Research Article

Vowel Acoustics in Dysarthria: Mapping to Perception

Kaitlin L. Lansford[a] and Julie M. Liss[a]

Purpose: The aim of the present report was to explore whether vowel metrics, demonstrated to distinguish dysarthric and healthy speech in a companion article (Lansford & Liss, 2014), are able to predict human perceptual performance.
Method: Vowel metrics derived from vowels embedded in phrases produced by 45 speakers with dysarthria were compared with orthographic transcriptions of these phrases collected from 120 healthy listeners. First, correlation and stepwise multiple regressions were conducted to identify acoustic metrics that had predictive value for perceptual measures. Next, discriminant function analysis misclassifications were compared with listeners' misperceptions to examine more directly the perceptual consequences of degraded vowel acoustics.
Results: Several moderate correlative relationships were found between acoustic metrics and perceptual measures,

with predictive models accounting for 18%–75% of the variance in measures of intelligibility and vowel accuracy. Results of the second analysis showed that listeners better identified acoustically distinctive vowel tokens. In addition, the level of agreement between misclassified-to-misperceived vowel tokens supports some specificity of degraded acoustic profiles on the resulting percept.
Conclusion: Results provide evidence that degraded vowel acoustics have some effect on human perceptual performance, even in the presence of extravowel variables that naturally exert influence in phrase perception.

Key Words: acoustics, dysarthria, speech perception, speech production

I n a companion article, we report (Lansford & Liss, 2014) that various vowel metrics designed to capture vowel reduction were capable of distinguishing between vowels produced by healthy and dysarthric speakers but not among dysarthria subtypes. This supports the general notion that dysarthric vowel production is statistically distinguishable from healthy speech and that some vowel metrics are more sensitive than others to these differences. The next question, and the focus of the present study, is whether these acoustic differences have predictable consequences for perception. That is, can the degraded acoustics predict perceptual performance in healthy, young listeners? Although this may seem intuitively likely, the current body of research on vowel production and perception in dysarthria has not converged on a definitive relationship between vowel acoustics and the resulting perception. Indeed, Weismer, Jeng, Laures, Kent, and Kent (2001) advanced the possibility that the degradations seen in vowels produced by individuals

with dysarthria may be more an overall consequence of motor speech disorder severity than an integral contributor to the intelligibility deficit.

A review of the existing literature indicates that investigations have focused on identifying dynamic and/or static acoustic vowel metrics that correlate with perceptual performance, namely, measures of intelligibility. For example, relationships between intelligibility and dynamic metrics of vowel formant pattern instability and reduced F2 slopes have been revealed in the literature (e.g., R. D. Kent, Weismer, Kent, & Rosenbeck, 1989; Y.-J. Kim, Weismer, Kent, & Duffy, 2009; Weismer et al., 2001; Weismer & Martin, 1992). Because such metrics that capture formant movement (specifically F2 movement) during vowel production have contributed greatly to current theories of vowel perception (Nearey, 1989; Strange, 1989a, 1989b), studying the effects of disordered formant movement on intelligibility in dysarthria is well motivated. Indeed, the relationship between movement of the second formant and intelligibility has been studied in the dysarthrias. For example, Weismer et al. (2001) found significant correlations between F2 slopes of /aɪ/, /ɔ/, and /ju/ and scaled sentence intelligibility estimates (r = .794, −.967, and .942, respectively) in patients with dysarthria secondary to amyotrophic lateral sclerosis (ALS) and Parkinson's disease (PD). Y.-J. Kim et al. (2009) reported a less robust,

[a]Arizona State University, Tempe

Correspondence to Kaitlin L. Lansford, who is now at Florida State University: klansford@fsu.edu

albeit significant, predictive relationship between F2 slopes and scaled estimates of intelligibility in speakers with dysarthria secondary to PD and stroke (r^2 from 13.9% to 14.3%).

Dysarthric vowel production is also characterized by centralization of formant frequencies and reduction in static vowel space area (K. Kent, Weismer, Kent, Vorperian, & Duffy, 1999). Because of the perceptual consequences of vowel space reduction described in studies of clear versus conversational speech (Bradlow & Bent, 2002; Payton, Uchanski, & Braida, 1994; Picheny, Durlach, & Braida, 1985; Uchanski, Choi, Braida, Reed, & Durlach, 1996), there is reason to believe that vowel space reduction in dysarthric speech should negatively impact intelligibility. However, investigations relating acoustic metrics approximating vowel space area (VSA; triangular and quadrilateral) to overall intelligibility in dysarthria have yielded mixed results. For example, Turner, Tjaden, and Weismer (1995) found that VSA derived from the vowel quadrilateral accounted for 46% of the variance in scaled intelligibility ratings in patients with ALS. The same was reported in an investigation of speakers with dysarthria secondary to either PD or ALS (Weismer et al., 2001). However, the authors concluded that the relationship appeared to be carried by the ALS speakers, as there was no distinguishable difference between PD and control vowel space areas. In children with dysarthria secondary to cerebral palsy (CP), VSA accounted for 64% of the variance in single word intelligibility scores (Higgins & Hodge, 2002). Similarly, a significant relationship between VSA and single word intelligibility scores in Mandarin speakers with CP ($r = .684$) has been reported (Liu, Tsao, & Kuhl, 2005). Conversely, Tjaden and Wilding (2004) demonstrated less impressive predictive power of VSA metrics in women with dysarthria secondary to multiple sclerosis (MS) or PD, as approximately 6%–8% of the variance in scaled intelligibility ratings were accounted for by a subset of acoustic metrics that included VSA and F2 slope of /aɪ/. In the male speakers, a different subset of metrics, which included F2 slope of /aɪ/ and /eɪ/ but not VSA, predicted 12%–21% of the variance in intelligibility scores (Tjaden & Wilding, 2004). In another investigation, VSA accounted for only 12% of the variance in scaled severity scores in speakers diagnosed with PD (McRae, Tjaden, & Schoonings, 2002). Thus, the extent to which VSA measures predicted intelligibility in these investigations would appear to be dependent on a number of factors, including gender of the speaker, nature of the underlying disease, and type of stimuli used in the investigation.

H. Kim, Hasegawa-Johnson, and Perlman (2011), motivated by such varied VSA findings, evaluated the ability of alternate measures of vowel working space including lax vowel space area, mean Euclidean distance between the vowels, F1 and F2 variability, and spectral overlap degree among the vowels to predict intelligibility scores obtained from speakers with dysarthria secondary to CP. Significant predictive relationships were revealed for VSA ($R^2 = .69$), mean distance between the vowels ($R^2 = .69$), and variability of F1 ($R^2 = .74$). However, a novel metric referred to as overlap degree demonstrated the strongest predictive relationship with intelligibility ($R^2 = .96$). Spectral overlap degree

was derived from the results of a per speaker classification analysis of vowel tokens into their respective categories. Vowel misclassification rates were interpreted to reflect the degree of spectral–temporal overlap among the vowels. The authors concluded that the degree of spectral–temporal overlap, captured by this metric, might offer more to the study of intelligibility deficits in dysarthria than traditional VSA metrics.

There have been many explorations of vowel acoustics and intelligibility in the dysarthrias, yet we still lack the information required to build an explanatory model. Overall severity of the speech disorder, particularly when mild, may contribute to a weak relationship between degraded vowel acoustics and intelligibility. Another possible reason for this weak relationship is that the dependent measures of intelligibility (e.g., scaled intelligibility estimates and words correct) may be insufficiently sensitive to deduce explanations. Perhaps, then, dependent measures that capture vowel accuracy more appropriately approximate the relationship between degraded vowel acoustics and vowel perception. Although this relationship has not been addressed directly in English speakers with dysarthria, results from related studies have not provided strong support for this argument (e.g., Bunton & Weismer, 2001; Liu et al., 2005; Whitehill, Ciocca, Chan, & Samman, 2006). For example, Liu and colleagues (2005) explored the relationship between VSA and vowel accuracy in young adult male Mandarin speakers with CP and found a significant correlation ($r = .63$). Similarly, Whitehill et al. (2006) demonstrated a significant relationship between VSA and vowel accuracy ($r = .32$) in Cantonese speakers with partial glossectomy. Bunton and Weismer (2001) evaluated the acoustic differences between correctly identified and misperceived (tongue-height errors) vowel tokens and found that they were not reliably distinguishable.

In a reanalysis of the Hillenbrand database, Neel (2008) focused her inquiry on the relationship between vowel acoustics and the perceptual identification accuracy of vowel tokens produced by healthy adult speakers. A host of derived vowel space measurements were regressed against the perceptual identification scores, and subsets of these metrics were found to account for only 9%–12% of the variance in the perceptual scores. The results of this analysis were influenced by a ceiling effect in the perceptual identification scores, as healthy control speakers were used. In a subsequent analysis, however, well-identified vowel tokens were found to be more distinctive in F1 and F2, duration, and formant movement over time as compared with poorly identified vowel tokens. Neel concluded that measurements of vowel distinctiveness among neighboring vowels, rather than VSA, might prove more useful in predicting vowel accuracy. This supports the notion that understanding the relationship between vowel acoustics and the corresponding percept is key to defining the contribution of vowel degradation to overall measures of intelligibility.

In the present report, we aimed to explore the relationship between degraded vowel acoustics and perceptual outcomes in a large and diverse cohort of dysarthric speakers producing phrase-level material by using a wide variety of

acoustic and perceptual measures. First, and in line with previous work, the correlative and predictive relationships between a number of established and novel vowel metrics and perceptual accuracy scores, including percentage of words correct and vowel accuracy, were evaluated (Analysis 1). Analysis 2 was designed to examine how the acoustics of a vowel influence its perception by comparing patterns of perceptual performance with the statistical classification of vowel token based strictly on acoustic data (discriminant function analysis).

Analysis 1

Study Overview

This investigation assessed the relationships between established and novel vowel metrics demonstrated to differentiate vowels produced by speakers with and without dysarthria (Lansford & Liss, 2014) and perceptual accuracy scores obtained from a transcription task, including intelligibility and vowel accuracy, in a heterogeneous cohort of dysarthric speakers producing read phrases. These relationships were first studied using correlation analysis, and then stepwise multiple regression analysis was used to generate predictive models of intelligibility and vowel accuracy.

Method

Speakers

Speech samples from 45 speakers with dysarthria, collected as part of a larger study and described in detail in the previous report (Lansford & Liss, 2014), were used in the present analysis. Briefly, the speakers were diagnosed with one of four types of dysarthria: ataxic dysarthria secondary to various neurodegenerative diseases (ataxic; $n = 12$), hypokinetic dysarthria secondary to idiopathic Parkinson's disease (PD; $n = 12$), hyperkinetic dysarthria secondary to Huntington's disease (HD; $n = 10$), or mixed flaccid-spastic dysarthria secondary to amyotrophic lateral sclerosis (ALS; $n = 11$). Speaker age, gender, and severity of impairment are provided in Table 1. Two trained speech-language pathologists affiliated with the Motor Speech Disorder (MSD) lab at Arizona State University (including the second author) independently rated severity of each speaker's impairment from a production of "The Grandfather Passage." Perceptual ratings of mild, moderate, and severe were corroborated by the intelligibility data (percent words correct) collected for this report. The disordered speakers were selected from the pool of speech samples on the basis of the presence of the cardinal features associated with their corresponding dysarthria.

Stimuli

The stimuli used in this investigation included 36 semantically anomalous, syntactically plausible phrases that alternated in phrasal stress pattern (18 produced by each speaker). These phrases were described in detail in the previous report (see Appendix A in Lansford & Liss, 2014, for a full list of the stimulus items). Briefly, the spectral and temporal characteristics of vowel tokens embedded in these phrases were measured to derive the vowel metrics included in

this investigation. In addition, these stimuli were used to collect the perceptual data used in this analysis.

Vowel Measurements and Derived Metrics

As detailed in Lansford and Liss (2014), the first and second formants were measured in hertz at each vowel's onset (20% of vowel duration), midpoint (50% of vowel duration), and offset (80% of vowel duration). In addition, total vowel duration (ms) was measured. These measures were then used to derive vowel metrics that capture mean working vowel space and formant movement over time. These metrics are presented in Table 2 and include traditional vowel space area metrics (triangular, quadrilateral, and lax), formant centralization ratio, mean dispersion of all vowel pairs (mean dispersion), mean dispersion of the front vowels (front dispersion), mean dispersion of the back vowels (back dispersion), mean dispersion of the corner vowels to the center vowel /ʌ/ (corner dispersion), mean dispersion of all vowels to the speaker's average F1 and F2 vowels across all vowels (global dispersion), mean F2 slope across all vowels, and mean F2 slope of the most dynamic vowels (dynamic F2 slope). They are described in greater detail in Lansford and Liss (2014). In addition to these metrics, we computed spectral overlap (H. Kim et al., 2011). This metric reflects the results of a per speaker classification analysis of vowel tokens into their vowel categories, and vowel misclassification rates were interpreted to represent the degree of spectral–temporal overlap among the vowels. Thus, for each speaker, the vowel tokens were classified as one of the 10 vowels via discriminant function analysis (DFA) using spectral and temporal per token measurements (see Table 2). The misclassification rate per speaker was interpreted to reflect the degree of spectral overlap of the vowels, as per H. Kim et al. (2011).

It is important to note that although some of these metrics have been used in previous investigations to explore their relationships with intelligibility decrements in dysarthria (e.g., VSA in Turner et al. [1995] and Weismer et al. [2001]; F2 slope in Y.-J. Kim et al. [2009] and Weismer et al. [2001]), the majority have not (e.g., dispersion metrics and formant centralization ratio). To our knowledge, none of these metrics has been used to explore dysarthric vowel accuracy in English.

Perceptual Task

Listeners. Listeners were 120 undergraduate and graduate students (115 female) recruited from the Arizona State University population. Listeners' ages ranged from 18 to 54 with a mean age of 24. They had no history of language or hearing disorders and were native speakers of English per self-report. All listeners received either partial course credit or monetary remuneration of $5 for their participation.

Materials. To permit investigation of listeners' perceptions of each vowel token per speaker and to minimize speaker-specific learning effects while simultaneously maximizing the limited stimuli, we created six listening blocks per dysarthria group. In each listening block, listeners heard three different phrases produced by the 12 speakers. The speaker–phrase composition of each listening block was

Table 1. Dysarthric speaker demographic information per stimulus set.

Speaker	Sex	Age	Medical etiology	Severity of speech disorder
			Set 1	
ALSF2	F	75	Amyotrophic lateral sclerosis	Severe
ALSF8	F	63	Amyotrophic lateral sclerosis	Moderate
ALSM1	M	56	Amyotrophic lateral sclerosis	Moderate
ALSM5	M	50	Amyotrophic lateral sclerosis	Mild
ALSM7	M	60	Amyotrophic lateral sclerosis	Severe
AF2	F	57	Multiple sclerosis/ataxia	Severe
AF6	F	57	Friedrich's ataxia	Moderate
AF7	F	48	Cerebellar ataxia	Moderate
AM1	M	73	Cerebellar ataxia	Severe
AM5	M	84	Cerebellar ataxia	Moderate
AM6	M	46	Cerebellar ataxia	Moderate
HDF5	F	41	Huntington's disease	Moderate
HDF6	F	57	Huntington's disease	Severe
HDM3	M	80	Huntington's disease	Moderate
HDM10	M	50	Huntington's disease	Severe
HDM12	M	76	Huntington's disease	Moderate
PDF1	F	64	Parkinson's disease	Mild
PDF7	F	58	Parkinson's disease	Moderate
PDF9	F	71	Parkinson's disease	Mild
PDM8	M	77	Parkinson's disease	Moderate
PDM9	M	76	Parkinson's disease	Moderate
PDM15	M	57	Parkinson's disease	Moderate
			Set 2	
ALSF5	F	73	Amyotrophic lateral sclerosis	Severe
ALSF7	F	54	Amyotrophic lateral sclerosis	Moderate
ALSF9	F	86	Amyotrophic lateral sclerosis	Severe
ALSM3	M	41	Amyotrophic lateral sclerosis	Mild
ALSM4	M	64	Amyotrophic lateral sclerosis	Moderate
ALSM8	M	46	Amyotrophic lateral sclerosis	Moderate
AF1	F	72	Cerebellar ataxia	Moderate
AF8	F	65	Cerebellar ataxia	Moderate
AF9	F	87	Cerebellar ataxia	Severe
AM3	M	79	Cerebellar ataxia	Moderate–severe
AM4	M	46	Cerebellar ataxia	Moderate
AM8	M	63	Cerebellar ataxia	Moderate
HDF1	F	62	Huntington's disease	Moderate
HDF3	F	37	Huntington's disease	Moderate
HDF7	F	31	Huntington's disease	Severe
HDM8	M	43	Huntington's disease	Severe
HDM11	M	56	Huntington's disease	Moderate
PDF3	F	82	Parkinson's disease	Mild
PDF5	F	54	Parkinson's disease	Moderate
PDF6	F	65	Parkinson's disease	Mild
PDM1	M	69	Parkinson's disease	Severe
PDM10	M	80	Parkinson's disease	Moderate
PDM12	M	66	Parkinson's disease	Severe

Note.   F = female; M = male.

counterbalanced such that perceptual data for each speaker's production of the 18 phrases were collected.

Procedure. Five listeners were randomly assigned to each of the six listening blocks per speaker group. Thus, the perceptual data set included 120 transcripts of the 36 phrases. All listeners were seated in front of a computer screen and keyboard and were fitted with Sennheiser HD 25 SP headphones. The task was completed in a quiet room in individual carrels, designed to minimize auditory and visual distractions. At the beginning of the experiment, the signal volume was set to a comfortable listening level by each listener and remained at the level for the duration of the task. The participants were instructed that they would hear a series of phrases produced by men and women with disordered speech. They were informed that although the phrases were composed of real English words, they would not necessarily make sense. The listeners were asked to type what they heard and were encouraged to guess if unsure. Immediately following presentation of each phrase, listeners were given the opportunity to transcribe what they heard. The phrases were presented in random order, and the task was untimed.

Transcript analysis. The transcripts collected from the 120 listeners were analyzed and scored independently by two trained members of the MSD lab for (a) number of words

Table 2. Derived vowel metrics.

Vowel metric	Description
Quadrilateral VSA	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/ are 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/).
Triangular VSA	Triangular vowel space area was constructed with the corner vowels (i, a, u). It was derived using the equation outlined by Sapir and colleagues (2010) and is expressed as ABS{[F1$_i$ × (F2$_a$ − F2$_u$) + F1$_a$ × (F2$_u$ − F2$_i$) + F1$_u$ × (F2$_i$ − F2$_a$)]/2}. ABS in this equation refers to absolute value.
Lax VSA	Lax vowel space area was constructed with the lax vowels /ɪ, ɛ, ʊ/. The equation used to derive triangular vowel space area was used to derive lax vowel space area.
FCR	This ratio, expressed as (F2$_u$ + F2$_a$ + F1$_i$ + F1$_a$)/(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, ɛ, 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 (Euclidian distance in F1 × F2 space).
Average F2 slope	The absolute values of the F2 slopes from vowel onset to offset were averaged across the entire vowel set.
Dynamic F2 slope	The absolute values of F2 slopes associated with the most dynamic vowels (æ, ʌ, ʊ) were averaged. Dynamic vowels were so designated based on the work of Neel (2008).
Spectral overlap	This metric is the vowel misclassification rate revealed by discriminant function analysis conducted for each speaker. The following formant and temporal metrics were used to classify each vowel per speaker: F1, F2, F0 at midpoint, vowel duration, and formant movement (Euclidean distance in F1 × F2 space) from vowel onset to midpoint to offset.

Note. VSA = vowel space area; FCR = formant centralization ratio.

correctly identified and (b) vowel accuracy. For the vowel accuracy score, tokens were regarded as correctly identified when the transcribed vowel matched the target, irrespective of word accuracy (e.g., "admit" transcribed as "permit," where the vowel of the strong syllable /ɪ/ was correctly transcribed). If the transcribed vowel matched the target, it was coded with a 1. Misperceived tokens were coded as zeros, and the identity of the misperceived vowel was noted for a subsequent analysis (e.g., if "meet" was transcribed as "met," vowel accuracy was coded as a 0, and the misperception was coded as an /ɛ/). Two independent transcript codes were compared for reliability. In the few cases that the coded transcripts did not match, the independent scorers convened and came to a consensus. No tokens were eliminated from the analysis because of lack of consensus.

Phrase intelligibility was calculated as a percentage of the number of words correct out of the number of words possible. Overall vowel accuracy was derived in two ways for subsequent analyses. First, token accuracy was computed by averaging the binary token identification scores across the five listeners. Thus, for each speaker, a total of 36 token accuracy scores (four tokens per nine vowels) were calculated. Next, vowel accuracy was computed by averaging the token accuracy scores for all of the vowels per speaker. The identity

of the vowel misperception also was entered into confusion matrices for subsequent analysis.

Results

Perceptual Data

Overall intelligibility and vowel identification scores obtained from the listeners of each dysarthric speaker are found in Table 3. To ensure the perceptual data obtained from this heterogeneous group of speakers, collected from two different sets of stimuli, could be analyzed together, means testing was completed. Specifically, t tests were conducted to ensure that the speakers assigned to Stimulus Sets 1 and 2 did not differ significantly on the perceptual measurements. Neither the percent total-words-correct intelligibility scores, $t(43) = -.304$, $p = .763$, nor target vowel accuracy, $t(43) = -.415$, $p = .681$, differed significantly. Intelligibility scores for Sets 1 and 2 speakers were 49% ($SD = .21$) and 50% ($SD = .20$), respectively; and mean vowel accuracy scores were 69% ($SD = .20$) and 71% ($SD = .17$) for Sets 1 and 2, respectively. Thus, the perceptual data obtained for Sets 1 and 2 were analyzed together in all subsequent analyses.

A one-way analysis of variance was conducted to evaluate the effect of dysarthria group on perceptual measures.

Table 3. Proportion of words and vowels correct per speaker.

Group and speaker	Words correct	Vowel accuracy
	Ataxic	
AF1	.59	.82
AF2	.38	.56
AF6	.72	.88
AF7	.61	.76
AF8	.68	.93
AF9	.19	.44
AM1	.26	.56
AM3	.44	.61
AM4	.64	.84
AM5	.49	.76
AM6	.47	.59
AM8	.63	.81
M (SD)	51 (.17)	.71 (.15)
	ALS	
ALSF2	.11	.28
ALSF5	.20	.43
ALSF7	.39	.61
ALSF8	.43	.68
ALSF9	.30	.53
ALSM1	.74	.85
ALSM3	.65	.81
ALSM4	.71	.87
ALSM5	.70	.89
ALSM7	.08	.24
ALSM8	.56	.70
M (SD)	.44 (.25)	.63 (.23)
	HD	
HDF1	.57	.77
HDF3	.65	.81
HDF5	.60	.83
HDF6	.19	.46
HDF7	.14	.32
HDM10	.26	.37
HDM11	.70	.83
HDM12	.67	.88
HDM3	.45	.64
HDM8	.48	.67
M (SD)	.47 (.21)	.66 (.21)
	PD	
PDF1	.74	.83
PDF3	.83	.92
PDF5	.60	.80
PDF6	.75	.91
PDF7	.64	.89
PDF9	.62	.82
PDM1	.13	.49
PDM10	.53	.83
PDM12	.36	.69
PDM15	.63	.83
PDM8	.37	.72
PDM9	.64	.90
M (SD)	.57 (.20)	.80 (.12)

Note. ALS = amyotrophic lateral sclerosis; HD = Huntington's disease; PD = Parkinson's disease.

The main effect of dysarthria group was not significant for intelligibility, $F(3, 41) = 0.825$, $p = .488$; or for vowel accuracy, $F(3, 41) = 2.137$, $p = .11$. Thus, the perceptual data obtained for all dysarthric speakers were combined to examine the acoustic correlates and predictors of intelligibility and vowel accuracy.

Correlation Analysis

To evaluate the relationships between traditional and alternative vowel space area metrics, dispersion metrics, and F2 slope vowel metrics and the perceptual outcome measures (intelligibility and vowel accuracy), we conducted Pearson correlation analysis. Several moderate relationships were revealed between the perceptual metrics and the alternate, dispersion, and F2 slope metrics (see Table 4 for detailed results). Notably, the formant centralization ratio (FCR; see Sapir, Ramig, Spielman, & Fox, 2010) and mean dispersion of the corner vowels to schwa demonstrated the strongest correlations with intelligibility. Likewise, the FCR and F2 slope of the most dynamic vowels were most strongly related to vowel accuracy.

Stepwise Multiple Regression

Because of the large set of acoustic variables, forward stepwise regression was conducted to identify a subset of vowel metrics that was predictive of intelligibility and vowel accuracy. In forward stepwise regression, one predictor variable is selected at a time to enter the regression model. At each step, the variable that yields the smallest *p* value upon entry into the model is entered (provided it is $p < .05$). Variables are entered until no other variable yields a significant *p* value. In addition, variables may be removed from the model at each step if inclusion of a new variable renders it insignificant ($p > .10$). The interdependency of the vowel metrics was investigated, and as expected, many moderate to strong correlations between the vowel space metrics were found to exist (see Appendix B in Lansford & Liss, 2014). To ensure that the regression results were not unduly inflated by the presence of multicollinearity, we computed the variance inflation factor (VIF) for the predictor variables included in each model. A VIF > 10 suggests the presence of severe multicollinearity in the model (Kutner, Nachtsheim, Neter, & Li, 2005). As reported in Table 5, none of the VIFs calculated for the variables entered into the final regression models exceeded 3.7. Thus, the presence of serious multicollinearity in the predictive models discussed below can be ruled out.

Table 4. Pearson correlations between vowel metrics and perceptual measures.

	Correlation coefficient per perceptual metric	
Vowel metric	Intelligibility	Vowel identification
Quadrilateral VSA	.401**	.412**
Triangular VSA	.203	.282
Lax VSA	−.029	.036
FCR	−.442**	−.526**
Mean dispersion	.317*	.364*
Front dispersion	.237	.308*
Back dispersion	.204	.218
Corner dispersion	.458**	.447**
Global dispersion	.335*	.392**
Average F2 slope	.401**	.461**
Dynamic F2 slope	.422**	.478**
Spectral overlap	−.379*	−.405**

*$p < .05$. ** $p < .01$.

Table 5. Predictive models of intelligibility and vowel accuracy (stepwise multiple regression results).

Variable entered	Beta	t	p	VIF
Intelligibility				
Omnibus				
Corner dispersion	0.318	2.437	.019	1.119
Mean F2 slope	0.298	2.334	.025	1.037
Spectral overlap	−0.290	−2.297	.027	1.048
Female speakers				
Dynamic slope	0.579	5.041	.000	1.105
Corner dispersion	0.378	3.320	.004	1.085
Spectral overlap	−0.319	−2.879	.010	1.032
Male speakers				
Corner dispersion	0.468	2.425	.024	1.0
Vowel accuracy				
Omnibus				
FCR	−0.834	−4.394	.000	2.991
Mean F2 slope	0.536	4.255	.000	1.256
Global dispersion	−0.455	−2.262	.029	3.366
Female speakers				
Dynamic slope	0.441	3.915	.001	1.292
Corner dispersion	0.329	3.087	.007	1.153
Spectral overlap	−0.463	−3.964	.001	1.386
Front dispersion	0.331	2.679	.016	1.553
Male speakers				
FCR	−1.169	−4.034	.001	3.655
VSA	−0.756	−2.608	.017	3.661
Mean F2 slope	0.337	2.215	.039	1.009

Note. VIF = variance inflation factor.

Because of the known spectral differences in vowels produced by male and female speakers, including greater vowel space area and mean vowel dispersion in female speakers (Hillenbrand, Getty, Clark, & Wheeler, 1995; Neel, 2008; Peterson & Barney, 1952), separate stepwise regressions were conducted for the female ($n = 22$) and male ($n = 23$) dysarthric speakers, in addition to the omnibus analyses. The acoustic data were not normalized for this experiment in order to preserve the ability of the various vowel space metrics to capture acoustic degradations.

Intelligibility

Results of the stepwise regression revealed a predictive model of intelligibility for the entire cohort of speakers that included mean dispersion of the corner vowels to the center vowel (/^/), average F2 slope, and spectral overlap degree, $R_a^2 = .33$, $F(3, 41) = 8.215$, $p < .001$. For female speakers with dysarthria, F2 slope of the most dynamic vowels, mean dispersion of the corner vowels to the center vowel, and spectral overlap degree were included in a predictive model of intelligibility, $R_a^2 = .749$, $F(3, 18) = 21.943$, $p < .001$. The predictive model for the male speakers was less impressive, $R_a^2 = .182$, $F(1, 21) = 5.881$, $p = .024$, and included mean dispersion of the corner vowels to the center (see Table 5 for summary).

Vowel Accuracy

FCR, average F2 slope, and global dispersion were selected by the stepwise regression to be included in the

predictive model of vowel accuracy, $R_a^2 = .47$, $F(3, 41) = 14.013$, $p < .001$. Thus, reduced formant centralization, greater excursion of the F2 slope, and increased mean dispersion of all vowels to each speaker's average F1 and F2 across all vowel categories were associated with increased vowel accuracy.

For female speakers with dysarthria, a subset of variables that included slope of the most dynamic vowels, mean dispersion of the corner vowels to /^/, spectral overlap, and mean dispersion of the front vowels was best predictive of vowel accuracy, $R_a^2 = .794$, $F(4, 17) = 21.18$, $p < .001$. FCR, VSA, and mean F2 slope were best predictive of vowel identification scores in male speakers, $R_a^2 = .495$, $F(3, 19) = 8.185$, $p < .001$. What is interesting, and not predicted, in this model is that vowel space area reduction in male speakers was associated with increased vowel accuracy. The relationships between the other predictor variables included in the model were in the predicted directions.

Discussion

In general, vowel space area decrements, irrespective of the measurement method, were associated with reduced intelligibility and vowel accuracy (except for male speakers). The intelligibility data in this experiment are in line with the results of previous studies conducted in dysarthria (e.g., Turner et al., 1995; Weismer et al., 2001). The regression analyses predicting vowel accuracy from subsets of acoustic variables accounted for more variance than models predicting intelligibility. Thus, the results of this analysis contribute to the literature by demonstrating vowel accuracy as an affected perceptual outcome measure of degraded vowel acoustics.

It is important to address the differential ability of the vowel metrics to account for the variance in intelligibility and vowel identification scores of male and female speakers. This finding, perhaps, is not very surprising given what we know about the spectral differences in vowels produced by male and female speakers (e.g., Hillenbrand et al., 1995; Peterson & Barney, 1952). Although both intelligibility and vowel accuracy were equivalent across the sexes, many significant acoustic differences emerged in this cohort of male and female dysarthric speakers (see Table 6). In line with previous findings (e.g., Neel, 2008), both vowel space area (triangular and quadrilateral) and mean dispersion of all vowel pairs were significantly larger in the female versus the male dysarthric speakers. Further, significant differences were found for the other metrics of dispersion and mean F2 slope for the most dynamic vowels. In addition, the standard deviations associated with these vowel metrics were generally larger for the female speakers than for the male speakers. So it is possible that the predictive models better accounted for the variance in the female speakers' scores because of the greater ranges of measures and variability in their acoustics.

The degree of variance accounted for by these acoustic metrics is notable given that the vowel metrics were derived from reduced vowel tokens embedded in connected (phrasal) speech, which introduces another source of articulatory undershoot (Lindblom, 1963). The results of this analysis provide evidence suggesting that degraded vowel acoustics

Table 6. By sex, perceptual and vowel metrics' means, standard deviations, and results of independent samples *t* tests.

Metric	n	M	SD	t(43)	p
Intelligibility				0.115	.909
F	22	0.497	0.223		
M	23	0.504	0.189		
Vowel accuracy				0.309	.759
F	22	0.694	0.205		
M	23	0.712	0.172		
Quadrilateral VSA				3.216	.002*
F	22	204,617.498	70,636.209		
M	23	146,322.300	49,589.526		
Triangular VSA				2.523	.015*
F	22	143,726.890	69,105.096		
M	23	98,046.023	51,448.739		
Lax VSA				1.321	.193
F	22	22,103.156	19,224.727		
M	23	15,365.787	14,780.968		
FCR				0.424	.674
F	22	1.180	0.111		
M	23	1.195	0.130		
Mean dispersion				2.849	.007*
F	22	356.555	59.432		
M	23	305.504	60.687		
Front dispersion				3.157	.003*
F	22	384.826	88.566		
M	23	308.169	73.987		
Back dispersion				3.231	.002*
F	22	308.250	75.584		
M	23	245.413	53.453		
Corner dispersion				3.446	.001*
F	22	476.298	93.938		
M	23	389.904	73.422		
Global dispersion				3.193	.003*
F	22	524.285	80.008		
M	23	445.679	84.890		
Average F2 slope				1.843	.072
F	22	1.720	0.606		
M	23	1.392	0.589		
Dynamic F2 slope				2.6	.013*
F	22	2.689	1.032		
M	23	1.966	0.825		
Spectral overlap				0.063	.950
F	22	0.557	0.123		
M	23	0.559	0.131		

*p < .05.

are associated with vowel perception and intelligibility; however, the conclusion that degraded vowel acoustics directly influence perception is premature. Analysis 2 was designed to shed light on this issue.

Analysis 2

Analysis 2 was designed specifically to determine whether the nature of the acoustic degradations in dysarthric vowels influences the resulting percept in a predictable way. With this goal in mind, a series of three analyses was conducted. First, we tested the hypothesis that listeners better identified vowel tokens with distinctive spectral and temporal properties than vowel tokens with less distinctive properties. Next, and to validate the findings of the first analysis, we assessed whether vowel tokens that were well identified by listeners were also

more acoustically distinctive than poorly identified vowel tokens. Finally, we conducted a point-by-point comparison of the misclassified vowel tokens (via DFA) with the listeners' specific misperception of those tokens. If the errors generated by these two very different levels of analysis (pure acoustics for the DFA versus human listener perception) were similar, this would suggest that vowel acoustics have a specific effect on perception and thus a specific contribution to the intelligibility reduction.

Method

Speakers and Stimuli

All disordered speakers and their recorded productions described in Analysis 1 were included in the present analysis.

Acoustic Metrics

The static and dynamic formant and temporal measurements associated with each vowel token (obtained in Analysis 1 and used to derive the vowel space measures) were the acoustic units of interest in this experiment. In previous reports of vowel categorization by acoustic features (see Hillenbrand et al., 1995; Peterson & Barney, 1952), inclusion of metrics of duration, fundamental frequency, and formant frequencies sampled at multiple time points significantly improved classification accuracy as compared with models that included only F1 and F2 sampled at the vowel's steady state. Thus, for each vowel token, the following formant and temporal metrics were included in the various analyses: first and second formant frequency information sampled at 20% (onset), 50% (midpoint), and 80% (offset) vowel duration; fundamental frequency (F0) sampled at 50% duration; total vowel duration; slope of the second formant from onset to offset; and formant movement (Euclidean distance) in F1 × F2 perceptual space captured in four ways: (a) from vowel onset to midpoint, (b) from midpoint to offset, (c) from onset to offset, and (d) sum of movement obtained from onset to midpoint and from midpoint to offset. The formant metrics were normalized using Labonov's method, a formant-intrinsic, vowel-extrinsic, and speaker-intrinsic procedure that has been demonstrated to eliminate interspeaker variation (Flynn, 2011).[1] The data were normalized for this experiment to improve classification accuracy of the discriminant function analysis.

Perceptual Metrics

The token accuracy scores, calculated from listener transcripts and described in Analysis 1, were used in this experiment. In addition to overall scores, correct token identifications and misperceptions for each speaker were coded and assembled into confusion matrices (see Table 7). Overall, vowel tokens were perceived with 71% accuracy.

Results

The static and dynamic formant metrics associated with each vowel token produced by the dysarthric speakers

---

[1]Flynn (2011) compared 20 methods of vowel normalization with respect to their ability to eliminate interspeaker variation. The methods were described to be vowel-, formant-, and speaker-intrinsic or extrinsic. Vowel-intrinsic methods use only the information from a single vowel token for normalization, whereas information from multiple vowel tokens, and at times from categorically different vowels, is considered by vowel-extrinsic methods. Likewise, formant-intrinsic methods use only the information contained in a given formant for normalization, but extrinsic methods use information from one or more other formants. Finally, speaker-intrinsic methods limit the normalization procedure to the information obtained for a given speaker. Speaker-extrinsic methods use information from a sample of speakers to normalize the vowel data and are rarely used. Procedures considered vowel-extrinsic and formant- and speaker-intrinsic (e.g., Bigham, 2008; Gertsman, 1968; Labonov, 1971; Watt & Fabricius, 2002) eliminated variability arising from interspeaker differences in vocal tract lengths and shapes better than many commonly used vowel-, formant-, and speaker-intrinsic methods (e.g., bark, mel, and log). Thus, normalization improved when the acoustic features of a speaker's entire vowel set are considered in the transformation of the individual vowel tokens.

were used to classify the tokens into their respective vowel categories via stepwise discriminant function analysis. The following variables were selected by the stepwise DFA to classify the 1,749 tokens in this order: F2 and F1 at midpoint, F2 slope, F1 at onset, vowel duration, F1 at offset, formant movement from onset to offset, F2 at offset and onset, sum of the formant movement from onset to midpoint and from midpoint to offset, F0, and formant movement from midpoint to offset. Classification accuracy of the vowel tokens was 65.1% (63.5% upon cross-validation; see Table 8 for the classification summary).

Subanalysis 1

An independent samples *t*-test analysis revealed that the perceptual scores associated with correctly classified tokens (*M* = .75, *SD* = .37) were significantly higher than those of misclassified tokens (*M* = .63, *SD* = .33), *t*(1658) = 6.455, *p* < .0001. Thus, correctly classified tokens (i.e., tokens with distinctive acoustic properties) were perceived with greater accuracy than tokens misclassified by DFA.

Table 7. Confusion matrix of correctly identified vowels tokens and perceptual errors.

Target vowel	Perceived vowel (%)									
	i	ɪ	e	ɛ	æ	a	o	u	ʌ	ʊ
i	74	6	3	3	1				1	
ɪ	3	66	6	7	3		1	2	2	1
e	2	4	80	3	1				1	
ɛ		6	1	75	3	1			4	
æ		3	2	15	65	2			2	
a		1		1	3	73		1	6	
o	1	1		1	1	5	69	3	6	1
u	2	3	1	3		1	5	62	2	8
ʌ		1		3	2	5	3		73	
ʊ		4					1		7	73

Note. Overall, vowel tokens were perceived with 71% accuracy. Agreements are presented along the diagonal and are shaded in gray.

Table 8. DFA classification summary of all vowel tokens.

Target vowel	Predicted vowel (%)									
	i	ɪ	e	ɛ	æ	a	o	u	ʌ	ʊ
i	88	2	6	2				2		1
ɪ	6	54	9	14	2		3	6	5	1
e	9	8	79	1	2					1
ɛ		23	1	49	16	1	1	2	8	1
æ		6	3	17	63	7	2	1	2	
a		1		1	7	69	8	2	9	3
o		1			1	4	70	6	15	3
u	8	12	1	1			15	58	1	3
ʌ		7		6	5	14	16	2	44	8
ʊ		4	3			2	5	5	1	81

Note. 65.1% of originally grouped vowels were correctly classified (63.5% upon cross-validation). Agreements are presented along the diagonal and are shaded in gray. DFA = discriminant function analysis.

Subanalysis 2

To validate the findings from the first subanalysis, vowel tokens identified with 100% accuracy (*n* = 768) and those with 60% or less accuracy (*n* = 638) by listeners were subjected to separate stepwise DFAs, in which the static and dynamic formant and temporal measurements were used to classify well-identified and poorly identified vowel tokens. The following 10 variables were selected by the stepwise DFA to classify well-identified vowel tokens: F2 and F1 at midpoint, F2 slope, vowel duration, F1 at onset, formant movement from onset to offset, F1 at offset, F2 at onset and offset, sum of the formant movement from onset to midpoint and from midpoint to offset. Well-identified vowel tokens were classified by DFA with 71.2% accuracy (69% upon cross-validation; see Table 9 for detailed classification results). The variables selected by the stepwise DFA to classify poorly identified vowel tokens were F2 and F1 at midpoint, F2 slope, vowel duration, F1 at onset and offset, formant movement from onset to midpoint, and F2 at offset. Poorly identified tokens were classified with 55.6% accuracy (51.6% upon cross-validation; see Table 9 for detailed classification results).

In an effort to identify parsimonious classification models of well- and poorly identified vowel tokens, a second set of DFAs that limited the inclusion of variables to the first four variables entered into each of the original DFAs—F1 and F2 at midpoint, F2 slope, and vowel duration—was conducted. The parsimonious models classified well-identified tokens with 67.6% accuracy (66.1% cross-validated accuracy) and poorly identified tokens with 49.8% accuracy (48.4% cross-validated accuracy).

Subanalysis 3

In this descriptive analysis, only those tokens misclassified by the DFA and misperceived by listeners were considered to evaluate the degree with which degraded vowel acoustics influenced the resulting percept. The level of agreement between classification and perceptual errors was interpreted to reflect the extent to which the nature of the degraded vowel acoustics affected the resulting percept.

A confusion matrix of misclassified (DFA) to misperceived (vowel accuracy) vowel tokens is found in Table 10. It is important to note that the classification results of the DFA were constrained, such that errors were limited to one of the nine other vowels. However, the perceptual data were collected from an unconstrained transcription task; thus, perceptual errors were not limited to the nine other vowels studied here. Examples of other perceptual errors made by listeners were diphthong or r-colored substitutions and vowel omissions. To constrain the perceptual data to match the constraints of the acoustic data, other perceptual errors were excluded from the calculations of percent agreement between misclassified tokens and misperceptions. Because of these constraints, a level of 10% agreement can be applied to reflect the likelihood an agreement between misclassified and misperceived vowels occurred by chance. Results of this analysis revealed agreement percentages that varied from 23% to 48% depending on the vowel. Thus, the results of this analysis suggest that the nature perceptual errors made by listeners were influenced by the vowel acoustic degradations.

Discussion

The results of this analysis demonstrated that listeners enjoyed a 12% benefit to vowel accuracy when transcribing

Table 9. Classification summary of well-identified and poorly identified vowel tokens.

Vowel	Predicted vowel (%)									
	i	ɪ	e	ɛ	æ	a	o	u	ʌ	ʊ
Well-identified target										
i	96		4							
ɪ		75	1	14			1	1	7	
e	6	4	87	3						
ɛ		24	1	45	21	1	1		7	
æ		7	3	17	66	1			6	
a					5	71	12	2	9	1
o		1				7	74	3	14	
u	10	14					8	65	1	1
ʌ		6		4	8	11	9	1	54	7
ʊ		6	6				6			81
Poorly identified target										
i	75	7	10	3	2			2		2
ɪ	12	34	16	12	7		4	11	2	2
e	17	6	67	4	4					2
ɛ		31		44	18		3	3	2	
æ		5	1	17	57	12	4	1		4
a				3	1	75	9	3		9
o		1		1		4	66	11	8	8
u	6	15	3	3			16	51		7
ʌ		5		5	3	17	20	5	37	8
ʊ		6	6		6		17			67

Note. 71.2% of well-identified vowels were correctly classified (69% upon cross-validation); 55.6% of poorly identified vowels were correctly classified (51.6% upon cross-validation). Agreements are presented along the diagonal and are shaded in gray.

Table 10. Percentage of misclassified to misperceived vowel agreement.

Classification error	Identification error (%)									
	i	ɪ	e	ɛ	æ	a	o	u	ʌ	ʊ
i	32	17	23	6	2		6	4	11	
ɪ	10	42	8	5	6	6	1	5	15	1
e	7	10	47	22	7	3		2	3	
ɛ	1	13	12	48	9	6		3	5	3
æ	2	8	2	10	38	6	2		27	4
a	2	7	2	41		27	7		14	
o	1	4	1	5		14	39		8	29
u	8	7		10	5	7	7	23	30	3
ʌ		14	3	11	8	13	6	5	34	6
ʊ	6	10	2	14	4	16	12	6	6	24

Note. Agreements are presented along the diagonal and are shaded in gray. The level of agreement exceeded the individual levels of disagreement for all misclassified vowels, with the exception of those tokens misclassified as /u/ and /a/. The larger disagreements are shaded in yellow.

acoustically distinctive vowel tokens (i.e., tokens correctly classified via discriminant function analysis) than when transcribing less distinctive tokens. This finding was corroborated by the results of the second subanalysis, which revealed that tokens perceived by listeners with 100% accuracy were classified via DFA with 20% greater accuracy than those tokens that presented perceptual challenges to listeners (identified with 0%–60% accuracy). These results support the notion that the relationship between degraded vowel acoustics and intelligibility in dysarthria may be causative and not simply correlative in nature. Further support of this notion was provided by the results of the qualitative analysis, which demonstrated several noteworthy agreements between the DFA misclassification and perceptual errors. The level of agreement varied from 23% to 48% depending on the vowel, and it exceeded the individual levels of disagreements for all vowel tokens, except for those tokens misclassified as /u/ and /a/. Vowel tokens misclassified (by DFA) as /u/ and /a/ were more frequently perceptually misperceived as /ʌ/ and /ɛ/, respectively. Thus, the relationship between the nature of the vowel degradation and the resulting percept is not perfect. However, this should be expected given the extravowel information provided by the phrasal stimuli (e.g., by neighboring consonants and syntax) that likely shaped listeners' vowel perceptions and misperceptions. So even with the extravowel influences, the degraded acoustic information associated with vowels produced by speakers with dysarthria was revealed to influence perception of those vowels in connected speech.

General Discussion

Compressed or reduced vowel space area has been demonstrated in dysarthria arising from various neurological conditions, including ALS, Parkinson's disease, and cerebral palsy (Liu et al., 2005; Tjaden & Wilding, 2004; Weismer et al., 2001). However, this has not been universally reported (e.g., see Sapir, Spielman, Ramig, Story, & Fox, 2007; Weismer et al., 2001). A possible limitation of previous studies attempting to relate degraded vowel acoustics to perception in dysarthria is that measures approximating overall intelligibility (e.g., scaled intelligibility estimates or percentage of words correct), not vowel accuracy, have been the perceptual units of interest. Because overall intelligibility is influenced by more information than what is provided by vowels (e.g., consonants, suprasegmental and linguistic features), it is plausible that this practice has prevented causative interpretation of the findings. Specifically, conclusions implicating degraded vowel acoustics as a contributory factor to the intelligibility disorder associated with dysarthria were deemed premature (Weismer et al., 2001).

In the present report, vowel space metrics capturing vowel reduction and centralization, reduced F2 slopes, and mean dispersion among the vowels demonstrated significant relationships with both intelligibility and vowel accuracy. These findings emerged not only for established metrics, such as VSA and mean dispersion, but also for recently introduced and novel metrics, including corner vowel to /ʌ/ dispersion, a novel metric capturing vowel centralization. Another

recently introduced metric, the FCR, which is proposed to minimize variability arising from interspeaker differences while maximizing sensitivity to vowel centralization, has been shown to differentiate between the vowel spaces of non-disordered and hypokinetic speakers (Sapir et al., 2010) but to date has not been used to predict intelligibility. Results of the present investigation suggest that the FCR was correlated with both intelligibility and vowel accuracy. The FCR considers only the formant information of three corner vowels, and its calculation is highly dependent on the formant information associated with /u/ (represented twice in the numerator). In this data set, /u/ tokens were fairly disparate, particularly along the F2 dimension and /a/ along the F1 dimension. It is possible that the instability of these tokens may be inflating the FCR. In our previous report (Lansford & Liss, 2014), the FCR was among the poorer delimiters of dysarthric and healthy control vowel spaces. Although it reliably categorized healthy control speakers (i.e., good specificity), it misclassified a large number of dysarthric speakers as healthy controls (i.e., poor sensitivity). The results of the present analysis, in concert with those of the previous report, suggest that the FCR warrants further investigation with regard to its usefulness in studying vowel reduction in dysarthria.

H. Kim et al. (2011) introduced a metric referred to as overlap degree that, when compared with VSA and other vowel metrics, accounted for the greatest amount of variance in intelligibility scores in speakers with dysarthria secondary to cerebral palsy. As reported by Kim and her colleagues, overlap degree is the misclassification rate of vowel tokens (/i/, /ɪ/, /ɛ/, /a/, /ʊ/, and /u/) categorized via DFA for each speaker. In the diverse population of dysarthric speakers studied here, spectral overlap degree did not reach the values from the Kim study ($R^2 = .96$), but it was (a) moderately correlated with vowel accuracy ($R = -.405$) and (b) included in a few of the predictive models of intelligibility and vowel accuracy. The discrepancy in predictive power is likely due to differences in perceptual task, stimuli, and subsets of vowels studied. Nevertheless, the results of the present investigation provide evidence supporting the use of recently introduced and novel vowel metrics that capture centralization and vowel distinctiveness to study dysarthric vowel perception.

The results of the first analysis linked degraded vowel acoustics to reduced perceptual outcome measures, including vowel accuracy. However, the direct implications of such degradations on the resulting percept were evaluated specifically in the second experiment. Results of the first subanalysis revealed that tokens that were more acoustically distinctive (i.e., correctly classified via DFA) were better identified. The second subanalysis validated these findings by demonstrating that well-identified tokens (i.e., those token identified with 100% accuracy) were classified with better accuracy than those tokens that presented perceptual challenges to the listener (i.e., tokens identified with 0%–60% accuracy). Taken together, these findings support the notion that the acoustic distinctiveness of vowel tokens may have a specific impact on intelligibility, as suggested by Neel (2008). Finally, some degree of agreement between the classification and perceptual errors was demonstrated for all vowels, suggesting that the

nature of the vowel degradation has some influence over its misperception. The level of agreement, however, was stronger for some vowels than for others. Specifically, misclassification–misperception agreement was strongest for front vowels that vary along the tongue-height (F1) dimension. As evidenced by the results of the DFA and the nature of the misclassification errors, front vowels possess a tight articulatory working space, raising the likelihood of eliciting perceptual errors. Thus, it follows that the acoustic features that led to misclassification of vowels by DFA in this tight working space similarly guided the nature of the perceptual errors.

It is important to note that the statistical treatment of the acoustic metrics was unconstrained by extravowel factors. For example, values extracted for the vowel in "push" could be misclassified as any other of the nine vowels, without regard for whether it created a real word (e.g., "pish"). The listeners, conversely, were told that the phrases consisted of real English words to tie their perceptions (and perceptual decisions) to the lexicon. This opens a host of other lexical influences that invariably influence perception of vowels embedded in phrases (e.g., lexical neighborhood densities, word frequency, phonotactic constraints, and context). With all of these extravowel influences, it might be expected that the contributions of vowel acoustics would be extremely limited in predicting perceptual performance in phrases for listeners. However, the results of the present analysis, which linked degraded vowel acoustics to the resulting percept in a heterogeneous cohort of speakers with dysarthria, suggest that the acoustic signature of vowels in dysarthria should be regarded as a cause of intelligibility deficit, along with other degraded source and filter acoustics.

The acoustic information carried by vowels serves important communicative functions. The formant structure provides cues to vowel identity, facilitating lexical access and lexical competition. The spectral information also serves to cue word boundaries, facilitating the task of lexical segmentation in connected speech. By establishing the link between vowel production errors and the nature of perceptual errors, targeted therapeutic interventions can be developed to improve vowel production. For example, reduced high–low vowel contrast (i.e., reduced distance or dispersion of front and/or back vowels) in a speaker with dysarthria may (a) produce perceptual errors along the same dimension, challenging the process of lexical activation; or (b) obscure strong–weak syllabic distinctions, hindering segmentation of the speech stream. Thus, a goal of speaker-directed therapy may be to increase spectral distinctiveness of neighboring vowel tokens along the affected dimension. The outcome measures of such an intervention could be specified to the communicative function. For example, when vowel reduction results in insufficient contrast between strong and weak vowels, the outcome measure of an intervention could be based on lexical segmentation analysis of naive listener transcripts. The results of this investigation also support the development of intervention techniques that aim to ease the communicative demands placed on the disordered speaker by maximizing the listener's ability to understand the degraded message. For example, when speaker-directed therapy is not feasible, as is the case for many patients diagnosed with progressive neurodegenerative disorders, caregivers, therapists, and other communicative partners could undergo perceptual training aimed to retune their perceptual boundaries for specific vowel tokens to accommodate less acoustically distinctive vowel tokens. Benefits to intelligibility following therapy or perceptual training are predicted by the outcomes of this investigation.

Conclusion

In the present set of experiments reported in this and our companion article (Lansford & Liss, 2014), a variety of acoustic vowel space metrics (e.g., VSA, dispersion, and F2 slope) were considered with regard to their abilities to (a) differentiate vowels produced with speakers with and without dysarthria (Lansford & Liss, 2014) and (b) predict perceptual outcomes. In addition, their contributions were evaluated within the context of a broad cohort of dysarthric speakers. With fairly equivalent groups of speakers diagnosed with the various dysarthria subtypes, exploration of dysarthria-specific effects on vowel production (represented acoustically) was possible (reported in Lansford & Liss, 2014). Results of the present investigation demonstrated that degraded vowel acoustics have some specific effect on human perceptual performance, even in the presence of extravowel variables that naturally exert influence in phrase perception. Thus, in this large cohort of speakers with varied dysarthria types and speech severity rating, the results of this work support the importance of vowel production deficits to the intelligibility disorder caused by dysarthria.

Acknowledgments

References

Bigham, D. (2008). *Dialect contact and accommodation among emerging adults in a university setting* (Doctoral dissertation). University of Texas at Austin.

Bradlow, A., & Bent, T. (2002). The clear speech effect for non-native listeners. *The Journal of the Acoustical Society of America, 112,* 272–284.

Bunton, K., & Weismer, G. (2001). The relationship between perception and acoustics for a high-low vowel contrast produced by speakers with dysarthria. *Journal of Speech, Language, and Hearing Research, 44,* 1215–1228.

Flynn, N. (2011). Comparing vowel formant normalization procedures. *York Working Papers in Linguistics (Series 2), 11,* 1–28.

Gertsman, L. (1968). Classification of self-normalized vowels. *IEEE Transactions on Audio Electroacoustics, AU-16,* 78–80.

Higgins, C., & Hodge, M. (2002). Vowel area and intelligibility in children with and without dysarthria. *Journal of Medical Speech & Language Pathology, 10,* 271–277.

Hillenbrand, J. M., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America, 97,* 3099–3111.

Kent, R. D., Weismer, G., Kent, J. F., & Rosenbek, J. C. (1989). Toward phonetic intelligibility testing in dysarthria. *Journal of Speech and Hearing Disorders, 54,* 482–499.

Kent, K., Weismer, G., Kent, J., Vorperian, H., & Duffy, J. (1999). Acoustic studies of dysarthric speech: Methods, progress and potential. *Journal of Communication Disorders, 32,* 141–186.

Kim, H., Hasegawa-Johnson, M., & Perlman, A. (2011). Vowel contrast and speech intelligibility in dysarthria. *Folia Phoniatrica et Logopaedica, 63,* 187–194.

Kim, Y.-J., Weismer, G., Kent, R. D., & Duffy, J. R. (2009). Statistical models of F2 slope in relation to severity of dysarthria. *Folia Phoniatrica et Logopaedica, 61,* 329–335.

Kutner, M., Nachtsheim, C., Neter, J., & Li, W. (2005). *Applied linear statistical models* (5th ed.). New York, NY: McGraw-Hill/Irwin.

Labanov, B. M. (1971). Classification of Russian vowels spoken by different speakers. *The Journal of the Acoustical Society of America, 49,* 606–608.

Lansford, K. L., & Liss, J. M. (2014). Vowel acoustics in dysarthria: Speech disorder diagnosis and classification. *Journal of Speech, Language, and Hearing Research.* Advance online publication. doi:10.1044/1092-4388(2013/12-0262)

Lindblom, B. (1963). Spectrographic study of vowel reduction. In R. D. Kent, J. L. Miller, & B. S. Atal (Eds.), *Papers in speech communication: Speech perception* (pp. 517–525). New York, NY: The Acoustical Society of America.

Liu, H. M., Tsao, F. M., & Kuhl, P. K. (2005). The effect of reduced vowel working space on speech intelligibility in Mandarin-speaking young adults with cerebral palsy. *The Journal of the Acoustical Society of America, 117,* 3879–3889.

McRae, P. A., Tjaden, K., & Schoonings, B. (2002). Acoustic and perceptual consequences of articulatory rate change in Parkinson disease. *Journal of Speech, Language, and Hearing Research, 45,* 35–50.

Nearey, T. M. (1989). Static, dynamic, and relational properties in vowel perception. *The Journal of the Acoustical Society of America, 85,* 2088–2112.

Neel, A. T. (2008). Vowel space characteristics and vowel accuracy. *Journal of Speech, Language, and Hearing Research, 51,* 574–585.

Payton, K., Uchanshki, R., & Braida, L. (1994). Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing. *The Journal of the Acoustical Society of America, 95,* 1581–1592.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America, 24,* 175–184.

Picheny, M., Durlach, N., & Braida, L. (1985). Speaking clearly for the hard of hearing: I. Intelligibility differences between clear and conversational speech. *Journal of Speech and Hearing Research, 28,* 96–103.

Sapir, S., Ramig, L., Spielman, J., & Fox, C. (2010). Formant centralization ratio (FCR) as an acoustic index of dysarthric vowel articulation: Comparison with vowel space area in Parkinson disease and healthy aging. *Journal of Speech, Language, and Hearing Research, 53,* 114–125.

Sapir, S., Spielman, J., Ramig, L., Story, B., & Fox, C. (2007). Effects of intensive voice treatment (the Lee Silverman Voice Treatment [LSVT]) on vowel articulation in dysarthric individuals with idiopathic Parkinson disease: Acoustic and perceptual findings. *Journal of Speech, Language, and Hearing Research, 50,* 899–912.

Strange, W. (1989a). Dynamic specification of coarticulated vowels spoken in sentence context. *The Journal of the Acoustical Society of America, 85,* 2135–2153.

Strange, W. (1989b). Evolving theories of vowel perception. *The Journal of the Acoustical Society of America, 85,* 2081–2087.

Tjaden, K., & Wilding, G. E. (2004). Rate and loudness manipulations in dysarthria: Acoustic and perceptual findings. *Journal of Speech, Language, and Hearing Research, 47,* 766–783.

Turner, G., Tjaden, K., & Weismer, G. (1995). The influence of speaking rate on vowel space and speech intelligibility for individuals with amyotrophic lateral sclerosis. *Journal of Speech and Hearing Research, 38,* 1001–1013.

Uchanski, R. M., Choi, S. S., Braida, L. D., Reed, C. M., & Durlach, N. I. (1996). Speaking clearly for the hard of hearing: IV. Further studies of the role of speaking rate. *Journal of Speech and Hearing Research, 39,* 494–509.

Watt, D., & Fabricius, A. (2002). Evaluation of a technique for improving the mapping of multiple speakers' vowel spaces in the F1~F2 plane. *Leeds Working Papers in Linguistics and Phonetics, 9,* 159–173.

Weismer, G., Jeng, J.-Y., Laures, J., Kent, R. D., & Kent, J. F. (2001). Acoustic and intelligibility characteristics of sentence production in neurogenic speech disorders. *Folia Phoniatrica et Logopaedica, 53,* 1–18.

Weismer, G., & Martin, R. (1992). Acoustic and perceptual approaches to the study of intelligibility. In R. D. Kent (Ed.), *Intelligibility in speech disorders: Theory measurement and management* (pp. 67–118). Amsterdam, the Netherlands: John Benjamin.

Whitehill, T. L., Ciocca, V., Chan, J. C.-T., & Samman, N. (2006). Acoustic analysis of vowels following glossectomy. *Clinical Linguistics & Phonetics, 20,* 135–140.