradigms were implemented similarly, using a visual display to represent the auditory stimulus materials. In both cases, stimulus presentation was controlled by specially designed software.

In addition to avoiding experimenter-defined category labels, the auditory free classification paradigm also provides an efficient alternative to time-intensive paired comparison perceptual similarity judgment or oddball-detection tasks. For example, whereas Clopper and Pisoni's (2007) free classification task with 48 talkers took 10–15 min for participants to complete, a paired comparison dialect similarity rating task with only 32 talkers took 50–60 min (Clopper, Levi, & Pisoni, 2006).

The purpose of the remainder of this article is to describe in detail a set of methods for the application of the auditory free classification technique to speech research, including methods for statistical analysis of the results. Unlike the techniques used in the literature reviewed above, the methods and techniques described below do not require any specialized software or programming skills beyond what is typically available in speech research laboratories, and can therefore be applied to new research questions by both students and more senior researchers. To make the methods more concrete, they are described in the context of an example set of data that was obtained in an unpublished follow-up study to the experiments reported by Clopper and Pisoni (2007). The stimulus materials in the present example were identical to those used in Clopper and Pisoni's Experiment 2. However, the methods differed slightly between their study and the present study: Whereas their listeners were allowed to make as many groups as they wanted, the participants in the present experiment were required to sort the stimulus materials into a fixed number of groups.

METHOD

Talkers

The set of 48 talkers included four men and four women from each of six dialect regions in the United States (New England, Mid-Atlantic, North, Midland, South, and West). The talkers were white native speakers of American English and ranged in age from 18 to 25 years at the time of recording. Each of the talkers had lived exclusively in his or her dialect region until the age of 18, and both parents of each talker were also from the same dialect region.

Regional dialects of American English are defined primarily in terms of phonological differences, particularly with respect to the vowel system (Labov, Ash, & Boberg, 2006). The New England, Midland, and Western dialects are characterized by a merger of the vowels in *cot* and *caught*. New England is also characterized by variable raising and fronting of the vowel in *bat*; the Midland dialect is characterized by fronting of the vowels in *boot* and *boat*; and the Western dialect is characterized by fronting of the vowel in *boot*. Together, the New England, Midland, and Western dialects make up the General American dialect (Labov, 1998). The Mid-Atlantic, Northern, and Southern dialects are characterized by more distinct vowel systems, and are, therefore, the "marked" dialects in American English (Clopper, Pisoni, & de Jong, 2005). The Mid-Atlantic dialect is characterized by raising and fronting of the vowel in *bat* and raising of the vowel in *caught* (Thomas, 2001). The Northern dialect is characterized by raising and fronting of the vowel in *bat*, fronting and lowering of the vowel in *cot*, backing and lowering of the vowel in *bet*, backing of the vowel in *but*, and lowering of the vowel in *caught* (Labov, 1998). The Southern dialect is characterized by fronting and raising of the vowels in *bet* and *bit*, lowering and backing of the vowels in *beet* and *bait*, and fronting of the vowels in *boot* and *boat* (Labov, 1998). Whereas descriptions of dialect variation in the United States tend to focus on the vowel system, consonantal and prosodic phenomena may also vary across regional dialects of American English.

Stimulus Materials

The auditory stimulus materials were meaningful English sentences, with one different utterance per talker. Digital movies were created to link the auditory stimulus materials to visual representations. The auditory track of each movie was the digital auditory signal that was being classified. The visual track of each movie was a single frame digital image. In order to identify the stimulus items visually, each movie had a unique image. In the regional dialect experiment, an image of the talker's initials was used to identify the stimulus items. Given that a talker's initials are unique, but unrelated to his or her regional background, the stimulus materials were meaningfully labeled for the experimenter, but not for the participants.

Digital movies can be created using any number of multimedia software packages, such as FinalCut Pro or iMovie. For each stimulus movie, the digital audio file was imported for the audio track and the digital image was imported for the video track. The duration of the video track was matched to the duration of the audio track for each movie, so that the same visual image was presented before, during, and after the movie played. The movies were exported into individual .avi files for playback in PowerPoint running on Windows. Although other methods for linking each auditory stimulus to a unique visual label are certainly possible, the creation of the digital movies proved to be the most effective method for obtaining the desired stimulus characteristics for playback in PowerPoint.

Listeners

Thirty-eight Indiana University undergraduates (20 women, 18 men) participated as listeners. All were monolingual native speakers of American English with no reported history of hearing or speech disorders. The participants ranged in age from 18 to 25 years. Twenty of the listeners had lived exclusively in the Midland dialect region until the age of 18. The remaining 18 listeners had lived in more than one dialect region before the age of 18. The listeners received partial course credit in an introductory psychology course for their participation.

Procedure

The stimulus materials were presented to participants using a single PowerPoint slide. The stimulus items were arranged in neat columns on the left-hand side of the slide, and a grid was drawn on the right-hand side of the slide. The size of the cells in the grid matched the size of the movies, such that one stimulus item fit in each cell of the grid. The movies played when they were double-clicked and could be dragged around the screen with the mouse. The slide was presented to participants using the Slide View option, so that participants could manipulate the location of the movies on the screen.

Participants were told that each square represented a different talker. They were asked to listen to all of the talkers and then group them on the grid according to where they thought the talkers were from. They were asked to put all of the talkers from the same region of the country in a group together; they could listen to and rearrange the talkers as many times as they wanted. In this experiment, they were also instructed to make exactly six groups, but they did not have to put the same number of talkers in each group. Crucially, for the free classification paradigm, the participants were not provided with any regional dialect category labels or specific instructions about the relevant acoustic–phonetic cues to dialect identity. Figures 1A and 1B show a sample PowerPoint slide with the 48 stimulus items and a 16 × 16 grid before and after the task was completed.

Design

The first step in data analysis was to transfer the participants' responses to a spreadsheet in Excel by copying the unique stimulus identifiers (e.g., the talkers' initials) as text into a grid of the same dimensions in the spreadsheet. The resulting grid could then be manipulated and processed; blank cells in the PowerPoint slide corresponded to empty cells in the Excel spreadsheet, and filled cells in the PowerPoint slide corresponded to cells with text identifiers in the Excel spreadsheet.

Several types of summary data were extracted from the Excel grids with a macro. Using a series of embedded loops, stimulus items in the same spatial group were assigned the same numerical index. These indices were then used to output the items included in each group, which were then counted to reveal the number of groups and the number of items per group. The indices were also used to produce an item-by-item similarity matrix, in which pairs of items with the same index were assigned a value of 1 and pairs of items with different indices were assigned a value of 0. For the example shown in Figure 1, the participant produced six groups of talkers, as instructed. Group 1 (at the top of the grid) includes the following 7 talkers: VO, PF, TE, IR, HR, MM, and CP. In the talker-by-talker similarity matrix, the value of the cell for VO and PF is 1, because they were grouped together, whereas the value of the cell for VO and HH is 0, because they were put in different groups. By summing these talker-by-talker matrices across listeners, a single-talker similarity matrix was constructed with gradient values of similarity between 0 and 38 (the total number of participants). Note that in this analysis, group membership is binary (same or different group) and the perceptual similarity of intragroup talkers is not evaluated. Given that the listeners in this experiment were simply instructed to produce groups based on dialect similarity, it would be inappropriate to analyze their responses in terms of intragroup distances.

The first set of analyses focused on overall classification strategy, including the mean, median, maximum, and minimum number of talkers per group. When the number of groups is not predetermined, a similar set of analyses on the number of groups produced (including mean, median, and range) can also be computed (see Clopper & Pisoni, 2007). The second set of analyses focused on the perceptual similarity of the talkers and included two types of models: clustering and multidimensional scaling (MDS). Clustering analyses are useful for interpreting how objects group together perceptually, whereas MDS analyses provide insight into the most salient perceptual dimensions of similarity. All of the analyses discussed below are based on the assumption that perceptual similarity is symmetric; the similarity of talker *i* to talker *j* is equivalent to the similarity of talker *j* to talker *i*. This assumption holds true for the similarity data obtained in the free classification task, because the number of times that *i* and *j* were grouped together across listeners reflects the similarity of *i* to *j* and *j* to *i*.

Two of the most commonly used clustering models are hierarchical and k-means. K-means clustering requires the experimenter to choose the desired number of clusters. Thus, in order to use k-means clustering, the experimenter must already have an idea about the number of perceptual categories that participants used to perform the task. Although the participants in the present study were asked to make six groups of talkers, they did not necessarily come into the experiment with a six-category perceptual structure of regional dialect variation in the United States. Thus, despite the task instructions, a six-means clustering analysis might not produce the best model of the perceptual dialect similarity space for these listeners. Hierarchical clustering produces solutions in which all intracluster distances are equal and are shorter than all intercluster distances. This type of structure is intuitively false for many kinds of data and numerically false for virtually all real data. On the other hand, a third alternative for clustering analyses, additive similarity tree models, do not require an a priori decision about the number of clusters, and do not assume that all objects are equidistant from the root of the tree; they are therefore preferable to k-means and hierarchical clustering, because they

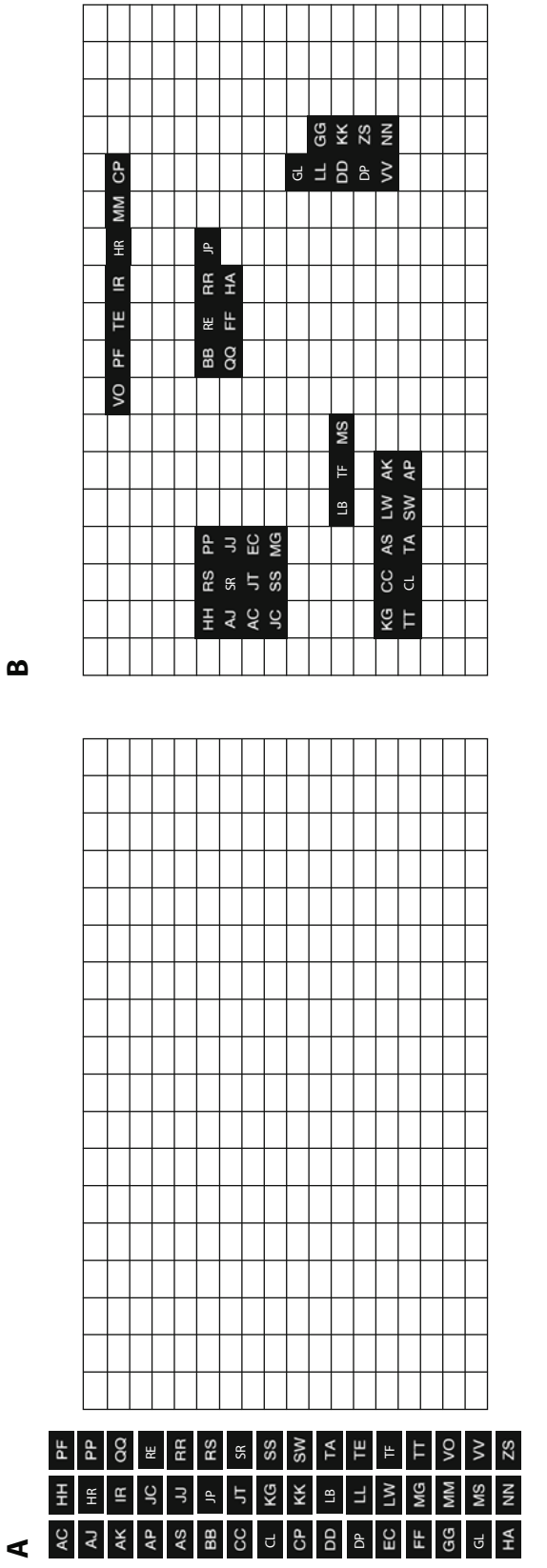

Figure 1. Sample PowerPoint presentation of the free classification task, with 48 stimulus items. A 16 × 16 grid (A) before the task and (B) after the task.

are able to model a wider range of relationships between objects. This flexibility is particularly important when complex stimulus materials are used and the potentially relevant dimensions of perceptual similarity are not all known or well specified in advance.

MDS analyses produce a different kind of spatial model of perceptual similarity. Whereas clustering analyses fit similarity data using an iterative pairwise distance calculation, MDS analyses produce the best fit for the entire distance matrix in a specified number of orthogonal dimensions. Thus, in some cases, objects that are close in one model might be farther away in the other. When conducting an MDS analysis, it is important to keep several aspects of the model in mind. First, the perceptual similarity space derived from the analysis is invariant with respect to rotation, reflection, and scale. The space can be rotated to find the most interpretable dimensions, but direct comparisons across two different solutions are impossible. Second, as the number of dimensions increases, the model returns a better fit to the data. Selecting the number of dimensions to interpret requires striking a balance between minimizing the stress (or badness of fit) of the solution and maximizing interpretability. Researchers will typically plot stress by dimensionality to look for the "elbow," or the number of dimensions past which stress no longer greatly improves (see Figure 3). A stress value under 0.1 is considered to be a good fit, but with complex stimulus materials such as speech, that level of model fit is often not attainable even with a relatively large number of dimensions.

When two or more groups of participants are used in a single experiment, differences between the groups can be assessed using individual differences scaling (INDSCAL) analyses (Carroll & Chang, 1970). In an INDSCAL analysis, one similarity matrix is created per listener and all of the individual matrices are then analyzed together. The INDSCAL analysis returns one similarity space for the entire set of listeners, as well as a set of dimension weights for each of the listeners. Like regular MDS analyses, the number of dimensions to interpret is selected on the basis of fit of the solution and interpretability of the dimensions. Unlike the solution in a regular MDS analysis, however, the similarity space obtained from an INDSCAL analysis cannot be rotated, and must be interpreted with respect to the dimensions returned by the model. The dimension weights can be analyzed statistically to compare the listener groups. For example, in a two-dimensional solution, one listener group may attend to Dimension 1 more than to Dimension 2, whereas a second listener group may attend to Dimension 2 more than to Dimension 1.

RESULTS

Across all of the listeners, groups ranged in size from 1 to 24 talkers, with a median of 8. Thus, although participants were instructed to produce exactly six groups of talkers, the groups varied considerably in size. In addition, none of the 38 listeners produced six groups with exactly eight talkers in each group.

An additive similarity tree analysis (Corter, 1982) of the talkers in the regional dialect classification experiment was conducted to examine talker similarity. The additive tree analysis returned two primary clusters: linguistically marked dialects (e.g., Mid-Atlantic, North, and South) and General American dialects (e.g., New England, Midland, and West). The marked cluster was further divided into Northern (Mid-Atlantic and North) and Southern clusters. The General American cluster was further divided into male and female clusters. A partial dendrogram of the additive tree solution is shown in Figure 2. In the earlier study, Clopper and Pisoni (2007) conducted clustering analyses on dialect similarity matrices instead of talker similarity matrices, so their results are not directly comparable to those obtained in the present experiment. However, the clustering analysis of the present data was interpretable in terms of dialect markedness, geography (northern vs. southern), and gender. This interpretation is consistent with the dimensions of perceptual dialect similarity obtained in MDS analyses of free classification (Clopper & Pisoni, 2007) and similarity rating (Clopper et al., 2006) tasks.

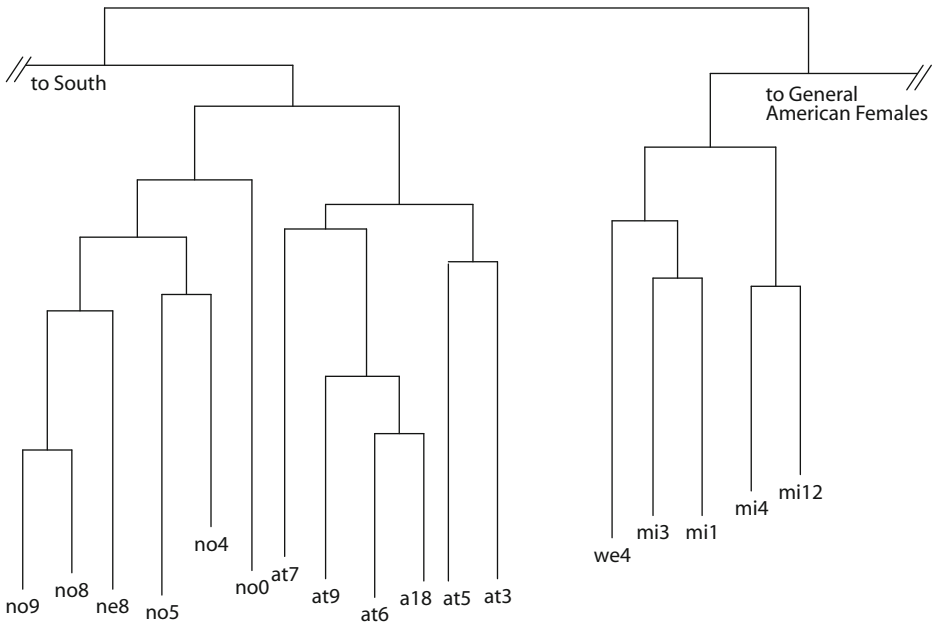

Figure 2. Partial dendrogram of the additive similarity tree clustering solution, including the marked Northern cluster and the male General American cluster. Talker labels indicate region of origin (at, Mid-Atlantic; mi, Midland; ne, New England; no, North; we, West) and gender (1–5 indicate male talkers, 6–0 indicate female talkers).

MDS analyses were conducted in one, two, three, and four dimensions, and the two-dimensional solution was selected for interpretation. Stress was substantially reduced from the one- to two-dimensional solution, but less reduced from two to three dimensions. Figure 3 shows a scree plot of the stress values obtained for each of the four MDS solutions (filled diamonds), with a slight elbow at two dimensions. In addition, the two-dimensional solution was interpretable without rotation, whereas the three-dimensional solution did not produce three interpretable dimensions under any rotation. As is shown in Figure 4, the first dimension of the two-dimensional solution separates the marked dialects (Mid-Atlantic, North, and South) from the General American dialects (New England, Midland, and West). The second dimension separates the northern dialects (New England, Mid-Atlantic, and North) from the non-northern dialects (Midland, South, and West).

INDSCAL analyses were conducted in two, three, and four dimensions. Dimension weights in a one-dimensional solution are undefined in the INDSCAL model, so a one-dimensional solution was not obtained. The two-dimensional solution was selected for interpretation. As is shown in Figure 3 (open circles), the stress value of the two-dimensional solution was low and did not substantially improve with the addition of a third dimension. In addition, the two-dimensional solution was interpretable. As is shown in Figure 5, the first dimension of the two-dimensional solution separates the marked dialects (Mid-Atlantic, North, and South) from the General American dialects (New England, Midland, and West). The second dimension separates the Northern dialects (New England, Mid-Atlantic, and North) from the non-Northern dialects (Midland, South, and West). These two dimensions are very similar to those obtained in the MDS analysis, except that the orientation of Dimension 1 is reversed across the two analyses. Since MDS and INDSCAL solutions are invariant with respect to reflection, this difference in orientation is insignificant. Thus, the overall perceptual similarity spaces obtained in the MDS and INDSCAL analyses were highly comparable.

The listener-specific dimension weights obtained in the INDSCAL analysis were analyzed to examine the effects of residential history on dialect classification performance. For all of the listeners, Dimension 1 was weighted more heavily ($M = .68$) than Dimension 2 ($M = .59$). Independent-samples t tests revealed no significant differences in the weights for Dimension 1 or Dimension 2 between the lifetime residents of the Midland dialect region and the listeners who had lived in multiple dialect regions. Taken together, the clustering, MDS, and INDSCAL analyses all suggest that linguistic markedness and geography are important dimensions of perceptual similarity for regional dialects of American English. In addition, the results of the INDSCAL analysis suggest that dialect classification was not affected by residential history for this group of undergraduate listeners.

DISCUSSION AND FUTURE DIRECTIONS

The present illustrative experiment replicated previous results reported by Clopper and Pisoni (2007) on the perceptual similarity of regional dialects of American English. In the earlier experiment, participants were permitted to make as many groups of talkers as they wished, but in the present experiment the participants were required to make exactly six groups of talkers. Despite this difference in the instructions, however, the results of both studies revealed two primary dimensions of perceptual dialect similarity: marked versus General American dialects, and Northern versus non-Northern dialects. Clopper and Pisoni also used an INDSCAL analysis to compare free classification performance across four different listener groups based on listener experience; and, as was the case in the present study, they did not observe significant differences due to residential history. However, Clopper and Bradlow (2007) found significant effects of experience on free classification of American English dialects when they compared native with nonnative listeners, using an INDSCAL analysis. Thus, the free classification task, in combination with clustering and scaling analyses, can re-

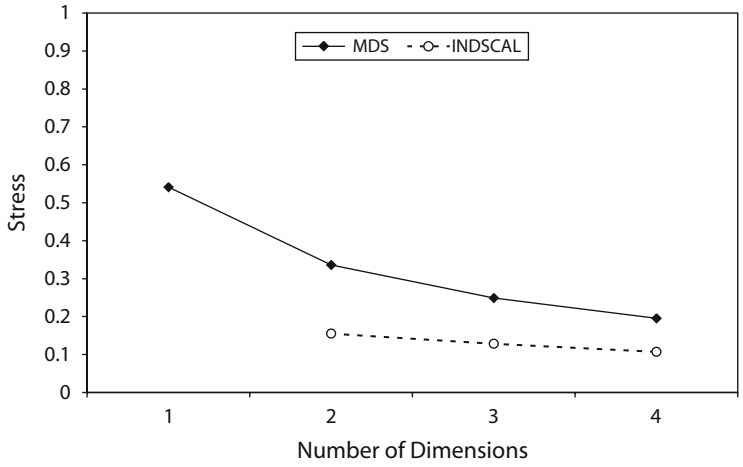

Figure 3. Scree plot showing the stress for each MDS solution from one to four dimensions and each INDSCAL solution from two to four dimensions.

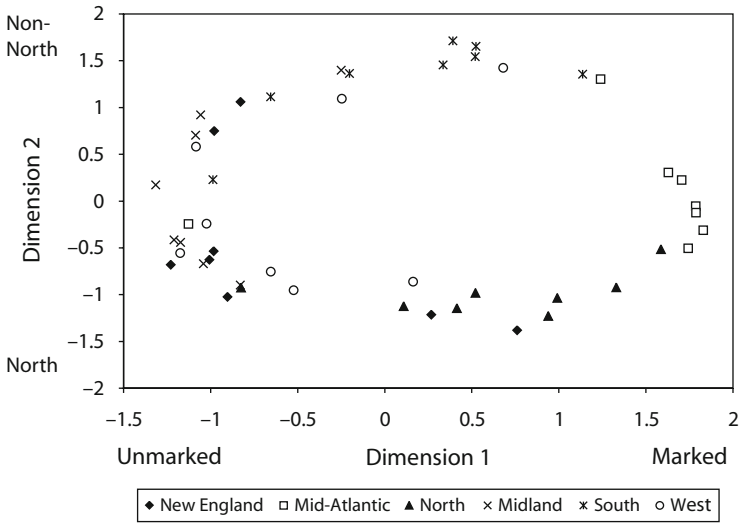

Figure 4. The two-dimensional MDS solution.

veal differences in perceptual similarity structures across listener groups.

The free classification paradigm is inherently flexible and permits a range of experimental designs. The number of groups or the number of items per group can be specified by the experimenter (Esposito, 2006) or left unspecified (McAdams et al., 2004), depending on the nature of the research question. In addition, the spatial nature of the task lends itself well to experiments designed to obtain more direct distance or spatial location measures. For example, the task could be modified to require participants to organize stimulus materials along a single dimension to obtain results similar to those in a magnitude estimation task (Granqvist, 2003). With respect to regional dialect variation, participants could be asked to indicate relative distance in a single dimension from General American English or relative position on a north–south axis to confirm the salience of those perceptual dimensions of similarity for undergraduate listeners. Alternatively, the two-dimensional space could be defined in terms of Cartesian or polar coordinates. In a regional dialect classification task, listeners could be asked to use a two-dimensional space with a map superimposed on it to indicate region of origin, to organize the talkers by perceptual similarity in two dimensions without the constraint of creating discrete groups, or to organize the talkers around a central point such as Standard American English. Analysis of data obtained using these kinds of experimental designs could include not only binary comparisons of group membership, but also physical Euclidean or polar distances between stimulus items in the classification space to refine our understanding of the perceptual dimensions of dialect similarity.

The auditory free classification task could also be applied to domains beyond regional dialect variation. For

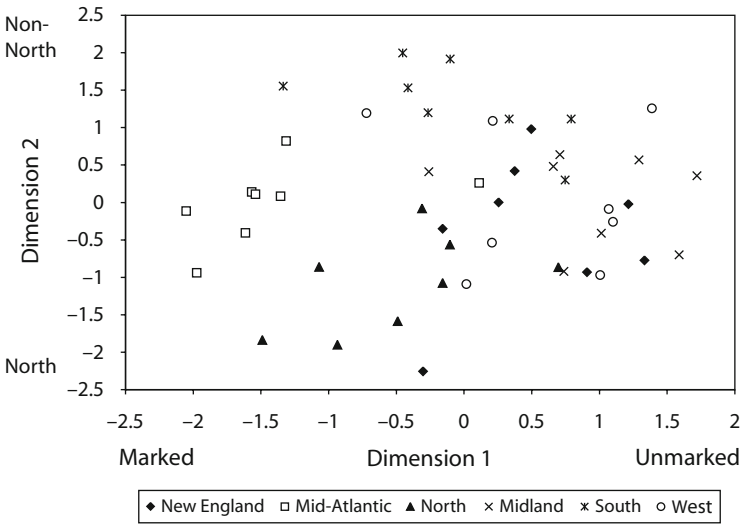

Figure 5. The two-dimensional INDSCAL solution.

example, numerous studies have examined the perceptual similarity of other speech and language phenomena, including voices (e.g., Walden, Montgomery, Gibeily, Prosek, & Schwartz, 1978), voice quality (e.g., Kreiman, Gerratt, & Berke, 1994), and vowel quality (e.g., Fox, 1983). Free classification has already been applied successfully to studies of the perceptual similarity of musical themes (McAdams et al., 2004), voice quality (Esposito, 2006), and synthetic speech (Granqvist, 2003), in addition to regional dialect variation. The free classification task can be used to obtain data similar to those obtained in paired comparison studies of perceptual similarity, but typically in much less time. In addition, the flexibility of the free classification paradigm, in terms of both specific instructions provided to the participants and the nature of the visual interface, provides an opportunity for researchers to obtain rich sets of data on perceptual similarity structures of many kinds of auditory stimulus materials.

The free classification paradigm can successfully be applied to auditory perception research using commonly available software tools. The rich data set obtained from free classification experiments can be examined using a range of statistical methods to produce converging evidence for the structure of perceptual similarity. These data can also be compared with data obtained in other commonly used paradigms, such as similarity rating, categorization, and discrimination tasks.

AUTHOR NOTE

This work was supported by NIH NRSA Postdoctoral Fellowship F32 DC007237 to Northwestern University and NIH Research Grant R01 DC00111 to Indiana University. Correspondence concerning this article should be addressed to C. G. Clopper, Department of Linguistics, Ohio State University, 222 Oxley Hall, 1712 Neil Avenue, Columbus, OH 43210 (e-mail: clopper.1@osu.edu).

REFERENCES

Carroll, J. D., & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an *n*-way generalization of "Eckart–Young" decomposition. *Psychometrika*, **35**, 238-319.

Clopper, C. G., & Bradlow, A. R. (2007, August). *Native and nonnative perceptual dialect similarity spaces*. Paper presented at the 16th International Congress of Phonetic Sciences, Saarbrücken, Germany.

Clopper, C. G., Levi, S. V., & Pisoni, D. B. (2006). Perceptual similarity of regional varieties of American English. *Journal of the Acoustical Society of America*, **119**, 566-574.

Clopper, C. G., & Pisoni, D. B. (2007). Free classification of regional dialects of American English. *Journal of Phonetics*, **35**, 421-438.

Clopper, C. G., Pisoni, D. B., & de Jong, K. (2005). Acoustic characteristics of the vowel systems of six regional varieties of American English. *Journal of the Acoustical Society of America*, **118**, 1661-1676.

Corter, J. E. (1982). ADDTREE/P: A PASCAL program for fitting additive trees based on Sattath and Tversky's ADDTREE algorithm. *Behavior Research Methods & Instrumentation*, **14**, 353-354.

Esposito, C. M. (2006). *The effects of linguistic experience on the perception of phonation.* Unpublished doctoral dissertation, University of California, Los Angeles.

Fox, R. A. (1983). Perceptual structure of monophthongs and diphthongs in English. *Language & Speech*, **26**, 21-60.

Gattuso, B., Smith, L. B., & Treiman, R. (1991). Classifying by dimensions and reading: A comparison of the auditory and visual modalities. *Journal of Experimental Child Psychology*, **51**, 139-169.

Granqvist, S. (2003). The visual sort and rate method of perceptual evaluation in listening tests. *Logopedics, Phoniatrics, Vocology*, **28**, 109-116.

Imai, S. (1966). Classification of sets of stimuli with different stimulus characteristics and numerical properties. *Perception & Psychophysics*, **1**, 48-54.

Imai, S., & Garner, W. R. (1965). Discriminability and preference for attributes in free and constrained classification. *Journal of Experimental Psychology*, **69**, 596-608.

Kreiman, J., Gerratt, B. R., & Berke, G. S. (1994). The multidimensional nature of pathologic vocal quality. *Journal of the Acoustical Society of America*, **96**, 1291-1302.

Labov, W. (1998). The three dialects of English. In M. D. Linn (Ed.), *Handbook of dialects and language variation* (pp. 39-81). San Diego: Academic Press.

Labov, W., Ash, S., & Boberg, C. (2006). *The atlas of North American English*. Berlin: Mouton de Gruyter.

Marcell, M. M., Borella, D., Greene, M., Kerr, E., & Rogers, S. (2000). Confrontation naming of environmental sounds. *Journal of Clinical & Experimental Neuropsychology*, **22**, 830-864.

McAdams, S., Vieillard, S., Houix, O., & Reynolds, R. (2004). Perception of musical similarity among contemporary thematic materials in two instrumentations. *Music Perception*, **22**, 207-237.

Medin, D. L., Wattenmaker, W. D., & Hampson, S. E. (1987). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive Psychology*, **19**, 242-279.

Smith, J. D., & Baron, J. (1981). Individual differences in the classification of stimuli by dimensions. *Journal of Experimental Psychology: Human Perception & Performance*, **7**, 1132-1145.

Thomas, E. R. (2001). *An acoustic analysis of vowel variation in New World English*. Durham, NC: Duke University Press.

Walden, B. E., Montgomery, A. A., Gibeily, G. J., Prosek, R. A., & Schwartz, D. M. (1978). Correlates of psychological dimensions in talker similarity. *Journal of Speech & Hearing Research*, **21**, 265-275.

(Manuscript received April 20, 2007;
revision accepted for publication December 7, 2007.)