

## Research Article

# Free-Classification of Perceptually Similar Speakers With Dysarthria

Kaitlin L. Lansford,<sup>a</sup> Julie M. Liss,<sup>a</sup> and Rebecca E. Norton<sup>a</sup>

**Purpose:** In this investigation, the construct of perceptual similarity was explored in the dysarthrias. Specifically, we employed an auditory free-classification task to determine whether listeners could cluster speakers by perceptual similarity, whether the clusters mapped to acoustic metrics, and whether the clusters were constrained by dysarthria subtype diagnosis.

**Method:** Twenty-three listeners blinded to speakers' medical and dysarthria subtype diagnoses participated. The task was to group together (drag and drop) the icons corresponding to 33 speakers with dysarthria on the basis of how similar they sounded. Cluster analysis and multidimensional scaling (MDS) modeled the perceptual dimensions underlying similarity. Acoustic metrics and perceptual judgments were used in correlation

analyses to facilitate interpretation of the derived dimensions.

**Results:** Six clusters of similar-sounding speakers and 3 perceptual dimensions underlying similarity were revealed. The clusters of similar-sounding speakers were not constrained by dysarthria subtype diagnosis. The 3 perceptual dimensions revealed by MDS were correlated with metrics for articulation rate, intelligibility, and vocal quality, respectively.

**Conclusions:** This study shows (a) feasibility of a free-classification approach for studying perceptual similarity in dysarthria, (b) correspondence between acoustic and perceptual metrics to clusters of similar-sounding speakers, and (c) similarity judgments transcended dysarthria subtype diagnosis.

The gold standard for classification of motor speech disorders, known as the Mayo Clinic approach, was set forth by Darley, Aronson, and Brown (1969a, 1969b, 1975) and was developed further by Duffy (2005). In their seminal work, Darley et al. (1969b) rated 38 dimensions of speech and voice observed in 212 patients with dysarthria arising from seven different neurological conditions. Seven subtypes of dysarthria (flaccid, spastic, ataxic, hypokinetic, hyperkinetic dystonia, hyperkinetic chorea, and mixed), each possessing unique but overlapping clusters of perceptual features, were delineated. The key to the classification system is that the underlying pathophysiology of each type of dysarthria is presumed responsible for the resulting clusters of perceptual features. For example, cerebellar lesions affect the timing, force, range, and direction of limb movements and result in

dysrhythmic, irregular, slow, and inaccurate actions. According to the Mayo Clinic approach, the speech features associated with ataxic speech (e.g., imprecise consonants, equal and even stress, and irregular articulatory breakdown) can be explained by the effects of cerebellar lesions on neuromuscular activity (as seen in the limbs). Thus, the explanatory relationship between locus of damage and the perceptual features associated with a dysarthria provides a valid and useful framework for clinical practice as well as research on motor speech disorders.

This expert, analytic evaluation of dysarthric speech is designed specifically to extract information relevant to differential diagnosis of dysarthria, which then serves as a source of corroborating information in the broader diagnosis of neurological disease or injury. However, such a level of analysis is unlikely to uncover unique, etiology-based communication disorders because, as Darley et al.'s (1969a, 1969b, 1975) work revealed, (a) not all speakers with a similar etiology exhibit similar speech symptoms, (b) speech symptoms within a given classification may differ along the severity dimension, and (c) there is considerable overlap in speech symptoms among the classification categories (e.g., imprecise consonants, slow rate). This gives rise to a gap in the ability to effectively leverage a dysarthria subtype

<sup>a</sup>Arizona State University, Tempe

Correspondence to Kaitlin L. Lansford: klansford@fsu.edu

Kaitlin L. Lansford is now at the School of Communication Sciences and Disorders, Florida State University.

Editor: Jody Kreiman

Associate Editor: Wolfram Ziegler

Received July 6, 2013

Revision received November 30, 2013

Accepted June 27, 2014

DOI: 10.1044/2014\_JSLHR-S-13-0177

**Disclosure:** The authors have declared that no competing interests existed at the time of publication.

diagnosis to identify appropriate treatment targets to address the resulting communication disorder. In the present report, we attempt to bridge this gap by exploring a paradigm that exploits a relatively simple level of analysis, namely, perceptual similarity. The hypothesis is that if listeners are able to identify clusters of similar sounding dysarthric speakers, listeners must be using perceptually salient features to accomplish the task. By extension, the features that underlie the perceptual clusters may be suitable candidates for treatment targets, to the extent that they contribute to the associated communication disorder. This line of reasoning is supported by work that demonstrates that dysarthric speech with similar acoustic-perceptual profiles challenges listener perceptual strategies (and outcomes) in specific ways (e.g., Borrie et al., 2012; Liss, 2007; Liss, Spitzer, Caviness, & Adler, 2002; Liss, Utianski, & Lansford, 2013).

The purpose of the present study was to investigate the construct of perceptual similarity in a heterogeneous cohort of speakers with dysarthria (i.e., speakers with varying dysarthria subtype diagnoses and severities). The speakers included in this study were recruited because the perceptual characteristics of their dysarthrias were consistent with the cardinal perceptual characteristics identified by Darley et al. (1969a, 1969b). Three research questions were addressed:

1. Can listeners cluster dysarthric speech samples without specific reference to speech features in an unconstrained free-classification task (see Clopper, 2008)?
2. Do the resulting clusters scale to meaningful or interpretable dimensions in the perceptual and acoustic domains?
3. To what extent do the freely classified clusters contain speakers with similar dysarthria subtype diagnoses?

## Method

### *Speakers and Stimuli*

Productions from 33 speakers were collected from a larger corpus of research in the Arizona State University Motor Speech Disorders laboratory (ASU MSD lab). Speakers were diagnosed with one of the following dysarthria subtypes by neurologists at the Mayo Clinic: ataxic dysarthria secondary to cerebellar degeneration ( $n = 11$ ), mixed flaccid-spastic dysarthria secondary to amyotrophic lateral sclerosis ( $n = 10$ ), hyperkinetic dysarthria secondary to Huntington's disease ( $n = 4$ ), and hypokinetic dysarthria secondary to Parkinson's disease ( $n = 8$ ). To be representative of previous research (Darley et al., 1969a, 1969b), speakers were selected on the basis of the presence of hallmark characteristics found within the Mayo Clinic classification system. Two speech-language pathologists (including the second author) concurred that the dysarthria type was consistent with the underlying medical diagnosis, and severity was rated to be moderate to severe (see Table 1).

All speaker stimuli were previously recorded and edited for use in a larger study conducted in the ASU MSD lab (e.g., Liss, Legendre, & Lotto, 2010; Liss et al., 2013, 2009). Each speaker read stimuli from visual prompts presented on a computer screen. All recordings utilized a head-mounted microphone (Plantronics DSP-100), and participants were seated in a sound-attenuating booth. Recordings were made using a custom script in TF32 (Milenkovic, 2004; 16 bit, 44 kHz) and were saved directly to disc for subsequent editing using commercially available software (SoundForge) to remove any noise or extraneous articulations before or after target utterances. For the purposes of this study, the sentence "The standards committee met this afternoon in an open meeting" was selected from the corpus of speech stimuli because of its diverse representation of speech sounds. Sentence durations across speakers were between 2.60 s and 13.544 s, with a mean duration of 6.486 s.

### *Listeners*

Twenty-three graduate students in communication disorders at ASU were recruited for this project. Participants were enrolled in a motor speech disorders class and had received basic instruction in both dysarthria and differential diagnosis. Listeners were native English speakers, passed a threshold hearing screening, and self-reported normal cognitive skills.

### *Procedure*

An auditory free-classification task, as detailed by Clopper (2008), was used to collect the similarity data. Free-classification is a perceptual sorting task, in which listeners are asked to group stimuli according to similarity. It was developed by cognitive psychologists interested in categorization of stimuli on the basis of perceptual dimensions undefined by the experimenter (Imai, 1966; Imai & Garner, 1965). Free-classification permits examination of perceptual similarity while avoiding experimenter-imposed categories and without naming distinctive perceptual characteristics. An attractive benefit of the free-classification method is that it is less time consuming than paired-comparison methods traditionally used to investigate perceptual similarity (Clopper, 2008). In the present study, the use of free-classification offered a faster and unconstrained listener task.

The stimulus materials (i.e., recordings of the sentence "The standards committee met this afternoon in an open meeting") produced by each of the speakers were embedded into a single PowerPoint slide and were presented to listeners. Each speaker's recording was randomly assigned a two-letter identifier (i.e., de-identified initials) to be used by listeners to keep track of the speakers during the free-classification task. The individual sound files and static images of the identifiers were merged in PowerPoint, such that when listeners double-clicked the image, the sound file played automatically. The merged files were placed neatly and randomly in three columns adjacent to a 16 × 16 cell

**Table 1.** Dysarthric speaker demographic information.

| Speaker | Gender | Age (years) | Dysarthria diagnosis | Etiology                  | Severity        |
|---------|--------|-------------|----------------------|---------------------------|-----------------|
| AF1     | F      | 72          | Ataxic               | Cerebellar ataxia         | Moderate        |
| AF2     | F      | 57          | Ataxic               | Multiple sclerosis/ataxia | Severe          |
| AF7     | F      | 48          | Ataxic               | Cerebellar ataxia         | Moderate        |
| AF8     | F      | 65          | Ataxic               | Cerebellar ataxia         | Moderate        |
| AF9     | F      | 86          | Ataxic               | Cerebellar ataxia         | Severe          |
| AM1     | M      | 73          | Ataxic               | Cerebellar ataxia         | Severe          |
| AM3     | M      | 79          | Ataxic               | Cerebellar ataxia         | Moderate–severe |
| AM4     | M      | 46          | Ataxic               | Cerebellar ataxia         | Moderate        |
| AM5     | M      | 84          | Ataxic               | Cerebellar ataxia         | Moderate        |
| AM6     | M      | 46          | Ataxic               | Cerebellar ataxia         | Moderate        |
| AM8     | M      | 63          | Ataxic               | Cerebellar ataxia         | Moderate        |
| ALSF2   | F      | 75          | Mixed                | ALS                       | Severe          |
| ALSF5   | F      | 73          | Mixed                | ALS                       | Severe          |
| ALSF6   | F      | 63          | Mixed                | ALS                       | Severe          |
| ALSF7   | F      | 54          | Mixed                | ALS                       | Moderate        |
| ALSF8   | F      | 63          | Mixed                | ALS                       | Moderate        |
| ALSF9   | F      | 86          | Mixed                | ALS                       | Severe          |
| ALSM1   | M      | 56          | Mixed                | ALS                       | Moderate        |
| ALSM4   | M      | 64          | Mixed                | ALS                       | Moderate        |
| ALSM7   | M      | 60          | Mixed                | ALS                       | Severe          |
| ALSM8   | M      | 46          | Mixed                | ALS                       | Moderate        |
| HDM8    | M      | 43          | Hyperkinetic         | HD                        | Severe          |
| HDM10   | M      | 50          | Hyperkinetic         | HD                        | Severe          |
| HDM11   | M      | 56          | Hyperkinetic         | HD                        | Moderate        |
| HDM12   | M      | 76          | Hyperkinetic         | HD                        | Moderate        |
| PDF5    | F      | 54          | Hypokinetic          | PD                        | Moderate        |
| PDF7    | F      | 58          | Hypokinetic          | PD                        | Moderate        |
| PDM8    | M      | 77          | Hypokinetic          | PD                        | Moderate        |
| PDM9    | M      | 76          | Hypokinetic          | PD                        | Moderate        |
| PDM10   | M      | 80          | Hypokinetic          | PD                        | Moderate        |
| PDM12   | M      | 66          | Hypokinetic          | PD                        | Severe          |
| PDM13   | M      | 81          | Hypokinetic          | PD                        | Moderate        |
| PDM15   | M      | 57          | Hypokinetic          | PD                        | Moderate        |

Note. A = ataxia; F = female; M = male; ALS = amyotrophic lateral sclerosis; HD = Huntington's disease; PD = Parkinson's disease.

grid in a single PowerPoint slide (see Figure 1). Each image was sized to fit precisely into one cell of the grid.

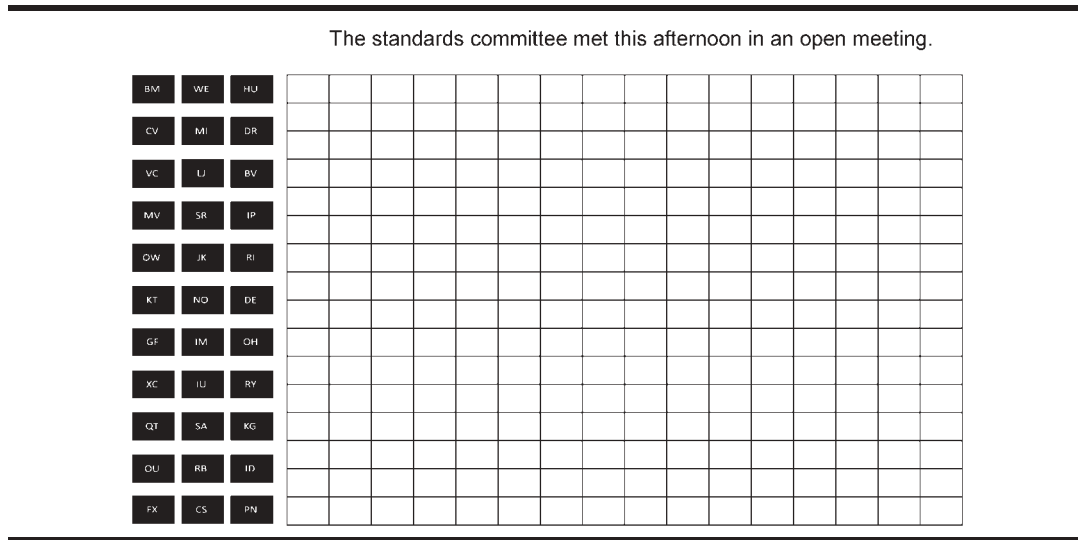
For the experimental task, listeners were seated in front of computers located in quiet listening cubicles. All computers were equipped with Sennheiser HD 280 sound-attenuating headphones and were calibrated using a digital sound level meter and a flat plate coupler. Volume was set individually on each computer, and participants did not adjust the volume. Listeners were informed that all of the speakers have dysarthria. However, the listeners did not know the underlying medical etiologies or dysarthria subtypes. Participants were instructed to listen to all of the merged sound files, via headphones, and to group the files in the grid (click, drag, drop) depending on how similar they sounded. Listeners were not provided any other instruction regarding how to make their judgments of similarity. They did not know the purpose of the study until they were debriefed. They were told that icons of the speakers perceived as sounding similar should be placed next to (touching) one another. Listeners were free to make as many groups as they deemed appropriate, with as many speakers in each group as needed (see Figure 2). There was no time limit imposed on the task,

and listeners were permitted to listen to and rearrange the speaker files as many times as necessary. Listeners recruited to participate in pilot testing indicated that the processing demands associated with the free-classification task taxed their working memory. Thus, listeners recruited for the present study were permitted, but were not required, to make notations as they made their similarity judgments in the notes space below the PowerPoint slide. These notes were saved for subsequent examination.

### *Acoustic and Perceptual Measurements*

Three sets of acoustic measures and one set of perceptual ratings were used in the correlation analyses (for detailed descriptions of the metrics, see Table 2). The sentence used in this study was one of five sentences produced by all speakers, whose classification results have been reported in previous work (Liss et al., 2010, 2009). Therefore, the first set of measures included previously published acoustic rhythm metrics (Liss et al., 2009) and envelope modulation metrics (Liss et al., 2010). The second set of metrics was designed to capture vowel space area and

**Figure 1.** Screen shot of the PowerPoint slide used for the free-classification task in its beginning position. Each of the initialed black icons located on the left side of the slide was paired with a specific speaker's sound file. When the icons were double-clicked, the sound file would play.



distinctiveness, and these data are reported in Lansford and Liss (2014a, 2014b).<sup>1</sup> The third new set of measures involved capturing the long-term average spectra (LTAS) of the sentences. Analysis of LTAS permits comprehensive exploration of the frequency distribution of a continuous speech sample and has been used in previous investigations of vocal quality (Leino, 2009; Shoji, Regenbogen, Yu, & Blaugrund, 1992), rhythmic disturbance in dysarthria (Utianski, Liss, Lotto, & Lansford, 2012), and perceived speech severity in dysarthria (Tjaden, Sussman, Liu, & Wilding, 2010).

Perceptual measures included scaled estimates of each speaker's overall severity, vocal quality, articulatory imprecision, nasal resonance, and prosodic disturbance.<sup>2</sup> The ratings of these dimensions were obtained via a visual analog task (Alvin Software; Hillenbrand, & Gayvert, 2005) completed by five speech-language pathologists (unaffiliated with the ASU MSD lab). The speakers' productions of the stimulus item used in the free-classification perceptual task were randomly presented, and the listeners were instructed to place a marker along a scale (ranging from *normal* to *severely abnormal*) that corresponded to their assessment of the speaker's level of impairment. Interrater reliabilities (Cronbach's alpha) for the ratings of severity, nasality, vocal quality, articulatory impression, and prosody were .936, .873, .898, .946, and .812, respectively. The ratings were normalized (*z* score) and averaged across listeners. Finally, intelligibility data (percentage of words correct on a transcription task)

<sup>1</sup>These vowel space metrics require inclusion of formant measures from vowels not present in the stimulus item used in the current experiment. Thus, formant measurements of vowels embedded in the phrases used in Lansford and Liss (2014a, 2014b) were used in the present analysis to facilitate calculation of the vowel space metrics.

<sup>2</sup>Listeners were instructed to judge prosodic disturbance without consideration of the speaker's overall speaking rate.

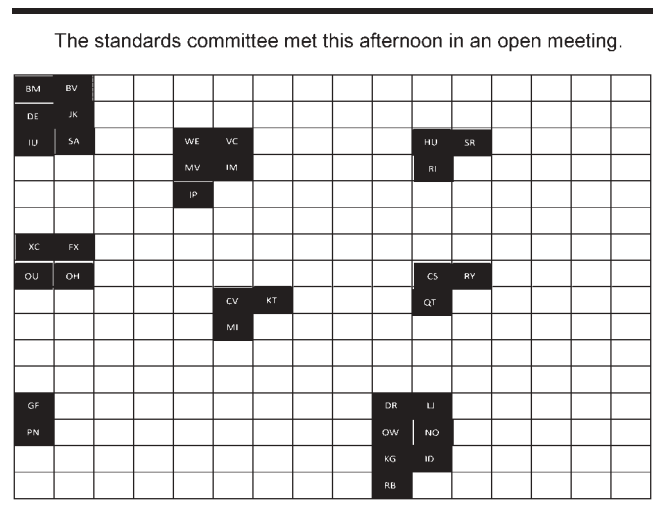
collected for these speakers (as reported in Liss et al., 2013) were included as a perceptual measure.

### Data Analysis

The PowerPoint slide with each listener's final speaker groupings was coded alphanumerically, and the final groupings were transferred into Microsoft Excel for subsequent analysis. Descriptive statistics were obtained to determine the mean, median, and range of numbers of listener-derived groups and speakers included in each group.

The similarity data obtained from each listener were arranged into a 33 × 33 speaker-similarity matrix in Excel (see the Appendix). A 1 was entered into cells corresponding

**Figure 2.** Screen shot of an example free-classification PowerPoint slide in its completed state, wherein icons touching one another in the grid were considered to be similar sounding.



**Table 2.** Descriptions of the acoustic and perceptual metrics.

| Metrics type             | Description  |
|--------------------------|--|
| Perceptual measures      |  |
| Intelligibility          | Percentage of words correct from a transcriptional task. Data were originally reported in Liss et al. (2013).  |
| Severity                 | Perceptual rating of global, integrated impression of dysarthria severity obtained from five SLPs.   |
| Vocal quality            | Perceptual rating of global, integrated impression of overall vocal quality obtained from five SLPs.   |
| Nasality                 | Perceptual rating of global, integrated impression of nasal resonance obtained from five SLPs.   |
| Articulatory imprecision | Perceptual rating of global, integrated impression of precision of articulatory gestures obtained from five SLPs.  |
| Prosody                  | Perceptual rating of global, integrated impression of speaker's rhythm, stress, and intonation obtained from five SLPs.  |
| Rhythm metrics           | Acoustic measures of vocalic and consonantal segment durations (Liss et al., 2009).  |
| ΔV                       | Standard deviation of vocalic intervals.   |
| ΔC                       | Standard deviation of consonantal intervals.   |
| %V                       | Percentage of utterance duration composed of vocalic intervals.  |
| VarcoV                   | Standard deviation of vocalic intervals divided by mean vocalic duration (× 100).  |
| VarcoC                   | Standard deviation of consonantal intervals divided by mean consonantal duration (× 100).  |
| VarcoVC                  | Standard deviation of vocalic + consonantal intervals divided by mean vocalic + consonantal duration (× 100).  |
| nPVI-V                   | Normalized pairwise variability index for vocalic intervals. Mean of the differences between successive vocalic intervals divided by their sum.  |
| rPVI-C                   | Pairwise variability index for consonantal intervals. Mean of the differences between successive consonantal intervals.  |
| rPVI-VC                  | Pairwise variability index for vocalic and consonantal intervals. Mean of the differences between successive vocalic and consonantal intervals.  |
| nPVI-VC                  | Normalized pairwise variability index for vocalic + consonantal intervals. Mean of differences between successive vocalic + consonantal intervals divided by their sum.  |
| Articulation rate        | Number of (orthographic) syllables produced per second, excluding pauses.  |
| EMS metrics              | The EMS variables were obtained for the full signal and for each of the octave bands (Liss et al., 2010).  |
| Peak frequency           | The frequency of the peak in the spectrum with the greatest amplitude. The period of this frequency is the duration of the predominant repeating amplitude pattern.  |
| Peak amplitude           | The amplitude of the peak described above (divided by overall amplitude of the spectrum).  |
| E3–E6                    | Energy in the region of 3–6 Hz (divided by overall amplitude of the spectrum). This is roughly the region of the spectrum (around 4 Hz) that has been correlated with intelligibility (Houtgast & Steeneken, 1985) and has been inversely correlated with segmental deletions (Tilsen & Johnson, 2008).  |
| Below 4                  | Energy in the spectrum from 0 to 4 Hz (divided by overall amplitude of the spectrum).  |
| Above 4                  | Energy in the spectrum from 4 to 10 Hz (divided by overall amplitude of the spectrum).   |
| Ratio 4                  | Below 4/above 4.   |
| LTAS metrics             | The measures of LTAS were normalized to RMS energy of entire signal and derived for 7 octave bands with center frequencies ranging from 125 to 8000 Hz (Utianski et al., 2012).  |
| N1RMS                    | RMS energy of entire signal and each octave band.  |
| nsd                      | Standard deviation of RMS energy (for 20-ms windows) of entire signal and each octave band.  |
| nRng                     | Range RMS energy (for 20-ms windows) of entire signal and each octave band.  |
| PV                       | Pairwise variability of RMS energy: mean difference between successive 20 ms windows of nRng. Computed for entire signal and for each octave band.   |
| Vowel metrics            | Vowel measures were derived from formant frequencies of vowel tokens embedded in six-syllable phrases produced by each speaker (Lansford & Liss, 2014a, 2014b).  |
| Quadrilateral VSA        | Vowel space area. Heron's formula was used to calculate the area of the irregular quadrilateral formed by the corner vowels (i, æ, a, u) in F1 × F2 space. Toward this end, the area (as calculated by Heron's formula) of the two triangles formed by the sets of vowels, /i/, /æ/, /u/, and /u/, /æ/, /a/, is summed. Heron's formula is as follows: $\sqrt{s(s-a)(s-b)(s-c)}$ , where s is the semiperimeter of each triangle, expressed as $s = \frac{1}{2}(a + b + c)$ , and a, b, and c each represent the Euclidean distance in F1 × F2 space between each vowel pair (e.g., /i/ to /æ/). |
| FCR                      | Formant centralization ratio. This ratio, expressed as $(F2_u + F2_a + F1_i + F1_u)/(F2_i + F1_a)$ , is thought to capture centralization when the numerator increases and the denominator decreases. Ratios greater than 1 are interpreted to indicate vowel centralization.  |
| Mean dispersion          | This metric captures the overall dispersion (or distance) of each pair of the 10 vowels, as indexed by the Euclidean distance between each pair in the F1 × F2 space.  |
| Front dispersion         | This metric captures the overall dispersion of each pair of the front vowels (i, ɪ, e, ε, æ). Indexed by the average Euclidean distance between each pair of front vowels in F1 × F2 space.  |
| Back dispersion          | This metric captures the overall dispersion of each pair of the back vowels (u, ʊ, o, a). Indexed by the average Euclidean distance between each pair of back vowels in F1 × F2 space.   |
| Corner dispersion        | This metric is expressed by the average Euclidean distance of each of the corner vowels, (i, æ, a, u), to the center vowel /ʌ/.  |
| Global dispersion        | Mean dispersion of all vowels to the global formant means (Euclidean distance in F1 × F2 space).   |

Note. SLPs = speech-language pathologists; EMS = envelope modulation spectra; LTAS = long-term average spectra; RMS: root-mean-square.

to two speakers grouped together by a listener. Likewise, a 0 was entered into the cells corresponding to speakers not grouped together. The individual listener's speaker-similarity matrices were summed and converted into a dissimilarity matrix for the subsequent analyses. First, the similarity data were subjected to an additive similarity tree cluster analysis described by Corter (1998) and used by Clopper (2008) to determine the number and composition of clusters of perceptually similar speakers. Multidimensional scaling (MDS) of the similarity data was completed to examine the salient perceptual dimensions underlying speaker similarity in this group of speakers with dysarthria. Correlation analysis was conducted to facilitate interpretation of the perceptual dimensions underlying similarity as revealed by the MDS. In addition, noncompulsory notes made by listeners as they completed the perceptual task were examined to determine whether the results of the quantitative analyses described above tracked to the acoustic and perceptual characteristics reported by the listeners to underlie speaker similarity in dysarthria.

## Results

### *Descriptive Analysis*

Listeners derived an average of 7.7 clusters ( $SD = 2.85$ ) of similar-sounding speakers, with a median of 7 and a range of 3–14 groups. The individual clusters of similar-sounding speakers included an average of 4.96 speakers ( $SD = 2.1$ ), with a median of 4 and a range of 1–13 speakers. See Figure 3 for the  $33 \times 33$  speaker-similarity matrix.

### *Cluster Analysis*

Additive similarity tree cluster analysis (Cortier, 1998) was used to identify clusters of similar sounding speakers. The results of the cluster analysis are best visualized via dendrogram representation of the similarity data, in which speakers were linked together one at a time at varying steps of the analysis (see Figure 3). Speakers that were most frequently grouped together by the listeners were linked first by the cluster analysis. Speakers joined existing clusters during subsequent steps of the analysis until all of the clusters joined to form a single group (see the top of Figure 3). The number of clusters revealed is experimenter defined. For the purposes of this initial foray into similarity in the dysarthrias, a six-cluster solution was selected. This solution was analyzed primarily because it most closely resembled the descriptive data (i.e., average number of groups derived by the listeners). Unfortunately, this solution left speaker AM3 without a cluster. He was, therefore, excluded from subsequent cluster-based analyses. It is important to note that the composition of each of the six clusters was not limited to a single dysarthria subtype (for cluster member distribution, see Table 3); however, one cluster contained all of the speakers diagnosed with Parkinson's disease and one speaker with Huntington's disease. Thus, these results support the notion that speaker similarity in dysarthria may transcend dysarthria diagnosis.

## MDS

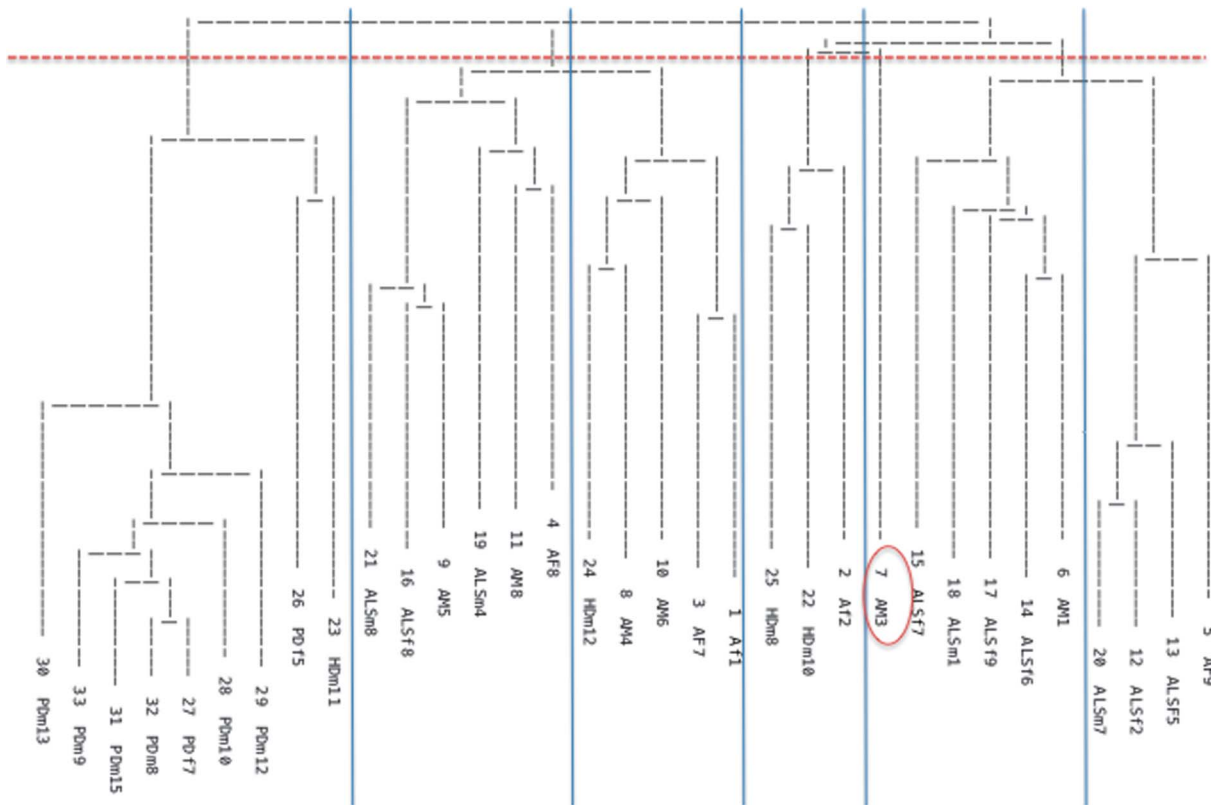
The similarity data were subjected to PROXSCAL MDS analysis (SPSS), and the normalized raw stress values obtained for models that included one to five dimensions were evaluated to determine the best fit of the data. Briefly, the stress of an MDS model refers to its overall stability. The normalized raw stress values of the one through five dimensional models were included in a scree plot to identify the point with which the addition of another dimension no longer substantially lowers stress (i.e., the "elbow" of the plot). The three-dimensional model (dispersion accounted for = .9907; normalized raw stress = .0096; and Stress 1 = .09621) was selected to facilitate visualization of the dimensions and to simplify subsequent interpretation. The clusters of similar-sounding speakers were plotted in the common space revealed by the MDS in Figures 4A, 4B, and 4C. In Figure 4A, the clusters were plotted in the two-dimensional space created by the first two dimensions derived by the MDS. Along the first dimension, Cluster 1 (composed mainly of speakers with Parkinson's disease) was clearly differentiated from the remaining five clusters. Furthermore, Clusters 2, 3, and 6 were well delineated in this space. Some overlap, though, is noted between Clusters 4 and 5. In Figure 4B, the clusters were plotted in the two-dimension space created by the first and third dimensions revealed by the MDS. Although substantial overlap of the clusters is evident in this representation of the common space, it is important to note that Clusters 4 and 5, indistinguishable in the space created by the first two dimensions, were well delineated by the third dimension.

### *Correlation Analysis*

A series of correlation analyses was conducted to interpret the abstract dimensions revealed by the MDS. Because of the large number of acoustic and perceptual measures considered in this analysis, the dependent variables that correlated most meaningfully with the dimensions are presented below. It is important to note that none of the vowel measures correlated significantly with any of the MDS dimensions. Thus, they will not be discussed in the following subsections.

*Dimension 1 (D1).* Across all acoustic and perceptual variables, there were strong correlations with those measures related to rate and rhythm (see Tables 4–7 for a full account of results). Most notably, D1 correlated strongly with the acoustic measure of articulation rate ( $r = -.888$ ). It also demonstrated a strong relationship with a measure of standard deviation of the durations of vocalic intervals ( $\Delta V$ ;  $r = .739$ ). In addition, several moderate relationships with the other segmental rhythm metrics were revealed. As reported in Liss et al. (2009), many of the rhythm metrics ( $\Delta V$  included) demonstrated strong relationships with articulation rate; therefore, it is not surprising that D1 correlated with the majority of these temporally based measures of rhythm.

**Figure 3.** Dendrogram derived from the cluster analysis. The dotted line corresponds to the solution selected for the present analysis. The solid lines demarcate cluster boundaries. One speaker, AM3 (circled in the figure), was not included in subsequent cluster-based analyses because of his late cluster linking.



As can be seen in Table 5, strong correlations also were found for envelope modulation spectra (EMS) measures that can be interpreted relative to articulation rate (Below 4 at 8000 Hz:  $r = .809$ ; and strong correlations with Below 4, Above 4, and Ratio 4 variables derived for most of the frequency bands). As reported by Liss et al. (2010), of all the EMS variables, Below 4 at 8000 Hz correlated most strongly with articulation rate ( $r = -.862$ ), but rate was also highly correlated with other 4-Hz variables in most of the frequency bands.

Conclusive support for D1 capturing articulation rate was not revealed for the LTAS measures (see Table 6). The strongest relationships between the LTAS measures

and D1 occurred for the pairwise variability measures in the 125-, 250-, and 500-Hz frequency bands. D1 was also weakly correlated with a few of the measures in the 1000-Hz band. Unlike EMS, very few of the LTAS measures were correlated with articulation rate, as would be expected. However, the LTAS variables that correlated significantly with D1 were also correlated with articulation rate, providing indirect support for the articulation rate hypothesis.

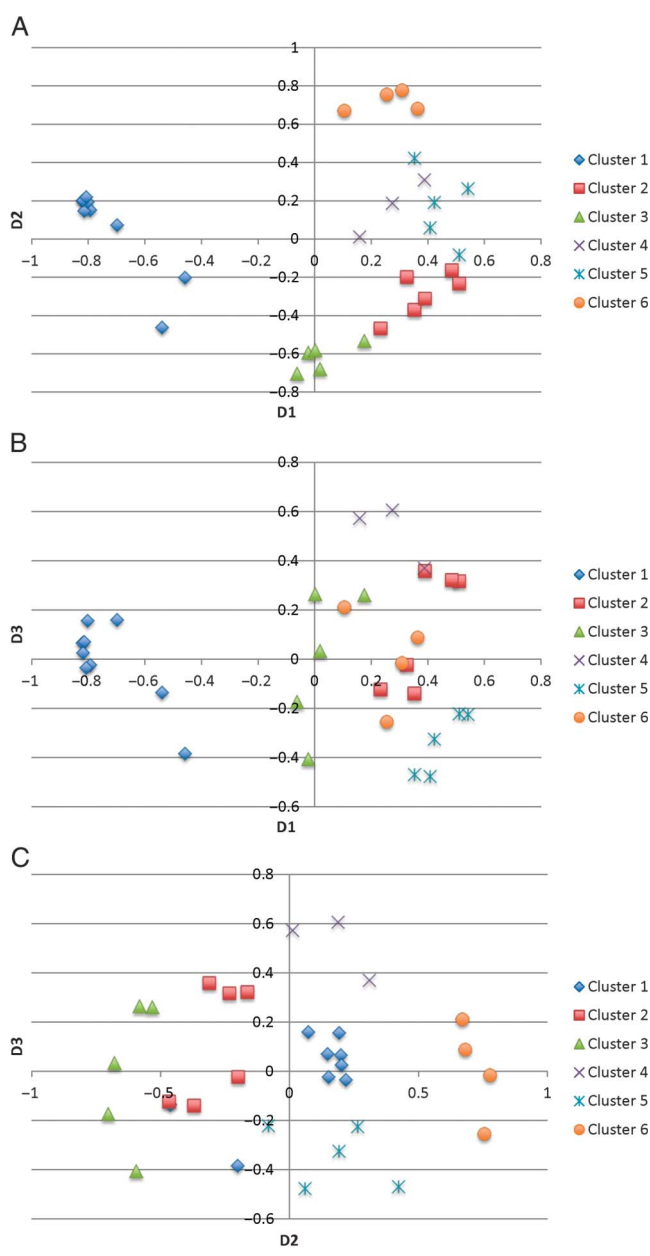
D1 was moderately to strongly correlated with all of the perceptual rating measures obtained from speech-language pathologists who scaled perceptual features of the sentences. Despite the strong relationships with the other perceptual ratings, D1 was not correlated with the intelligibility data collected for these speakers (see Table 7). Interestingly, articulation rate was moderately to strongly correlated with all of the perceptual rating measures ( $r$ s ranging from  $-.57$  to  $-.78$ ), but it was not correlated with intelligibility ( $r = .243$ ). Thus, it is possible that the relationship between D1 and articulation rate was responsible for the significant relationships between D1 and the perceptual ratings measures.

*Dimension 2 (D2).* Of all of the acoustic and perceptual variables, D2 was most strongly related to intelligibility ( $r = -.646$ ). D2 was also correlated with the perceptual ratings measures of severity, nasality vocal quality, articulatory

**Table 3.** Listener derived clusters with members.

| Cluster | Speakers  |
|---------|---|
| 1       | PDF5, PDF7, PDM8, PDM9, PDM10, PDM12, PDM13, PDM15, HDM11 |
| 2       | AF8, AM5, AM8, ALSF8, ALSM4, ALSM8                        |
| 3       | AF1, AF7, AM4, AM6, HDM12                                 |
| 4       | AF2, HDM8, HDM10  |
| 5       | ALSF6, ALSF7, ALSF9, ALSM1, AM1                           |
| 6       | AF9, ALSF2, ALSF5, ALSM7                                  |

**Figure 4.** (A) Listener-derived clusters plotted in the perceptual space created by the first two dimensions derived by multidimensional scaling (Dimension 1 [D1] and Dimension 2 [D2]). (B) Listener-derived clusters plotted in the perceptual space created by D1 and Dimension 3 (D3). (C) Listener-derived clusters plotted in the perceptual space created by D2 and D3.



imprecision, and prosody. The perceptual rating measures were significantly intercorrelated and were also significantly correlated with both articulation rate and intelligibility in this cohort of speakers. Intelligibility and articulation rate, however, were not correlated.

Like D1, D2 exhibited a few correlations with duration-based measures related to rate and rhythm, but less robustly. The strongest correlations from the rhythm metrics included

**Table 4.** Significant correlations between multidimensional scaling dimensions and temporal-based measures of rate and rhythm.

| Dimension 1 |          | Dimension 2 |          |
|-------------|----------|-------------|----------|
| Variable    | <i>r</i> | Variable    | <i>r</i> |
| Rate        | -.888    | nPVI-V      | -.583    |
| $\Delta V$  | .739     | %V          | .536     |
| rPVI-VC     | .55      | VarcoV      | -.447    |
| VarcoV      | -.547    |             |          |
| rPVI-C      | .543     |             |          |
| nPVI-V      | -.392    |             |          |
| VarcoVC     | -.368    |             |          |
| %V          | .346     |             |          |

*Note.* The metrics in the table are rank ordered by level of significance (all *ps* < .05). Insignificant correlations are not included.

a pair of intercorrelated measures of standard deviation of vocalic intervals that have been rate normalized (normalized pairwise variability index for vocalic intervals [nPVI-V] and standard deviation of vocalic intervals divided by mean vocalic duration  $\times 100$  [VarcoV];  $r = -.583$  and  $-.447$ , respectively) and a measure of the proportion of the signal that is composed of vocalic intervals (percentage of utterance duration composed of vocalic intervals [%V];  $r = .536$ ). These measures were all significantly correlated with intelligibility (absolute *r* ranging from .395 to .512).

With regard to EMS metrics, D2 correlated significantly with many of the same variables that were correlated with the D1, albeit less strongly. A notable deviation from this pattern of results, however, was the moderate relationships between the D2 and the E3–E6 (energy in the region of 3–6 Hz) variables in most of the frequency bands. Interestingly, the E3–E6 variables were not correlated with D1. Liss et al. (2010) derived these variables largely because the energy in this region of the spectrum has been shown to be correlated with intelligibility (Houtgast & Steeneken, 1985). Indeed, the E3–E6 variables, particularly in the higher frequency bands (e.g., 4000- and 8000-Hz bands), were significantly correlated with the intelligibility data collected for these speakers (e.g.,  $r = .561$  and  $.597$ , respectively). It should also be noted that many significant correlations were found between the EMS variables (particularly Below 4, After 4, and Ratio 4) and the perceptual measures, including intelligibility and the ratings of severity, vocal quality, nasality, articulatory imprecision, and prosody.

For the LTAS metrics, D2 correlated significantly with all of the variables in the 4000-Hz band ( $r$  ranging from  $-.559$  to  $-.623$ ). In addition, slightly less significant relationships were revealed for the LTAS variables in the 8000-Hz band and D2. All of these LTAS variables were moderately related to the perceptual measures, including intelligibility and severity.

*Dimension 3 (D3).* Overall, D3 generally had weaker correlations with all acoustic and perceptual measures than did D1 and D2. None of the duration-based measures of rate and rhythm correlated significantly with D3. Modest



**Table 5.** Significant correlations between multidimensional scaling dimensions and measures of EMS.

| Dimension 1                 |                | Dimension 2                 |                | Dimension 3         |          |
|-----------------------------|----------------|-----------------------------|----------------|---------------------|----------|
| Variable                    | <i>r</i> range | Variable                    | <i>r</i> range | Variable            | <i>r</i> |
| Below 4                     | .391 to .809   | E3–E6                       | –.433 to –.623 | Peak amplitude 2000 | –.469    |
| Above 4                     | –.529 to –.731 | Peak frequency <sup>a</sup> | .346 to .568   | Peak amplitude 1000 | –.449    |
| Ratio 4                     | .539 to .692   | Above 4 <sup>b</sup>        | –.436 to –.568 |                     |          |
| Peak frequency              | .41 to .605    | Ratio 4                     | .361 to .602   |                     |          |
| Peak amplitude <sup>c</sup> | .383 to –.695  |                             |                |                     |          |

*Note.* Because of the large number of correlated variables, the results for Dimensions 1 and 2 are summarized, and the range of correlation coefficients is reported. The metrics in the table are rank ordered by level of significance (all  $ps < .05$ ). Insignificant correlations are not included.

<sup>a</sup>Not significant for the 250-Hz band or for the full spectrum. <sup>b</sup>Significant correlations found for the 250-, 2000-, 4000-, and 8000-Hz bands.

<sup>c</sup>Significant correlations found for the 125-, 250-, 1000-, and 8000-Hz bands and for the full spectrum.

relationships between D3 and an EMS measure of the dominant modulation rate were revealed for the 1000- and 2000-Hz bands (peak amplitude;  $r = -.449$  and  $-.469$ , respectively). Although these metrics are thought to reflect, in part, speech rhythm, there is no straightforward interpretation in this context. None of the vowel space measures correlated significantly with D3. However, D3 was significantly correlated with all of the LTAS measures in the 8000-Hz band. Although no direct interpretation for the metrics derived in this octave band exists, it has been demonstrated that the spectral peaks of most English fricatives are found in the higher frequency bands of the spectrum (voiceless > voiced; Hughes & Halle, 1956; Jongman, Waylung, & Wong, 2000; Maniwa, Jongman, & Wade, 2009). In addition, increased energy in high frequency LTAS (>5000 Hz) has been found for speakers with breathy vocal quality (Shoji et al., 1992). This finding was corroborated by the results of a recent study that found the energy in high frequency LTAS for softly produced speech was relatively greater than that of loudly produced speech when LTAS for the conditions was normalized for overall sound pressure level (Monson, Lotto, & Story, 2012).

Of all of the perceptual measures, D3 correlated only with the rating of vocal quality ( $r = -.414$ ). Recall that the combined results of the cluster and MDS analyses demonstrated that Clusters 4 and 5 were not well delineated by the first two dimensions (see Figure 4A); however, with inclusion of the third dimension, they were separated. A post hoc analysis (i.e., one-way analysis of variance with multiple comparisons) was conducted to determine whether the clusters of speakers, particularly Clusters 4 and 5, possessed significantly different ratings of vocal quality. Indeed, a main effect,  $F(5, 31) = 20.838$ ,  $p < .0001$ , of cluster group on vocal quality rating was revealed. Inspection of the cluster means (see Table 8) revealed speakers belonging to Cluster 5 had the highest vocal quality ratings ( $M = 1.17$ ,  $SD = 0.4$ ), and Cluster 4 speakers had the lowest ( $M = 0.03$ ,  $SD = 0.21$ ). Bonferroni-corrected multiple comparisons demonstrated that the vocal quality ratings of the Cluster 5 speakers were significantly higher

(meaning more impaired) than those of the other clusters, with the exception of Cluster 6. Thus, the results of this post hoc analysis provide some evidence to support interpretation of D3 as one that may capture some aspects of vocal quality.

### Examination of Listener Notes

To assuage the processing and working memory demands placed on the listeners by the free-classification task, listeners were permitted to take notes as they grouped together similar-sounding speakers. This afforded an opportunity to qualitatively evaluate the perceptual relevance of the dimensions revealed by the MDS analysis. In total, 21 of the 23 listeners elected to make notations as they completed the task. We found that 100% of these listeners mentioned rate and rhythm of speech in their notations. This finding corresponds with the quantitative results that demonstrated that metrics capturing rate and rhythm were significantly correlated with a primary dimension underlying similarity in dysarthria. In addition, approximately 66% of listeners mentioned intelligibility in their notes. Again, this finding tracks to the results of our quantitative approach that revealed intelligibility was salient to similarity judgments. The third dimension revealed by MDS was correlated with the perceptual rating measure of vocal quality. Indeed, 85.7% of the listeners mentioned vocal quality characteristics in their notes. Other perceptual features mentioned by the listeners included the following: articulatory imprecision (62%), severity (23.8%), resonance (23.8%), prosody (23.8%), respiratory differences (19%), variable loudness (14.3%), pitch breaks (9%), word boundary errors (4.7%), and overall “bizarreness” (4.7%).

### Discussion

The contributions of the present study are threefold. First, results demonstrate proof-of-concept for the use of an auditory free-classification task in the study of perceptual similarity in the dysarthrias. Because the paradigm does not rely on a predetermined set of clustering variables,

**Table 6.** Significant correlations between multidimensional scaling dimensions and measures of LTAS.

| Dimension 1 |          | Dimension 2 |          | Dimension 3 |          |
|-------------|----------|-------------|----------|-------------|----------|
| Variable    | <i>r</i> | Variable    | <i>r</i> | Variable    | <i>r</i> |
| PV125       | -.583    | nsd4000     | -.623    | N1RMS8000   | -.493    |
| PV250       | -.52     | N1RMS4000   | -.599    | nsd8000     | -.474    |
| PV500       | -.517    | PV4000      | -.582    | PV8000      | -.447    |
| N1RMS1000   | .383     | nRng4000    | -.559    | nRng8000    | -.446    |
| nRng1000    | .377     | PV8000      | -.496    |             |          |
| nsd1000     | .367     | nsd8000     | -.462    |             |          |
|             |          | N1RMS8000   | -.451    |             |          |
|             |          | nRng8000    | -.419    |             |          |

*Note.* The metrics in the table are rank ordered by level of significance (all *ps* < .05). Insignificant correlations are not included. PV = pairwise variability.

listeners cluster by whatever similarities are salient to them. Until now it was not known whether dysarthria was amenable to free-classification by judges with minimal experience with dysarthria. The second contribution is the demonstration that the clusters were made along three dimensions and that these dimensions corresponded with independent acoustic and perceptual measures. The third contribution is that the results of this perceptual similarity task transcended dysarthria subtype diagnoses, providing support for the notion that the paradigm may provide clusters more closely linked with the nature of the communication disorder. These contributions and some of their limitations are detailed below.

### ***Proof-of-Concept***

Free-classification methods have been used to investigate perceptual similarity of environmental sounds (Guastavino, 2007; Gygi, Kidd, & Watson, 2007), musical themes (McAdams, Vieillard, Houix, & Reynolds, 2004), and American-English regional dialects (Clopper & Bradlow, 2009; Clopper & Pisoni, 2007). To our knowledge, perceptual similarity has not been previously directly assessed in the dysarthrias. Therefore, it was necessary to determine the appropriateness of free-classification techniques in the investigation of perceptual similarity in dysarthria. A feature of free-classification that made it appealing for the current study is that it liberates participants from experimenter-defined categories. For example, Clopper and Pisoni (2007) found that in using a free-classification task to investigate regional American-English dialects, listeners were able to make finer distinctions between dialectal speech patterns when specific labels were not experimenter imposed. The ability of the statistical analyses to adequately model the similarity data in concert with the finding that the perceptual dimensions underlying similarity correlated meaningfully with acoustic and perceptual metrics supports the use of free-classification methodology as a viable tool for the study of perceptual similarity in dysarthria.

In this initial assessment of perceptual similarity in dysarthria, it was necessary to make a variety of methodological decisions that were undoubtedly contributors to the cluster outcomes. Our targeted listeners were graduate students in communication disorders with basic familiarity with dysarthrias and the Mayo classification scheme—but with limited clinical exposure. We selected this group of listeners because they were expected to have more finely honed perceptual judgment skills than truly naïve participants, but they were expected to have less honed skills than those of clinicians experienced in the Mayo Clinic approach to differential diagnosis of dysarthria. Although supported by intuition, this assumption must be verified in the context of experimental design. To identify ecologically valid parameters contributing to similarity, it will be important to explore how listener variables—such as clinical sophistication, experience/exposure, perceptual astuteness/awareness, or even listening strategies—influence judgments of perceptual similarity. Toward this end, clustering data elicited from practicing speech-language pathologists on these same stimuli are presently being analyzed, the results of which will partially inform this question.

A second methodological decision was to use speech samples, specifically a single sentence, from speakers who ranged in speech impairment from moderate to severe and who were selected in a larger investigation because their speech exhibited perceptual characteristics associated with their dysarthria subtype diagnosis. There is every reason to believe that clustering decisions were influenced by both the speech sample used in the task and the constellation of speakers to be clustered, as this is a comparative task. Thus, it will be critical in future studies to assess the stability of perceptual decisions for a given speaker across a variety of speech samples and across groups that vary in speaker composition. Optimally, a free-classification paradigm will reveal the most perceptually salient parameters for any given group of speakers, irrespective of speech sample material, and that individual members of the clusters will be similar on these parameters. Computational modeling methods conducted on sets of clustering data will be important for establishing characteristics that influence judgment stability.

### ***Interpretability of the Dimensions Underlying Similarity***

MDS of the similarity data uncovered a minimum of three salient perceptual dimensions underlying similarity in this cohort of speakers. In addition, the similarity-based clusters of speakers were well delineated in these dimensions (see Figure 4). The results of the correlation analyses, which compared the abstract MDS dimensions with a host of acoustic and perceptual measures, provided important information that facilitated their interpretation. The interpretations of the first two dimensions were fairly straightforward: D1 correlated strongly with measures capturing articulation rate, and D2 correlated with measures capturing overall intelligibility. Interpretation of D3 was less

**Table 7.** Significant correlations between multidimensional scaling dimensions and perceptual measures of intelligibility, severity, vocal quality, nasality, articulatory imprecision, and prosody.

| Dimension 1            |          | Dimension 2            |          | Dimension 3   |          |
|------------------------|----------|------------------------|----------|---------------|----------|
| Variable               | <i>r</i> | Variable               | <i>r</i> | Variable      | <i>r</i> |
| Vocal quality          | .716     | Intelligibility        | -.646    | Vocal quality | -.414    |
| Severity               | .702     | Severity               | .632     |               |          |
| Nasality               | .632     | Prosody                | .624     |               |          |
| Articulatory precision | .544     | Articulatory precision | .622     |               |          |
| Prosody                | .544     | Nasality               | .55      |               |          |
|                        |          | Vocal quality          | .434     |               |          |

Note. The metrics in the table are rank ordered by level of significance (all  $p$ s < .05). Insignificant correlations are not included.

clear; however, results of a post hoc analysis demonstrated it is probably related to vocal quality characteristics. A number of features included in the listeners' notations (e.g., pitch breaks and resonance) were not revealed as contributing to similarity by the quantitative approaches used in the present analysis. It is important to note that the statistical techniques used in this investigation were largely linear, and it is likely that listeners' judgments of similarity are not always amenable to such approaches (e.g., potential binary decisions made by listeners regarding the absence or presence of a perceptual feature in a speaker or cluster of speakers). Thus, alternative techniques (e.g., logistic regression) should be considered as this line of research progresses. In addition, although a large number of acoustic and perceptual features were considered in this preliminary step, it was in no way exhaustive. The perceptual ratings of severity, vocal quality, nasality, articulatory imprecision, and prosody were useful in this analysis but are subjective and vulnerable to poor intra- and interrater reliability (e.g., Kreiman & Gerratt, 1988). Bunton, Kent, Duffy, Rosenbek, and Kent (2007) investigated intrarater and interrater agreement for the Mayo Clinic system's perceptual indicators (i.e., the 38 dimensions of speech and voice originally outlined by the Mayo Clinic) and found listener agreement to be highest when ratings of each dimension were at the endpoints of a 7-point scale (e.g., *normal* or *very severe deviation from normal*). In other words, there was greater variability in listener agreement in the middle of the scale. Thus, a rating scale that denotes the absence or presence of a perceptual feature may prove useful in subsequent investigations of similarity in dysarthria.

**Table 8.** Mean vocal quality ratings (z-score normalized) for each similarity-based cluster.

| Cluster | <i>N</i> | <i>M</i> | <i>SD</i> |
|---------|----------|----------|-----------|
| 1       | 9        | -0.7463  | 0.4015    |
| 2       | 6        | 0.0413   | 0.3566    |
| 3       | 5        | -0.5418  | 0.2325    |
| 4       | 3        | 0.0292   | 0.2114    |
| 5       | 5        | 1.1699   | 0.4013    |
| 6       | 4        | 0.7066   | 0.786     |

### **Relationship Between Dysarthria Diagnosis and Similarity-Based Clusters**

Given that the speakers used in this investigation were recruited because their speech exhibited the hallmark characteristics of their dysarthria diagnosis, one might expect that perceptual clustering would mirror dysarthria subtype more often than not. With the exception of Cluster 1, which was composed primarily of speakers with hypokinetic dysarthria with intact or fast speaking rate, this generally did not occur. Thus, the results of this analysis suggest that if we had sampled a random group of speakers with dysarthria (i.e., without selection of speakers on the basis of perceptual features or dysarthria diagnosis), a similarity-clustering paradigm would be successful in identifying speakers with common acoustic speech features. However, it is important to note that although the clusters were not constrained by dysarthria subtype, influence of disease process on perceived acoustic similarity was evident. For example, Clusters 2–6, each composed of a mixture of speakers with hyperkinetic, ataxic, or mixed flaccid-spastic dysarthria, were well distinguished along the intelligibility dimension (D2). Examination of the speakers' severity ratings revealed that all speakers belonging to Clusters 2 and 3, represented at one end of the D2/intelligibility continuum, were diagnosed with moderate dysarthria, and at the other end of the continuum was Cluster 6, composed of four speakers with severe dysarthria. Further evidence of disease process on listeners' judgments of similarity can be found for Clusters 4 and 5. Recall, Clusters 4 and 5 blurred along the first two dimensions but were differentiated by the third/vocal quality dimension, and the results of the post hoc analysis discussed in the Results section suggested that Cluster 5 speakers had more abnormal vocal quality than Cluster 4 speakers. Cluster 5 was composed of one speaker with ataxic dysarthria and four speakers with mixed dysarthria (secondary to amyotrophic lateral sclerosis), and Cluster 4 was composed of a single speaker with ataxic dysarthria and two speakers with hyperkinetic dysarthria. Given that strained-strangled vocal quality is a hallmark of mixed flaccid-spastic dysarthria and that these speakers were recruited because of the presence of such characteristics, it follows that vocal quality abnormalities would be greater for Cluster 5 than for Cluster 4.

The results of the present analysis are consistent with a taxonomical approach to dysarthria diagnosis, which has been offered as an alternative to classification (Weismer & Kim, 2010). Weismer and Kim (2010) proposed a taxonomical approach to studying dysarthria subtypes, in which the goal is to identify a core set of deficits (i.e., similarities) common to most, if not all, speakers with dysarthria. With respect to the present report, identification of perceptual similarities among dysarthric speech would facilitate (a) the detection of differences that reliably distinguish different types of motor speech disorders irrespective of damaged component of motor control and (b) systematic investigation of the perceptual challenges associated with the defining features of dysarthria. Indeed, the acoustic and perceptual dimensions underlying similarity in this cohort of speakers—speaking rate, intelligibility, and vocal quality—are speech features that generally unite speakers with dysarthria. Thus, the present investigation represents the first phase of research that explores the use of a taxonomical approach to understanding and defining dysarthria. The results of the cluster analysis, which identified six clusters of similar-sounding speakers, were experimenter defined. This solution was selected largely because it reflected the mean number of speaker groups identified by the listeners. However, before all six clusters were united into a single group, Clusters 2 and 3 merged, as did Clusters 4, 5, and 6, forming three discrete groups (see Figure 3). Uncovering the acoustic and perceptual features that unite a more parsimonious clustering of speakers is the goal in developing a taxonomical approach. Thus, as this line of research advances, realization of this goal will become requisite.

## Conclusion

Results of the present investigation reveal (a) feasibility of a free-classification approach for studying perceptual similarity in dysarthria, (b) correspondence between acoustic and perceptual metrics to clusters of similar-sounding speakers, and (c) impressions of perceptual similarity transcended dysarthria subtype. Together, these findings support future investigation of the link between perceptual similarities and the resulting communication disorders and targets for interventions.

## Acknowledgments

This research was supported by National Institute on Deafness and Other Communication Disorders Grants R01 DC006859 and F31 DC10093. We gratefully acknowledge Rene Utianski, Dena Berg, Angela Davis, and Cindi Hensley for their contributions to this research.

## References

- Borrie, S. A., McAuliffe, M. J., Liss, J. M., Kirk, C., O'Beirne, G. A., & Anderson, T. (2012). Familiarisation conditions and the mechanisms that underlie improved recognition of dysarthric speech. *Language and Cognitive Processes, 27*, 1039–1055.
- Bunton, K., Kent, R. D., Duffy, J. R., Rosenbek, J. C., & Kent, J. F. (2007). Listener agreement for auditory-perceptual ratings of dysarthria. *Journal of Speech, Language, and Hearing Research, 50*, 1481–1495. doi:10.1044/1092-4388(2007/102)
- Clopper, C. G. (2008). Auditory free classification: Methods and analysis. *Behavior Research Methods, 40*, 575–581. doi:10.3758/BRM.40.2.575
- Clopper, C., & Bradlow, A. (2009). Perception of dialect variation in noise: Intelligibility and classification. *Language and Speech, 51*, 175–198.
- Clopper, C. G., & Pisoni, D. B. (2007). Free classification of regional dialects of American English. *Journal of Phonetics, 35*, 421–438. doi:10.1016/j.wocn.2006.06.001
- Corter, J. E. (1998). An efficient metric combinatorial algorithm for fitting additive trees. *Multivariate Behavioral Research, 33*, 249–272. doi:10.1207/s15327906mbr3302\_3
- Darley, F. L., Aronson, A. E., & Brown, J. R. (1969a). Clusters of deviant speech dimensions in the dysarthrias. *Journal of Speech and Hearing Research, 12*, 462–496.
- Darley, F. L., Aronson, A. E., & Brown, J. R. (1969b). Differential diagnostic patterns of dysarthria. *Journal of Speech and Hearing Research, 12*, 246–269.
- Darley, F. L., Aronson, A. E., & Brown, J. R. (1975). *Motor speech disorders*. Philadelphia, PA: Saunders.
- Duffy, J. R. (2005). *Motor speech disorders: Substrates, differential diagnosis, and management* (2nd ed.). St. Louis, MO: Elsevier.
- Guastavino, C. (2007). Categorization of environmental sounds. *Canadian Journal of Experimental Psychology, 61*, 54–63. doi:10.1037/cjep2007006
- Gygi, B., Kidd, G. R., & Watson, C. S. (2007). Similarity and categorization of environmental sounds. *Perception & Psychophysics, 69*, 839–855. doi:10.3758/BF03193921
- Hillenbrand, J. M., & Gayvert, R. T. (2005). Open source software for experiment design and control. *Journal of Speech, Language, and Hearing Research, 48*, 45–60.
- Houtgast, T., & Steeneken, J. M. (1985). A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. *The Journal of the Acoustical Society of America, 77*, 1069–1077.
- Hughes, G. W., & Halle, M. (1956). Spectral properties of fricative consonants. *The Journal of the Acoustical Society of America, 28*, 303–310. doi:10.1121/1.1908271
- 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.
- Jongman, A., Waylung, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America, 108*, 1252–1263.
- Kreiman, J., & Gerratt, B. (1988). Validity of rating scale measures for voice quality. *The Journal of the Acoustical Society of America, 104*, 1598–1608.
- Lansford, K. L., & Liss, J. M. (2014a). Vowel acoustics in dysarthria: Mapping to perception. *Journal of Speech, Language, and Hearing Research, 57*, 68–80. doi:10.1044/1092-4388(2013/12-0263)
- Lansford, K. L., & Liss, J. M. (2014b). Vowel acoustics in dysarthria: Speech disorder diagnosis and classification. *Journal of Speech, Language, and Hearing Research, 57*, 57–67. doi:10.1044/1092-4388(2013/12-0262)
- Leino, T. (2009). Long-term average spectrum in screening of voice quality in speech: Untrained male university students. *Journal of Voice, 23*, 671–676.

- Liss, J. M.** (2007). Perception of dysarthric speech. In G. Weismer (Ed.), *Motor speech disorders: Essays for Ray Kent* (pp. 187–219). San Diego, CA: Plural.
- Liss, J. M., Legendre, S., & Lotto, A. J.** (2010). Discriminating dysarthria type from envelope modulation spectra. *Journal of Speech, Language, and Hearing Research, 53*, 1246–1255.
- Liss, J. M., Spitzer, S., Caviness, J. N., & Adler, C.** (2002). The effects of familiarization on intelligibility and lexical segmentation in hypokinetic and ataxic dysarthria. *The Journal of the Acoustical Society of America, 112*, 3022–3030.
- Liss, J. M., Utianski, R. L., & Lansford, K. L.** (2013). Cross-linguistic application of English-centric rhythm descriptors in motor speech disorders. *Folia Phoniatrica et Logopaedica, 65*, 3–19. doi:10.1159/000350030
- Liss, J. M., White, L., Mattys, S. L., Lansford, K. L., Lotto, A. J., Spitzer, S. M., & Caviness, J. N.** (2009). Quantifying speech rhythm abnormalities in the dysarthrias. *Journal of Speech, Language, and Hearing Research, 52*, 134–1352. doi:10.1044/1092-4388(2009/08-0208)
- Maniwa, K., Jongman, A., & Wade, T.** (2009). Acoustic characteristics of clearly spoken English fricatives. *The Journal of the Acoustical Society of America, 125*, 3962–3973.
- 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.
- Milenkovic, P. H.** (2004). TF32 [Computer software]. Madison: University of Wisconsin—Madison, Department of Electrical and Computer Engineering.
- Monson, B. B., Lotto, A. J., & Story, B. H.** (2012). Analysis of high-frequency energy in long-term average spectra of singing, speech, and voiceless fricatives. *The Journal of the Acoustical Society of America, 132*, 1754–1764.
- Shoji, K., Regembogen, E., Yu, J. D., & Blaugrund, S. M.** (1992). High-frequency power ratio of breathy voice. *Laryngoscope, 102*, 267–271.
- Tilsen, S., & Johnson, K.** (2008). Low-frequency Fourier analysis of speech rhythm. *The Journal of the Acoustical Society of America, 124*, EL34–EL39.
- Tjaden, K., Sussman, J., Liu, G., & Wilding, G.** (2010). Long-term average spectral measures of dysarthria and their relationship to perceived severity. *Journal of Medical Speech-Language Pathology, 18*, 125–132.
- Utianski, R. L., Liss, J. M., Lotto, A. J., & Lansford, K. L.** (2012, February). *The use of long-term average spectra (LTAS) in discriminating dysarthria types*. Paper presented at the Conference on Motor Speech, Santa Rosa, CA.
- Weismer, G., & Kim, Y.** (2010). Classification and taxonomy of motor speech disorders: What are the issues? In B. Maassen & P. van Lieshout (Eds.), *Speech motor control: New developments in basic and applied research* (pp. 229–241). Cambridge, England: Oxford University Press.

## Appendix

Pooled speaker similarity matrix.

| Speaker | 1  | 2  | 3  | 4  | 5  | 6  | 7 | 8  | 9  | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 |   |
|---------|----|----|----|----|----|----|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|---|
| AF1     | 0  | 0  | 14 | 8  | 0  | 2  | 4 | 6  | 0  | 6  | 9  | 1  | 1  | 1  | 4  | 2  | 3  | 3  | 3  | 1  | 2  | 0  | 6  | 10 | 3  | 8  | 0  | 1  | 1  | 2  | 0  | 0  | 0  |   |
| AF2     | 0  | 0  | 2  | 4  | 8  | 7  | 6 | 3  | 5  | 2  | 4  | 3  | 4  | 5  | 4  | 5  | 3  | 3  | 7  | 4  | 5  | 11 | 0  | 1  | 7  | 1  | 1  | 0  | 0  | 1  | 1  | 0  | 0  |   |
| AF7     | 14 | 2  | 0  | 6  | 1  | 3  | 5 | 8  | 2  | 8  | 7  | 0  | 1  | 0  | 3  | 4  | 1  | 1  | 2  | 1  | 7  | 0  | 9  | 12 | 1  | 4  | 0  | 1  | 1  | 2  | 0  | 0  | 0  |   |
| AF8     | 8  | 4  | 6  | 0  | 0  | 7  | 6 | 7  | 10 | 6  | 12 | 2  | 1  | 5  | 9  | 8  | 6  | 7  | 9  | 1  | 8  | 3  | 2  | 11 | 4  | 4  | 0  | 1  | 0  | 2  | 0  | 0  | 0  |   |
| AF9     | 0  | 8  | 1  | 0  | 0  | 3  | 1 | 2  | 2  | 2  | 0  | 9  | 11 | 5  | 3  | 3  | 3  | 1  | 3  | 10 | 2  | 7  | 2  | 1  | 2  | 2  | 1  | 2  | 2  | 1  | 0  | 2  | 2  |   |
| AM1     | 2  | 7  | 3  | 7  | 3  | 0  | 8 | 1  | 7  | 4  | 3  | 6  | 3  | 13 | 8  | 5  | 11 | 10 | 5  | 2  | 6  | 3  | 1  | 0  | 4  | 3  | 0  | 1  | 0  | 0  | 0  | 1  | 1  |   |
| AM3     | 4  | 6  | 5  | 6  | 1  | 8  | 0 | 3  | 4  | 2  | 5  | 3  | 1  | 5  | 4  | 3  | 3  | 6  | 3  | 1  | 2  | 3  | 2  | 2  | 5  | 0  | 2  | 2  | 2  | 3  | 1  | 1  | 2  |   |
| AM4     | 6  | 3  | 8  | 7  | 2  | 1  | 3 | 0  | 6  | 12 | 6  | 1  | 1  | 2  | 4  | 3  | 1  | 2  | 3  | 1  | 5  | 4  | 5  | 13 | 4  | 4  | 1  | 0  | 1  | 0  | 2  | 1  | 1  |   |
| AM5     | 0  | 5  | 2  | 10 | 2  | 7  | 4 | 6  | 0  | 8  | 7  | 1  | 4  | 2  | 4  | 15 | 6  | 4  | 8  | 2  | 15 | 6  | 0  | 7  | 4  | 0  | 1  | 0  | 0  | 2  | 0  | 0  | 0  |   |
| AM6     | 6  | 2  | 8  | 6  | 2  | 4  | 2 | 12 | 8  | 0  | 6  | 0  | 2  | 1  | 7  | 6  | 2  | 3  | 5  | 1  | 12 | 3  | 4  | 9  | 4  | 3  | 0  | 1  | 1  | 2  | 0  | 0  | 0  |   |
| AM8     | 9  | 4  | 7  | 12 | 0  | 3  | 5 | 6  | 7  | 6  | 0  | 1  | 0  | 3  | 5  | 6  | 4  | 4  | 11 | 1  | 6  | 4  | 3  | 10 | 6  | 5  | 0  | 1  | 1  | 3  | 0  | 0  | 0  |   |
| ALSF2   | 1  | 3  | 0  | 2  | 9  | 6  | 3 | 1  | 1  | 0  | 1  | 0  | 15 | 10 | 1  | 1  | 6  | 5  | 1  | 19 | 0  | 1  | 0  | 0  | 2  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |   |
| ALSF5   | 1  | 4  | 1  | 1  | 11 | 3  | 1 | 1  | 4  | 2  | 0  | 15 | 0  | 5  | 4  | 3  | 6  | 3  | 2  | 19 | 4  | 3  | 0  | 1  | 2  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 0  |   |
| ALSF6   | 1  | 5  | 0  | 5  | 5  | 13 | 5 | 2  | 2  | 1  | 3  | 10 | 5  | 0  | 7  | 1  | 10 | 11 | 4  | 6  | 0  | 2  | 0  | 0  | 2  | 3  | 0  | 0  | 0  | 0  | 0  | 0  | 0  |   |
| ALSF7   | 4  | 4  | 3  | 9  | 3  | 8  | 4 | 4  | 4  | 7  | 5  | 1  | 4  | 7  | 0  | 7  | 10 | 11 | 9  | 3  | 7  | 2  | 0  | 2  | 5  | 3  | 0  | 1  | 0  | 1  | 1  | 0  | 0  |   |
| ALSF8   | 2  | 5  | 4  | 8  | 3  | 5  | 3 | 3  | 15 | 6  | 6  | 1  | 3  | 1  | 7  | 0  | 5  | 2  | 7  | 3  | 14 | 4  | 1  | 8  | 4  | 0  | 1  | 0  | 0  | 1  | 0  | 0  | 0  |   |
| ALSF9   | 3  | 3  | 1  | 6  | 3  | 11 | 3 | 1  | 6  | 2  | 4  | 6  | 6  | 10 | 10 | 5  | 0  | 10 | 5  | 6  | 3  | 5  | 0  | 1  | 3  | 2  | 0  | 0  | 0  | 1  | 0  | 0  | 0  |   |
| ALSM1   | 3  | 3  | 1  | 7  | 1  | 10 | 6 | 2  | 4  | 3  | 4  | 5  | 3  | 11 | 11 | 2  | 10 | 0  | 6  | 2  | 4  | 3  | 0  | 3  | 4  | 5  | 0  | 1  | 0  | 1  | 0  | 0  | 0  |   |
| ALSM4   | 3  | 7  | 2  | 9  | 3  | 5  | 3 | 3  | 8  | 5  | 11 | 1  | 2  | 4  | 9  | 7  | 5  | 6  | 0  | 1  | 9  | 4  | 5  | 4  | 4  | 5  | 1  | 1  | 1  | 1  | 1  | 2  | 2  |   |
| ALSM7   | 1  | 4  | 1  | 1  | 10 | 2  | 1 | 1  | 2  | 1  | 1  | 19 | 19 | 6  | 3  | 3  | 6  | 2  | 1  | 0  | 1  | 2  | 0  | 1  | 2  | 0  | 1  | 2  | 0  | 0  | 0  | 0  | 0  | 0 |
| ALSM8   | 2  | 5  | 7  | 8  | 2  | 6  | 2 | 5  | 15 | 12 | 6  | 0  | 4  | 0  | 7  | 14 | 3  | 4  | 9  | 1  | 0  | 5  | 1  | 5  | 6  | 0  | 0  | 1  | 0  | 2  | 1  | 0  | 0  | 0 |
| HDM10   | 0  | 11 | 0  | 3  | 7  | 3  | 3 | 4  | 6  | 3  | 4  | 1  | 3  | 2  | 2  | 4  | 5  | 3  | 4  | 2  | 5  | 0  | 1  | 2  | 11 | 0  | 0  | 1  | 2  | 1  | 0  | 0  | 0  |   |
| HDM11   | 6  | 0  | 9  | 2  | 2  | 1  | 2 | 5  | 0  | 4  | 3  | 0  | 0  | 0  | 1  | 0  | 0  | 5  | 0  | 1  | 1  | 0  | 7  | 2  | 9  | 5  | 3  | 3  | 3  | 3  | 6  | 6  |    |   |
| HDM12   | 10 | 1  | 12 | 11 | 1  | 0  | 2 | 13 | 7  | 9  | 10 | 0  | 1  | 0  | 2  | 8  | 1  | 3  | 4  | 1  | 5  | 2  | 7  | 0  | 2  | 5  | 1  | 2  | 1  | 1  | 1  | 1  | 0  |   |
| HDM8    | 3  | 7  | 1  | 4  | 2  | 4  | 5 | 4  | 4  | 4  | 6  | 2  | 2  | 2  | 5  | 4  | 3  | 4  | 4  | 2  | 6  | 11 | 2  | 2  | 0  | 0  | 0  | 1  | 1  | 3  | 2  | 0  | 0  |   |
| PDF5    | 8  | 1  | 4  | 4  | 2  | 3  | 0 | 4  | 0  | 3  | 5  | 1  | 0  | 3  | 3  | 0  | 2  | 5  | 5  | 0  | 0  | 0  | 9  | 5  | 0  | 0  | 6  | 6  | 6  | 5  | 5  | 7  | 7  |   |
| PDF7    | 0  | 1  | 0  | 0  | 1  | 0  | 2 | 1  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 0  | 1  | 0  | 0  | 0  | 5  | 1  | 0  | 6  | 0  | 19 | 16 | 13 | 20 | 22 | 19 |   |
| PDM10   | 1  | 0  | 1  | 1  | 2  | 1  | 2 | 0  | 0  | 1  | 1  | 0  | 0  | 0  | 1  | 0  | 0  | 1  | 1  | 0  | 1  | 0  | 3  | 2  | 1  | 6  | 19 | 0  | 17 | 16 | 18 | 19 | 18 |   |
| PDM12   | 1  | 0  | 1  | 0  | 2  | 0  | 2 | 1  | 0  | 1  | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 0  | 1  | 0  | 1  | 3  | 1  | 1  | 6  | 16 | 17 | 0  | 16 | 16 | 16 |   |
| PDM13   | 2  | 1  | 2  | 2  | 1  | 0  | 3 | 0  | 2  | 2  | 3  | 0  | 0  | 0  | 1  | 1  | 1  | 1  | 1  | 1  | 0  | 2  | 2  | 3  | 1  | 3  | 5  | 13 | 16 | 16 | 0  | 12 | 13 |   |
| PDM15   | 0  | 1  | 0  | 0  | 0  | 0  | 1 | 2  | 0  | 0  | 0  | 0  | 1  | 0  | 1  | 0  | 0  | 0  | 1  | 0  | 1  | 1  | 3  | 1  | 2  | 5  | 20 | 18 | 16 | 12 | 0  | 21 | 19 |   |
| PDM8    | 0  | 0  | 0  | 0  | 2  | 1  | 1 | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 2  | 0  | 0  | 0  | 6  | 1  | 0  | 7  | 22 | 19 | 16 | 13 | 21 | 0  | 20 |   |
| PDM9    | 0  | 0  | 0  | 0  | 2  | 1  | 2 | 1  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 2  | 0  | 0  | 0  | 6  | 0  | 0  | 7  | 19 | 18 | 18 | 15 | 19 | 20 | 0  |   |

Note. The frequency with which each pair of speakers was judged to be similar by listeners is shown in each cell. This matrix was used to conduct the cluster and multidimensional scaling analyses. A = ataxia; F = female; M = male; ALS = amyotrophic lateral sclerosis; HD = Huntington's disease; PD = Parkinson's disease.