Degraded Vowel Acoustics and

the Perceptual Consequences in Dysarthria

by

Kaitlin L. Lansford

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved December 2011 by the
Graduate Supervisory Committee:

Julie M. Liss, Chair
Tamiko Azuma
Michael Dorman
Andrew Lotto

ARIZONA STATE UNIVERSITY

May 2012

ABSTRACT

Distorted vowel production is a hallmark characteristic of dysarthric speech, irrespective of the underlying neurological condition or dysarthria diagnosis. A variety of acoustic metrics have been used to study the nature of vowel production deficits in dysarthria; however, not all demonstrate sensitivity to the exhibited deficits. Less attention has been paid to quantifying the vowel production deficits associated with the specific dysarthrias. Attempts to characterize the relationship between naturally degraded vowel production in dysarthria with overall intelligibility have met with mixed results, leading some to question the nature of this relationship. It has been suggested that aberrant vowel acoustics may be an index of overall severity of the impairment and not an “integral component” of the intelligibility deficit. A limitation of previous work detailing perceptual consequences of disordered vowel acoustics is that overall intelligibility, not vowel identification accuracy, has been the perceptual measure of interest. A series of three experiments were conducted to address the problems outlined herein. The goals of the first experiment were to identify subsets of vowel metrics that reliably distinguish speakers with dysarthria from non-disordered speakers and differentiate the dysarthria subtypes. Vowel metrics that capture vowel centralization and reduced spectral distinctiveness among vowels differentiated dysarthric from non-disordered speakers. Vowel metrics generally failed to differentiate speakers according to their dysarthria diagnosis. The second and third experiments were conducted to evaluate the relationship between degraded vowel acoustics and the resulting percept. In the second experiment, correlation and regression analyses revealed vowel metrics that capture vowel centralization and distinctiveness and movement of the second formant frequency were most predictive of vowel identification accuracy and overall intelligibility. The third experiment was conducted to evaluate the extent to which the nature of the acoustic degradation predicts the resulting percept. Results suggest distinctive vowel tokens are better identified and, likewise, better-identified tokens are more distinctive. Further, an above-chance level agreement between nature of vowel misclassification and misidentification errors was demonstrated for all vowels, suggesting degraded vowel acoustics are not merely an index of severity in dysarthria, but rather are an integral component of the resultant intelligibility disorder.

DEDICATION

To my husband and children, both born and in utero, with love.

Andres ~ You have been unwavering in your support, love and patience. The words “thank you” do not adequately express my gratitude.

Mia ~ You are my sunshine…

My dissertation baby ~ Thank you, baby girl, for staying put and not misbehaving!

ACKNOWLEDGEMENTS

I would like to take this opportunity to formally thank all those who offered me guidance, support and love as I completed this culminating experience. First and foremost, I would like to acknowledge my advisor, mentor and future collaborator, Dr. Julie Liss. Thank you, Julie, for training me well and always having my back. Without your unfaltering support, even through the darkest of days, I know that I would not be where I am today. I would also like to thank the members of my committee, Drs. Tamiko Azuma, Michael Dorman and Andrew Lotto, for their roles in influencing and enhancing my course of study and research endeavors.

A truly special thanks goes out to my lab ladies, Dena Berg, Angela Davis, Cindi Hensley and Rebecca Norton. Not only were these women instrumental contributors to my project, but also they were my unflagging cheerleaders. The Motor Speech Disorders Lab is now a highly productive environment that is infused with humor, friendship and just the slightest hint of debauchery. This transformation is largely due to the unique and delightful personalities offered by each of these women.

To my closest colleague, Rene Utianski, thank you for your patience, collaboration and friendship. Without all three, these last couple of years would have been a struggle (or at least more of one)! I'd also like to thank my original cohort members, Anthony Koutsoftas and Virginia Dubasik, for inspiring me to be a better scientist and for all of the laughs along the way.

Words cannot adequately express the gratitude I have for my family and friends. In particular, I’d like to thank my husband, parents and daughter for their tolerance and encouragement, and my closest friend, Kendra Flory, for honoring the self-imposed “no fun” policy, put into full effect a couple of months ago.

Finally, I’d like to acknowledge the financial support afforded to me by the Ruth L. Kirschstein National Service Research Award (NRSA) awarded by NIH/NIDCD (F31DC010093). In addition, the Graduate Research Support Program grant awarded by the Graduate Professional Student Association at Arizona State University funded a portion of this project.

TABLE OF CONTENTS

Page

Page

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

A COMPREHENSIVE REVIEW OF VOWEL PERCEPTION

Introduction

In optimal listening conditions, spoken language is processed with considerable ease. The contributions of segmental (e.g., acoustic-phonetic), suprasegmental (e.g., prosodic) and linguistic (e.g., lexical, sublexical and syntactic) information to segmentation and perception of spoken language have been a focus of speech perception investigations for the past several decades. The relative importance of segmental information offered by vowels and consonants to overall word recognition has been the source of recent debate (e.g., Cole et al., 1996; Fogerty & Kewley-Port, 2009; Kewley-Port, Burkle & Lee, 2007; Owren & Cardillo, 2006). Traditionally, information carried by consonants was considered to be the most crucial segmental component of spoken language processing (Owens, Talbot & Schubert, 1968). Indeed, this view is supported by the written language processing literature (Lee, Rayner & Pollatsek, 2001; see Shimron, 1993 for a review), primarily owing to greater number of consonants as compared to vowels in the English language. However, evidence from recent investigations challenges this traditional notion with respect to spoken language processing. For example, Kewley-Port, Burkle and Lee (2007) replaced either the vocalic or consonantal segments of sentences with noise, rendering each sentence as containing only consonant or vowel information, respectively. The authors found a 2:1 advantage to intelligibility for the vowel-only sentences for both healthy young adults and elderly adults with hearing loss. This finding, also supported by

Cole et al. (1996) and Fogerty and Kewley-Port (2009), suggests the absence of vowels from a speech signal is more detrimental to recovering the intended message than the absence of consonants.

These results should not be surprising, though, as the information contained in vowel segments, particularly in vowel transitions, cue listeners not only to identification of vowels, but also to neighboring consonants via coarticulation (Cooper, Delattre, Liberman, Borst & Gerstman, 1952; Liberman, Cooper, Shankweiler & Studdert-Kennedy, 1967). This observed vowel superiority effect, however, may be limited to processing of sentential information, as conflicting results have been found for monosyllabic (Fogerty & Humes, 2010) and multisyllabic (Owren & Cardillo, 2006) words. However, it is important to note that while the relative potency of the segmental information offered by vowels and consonants to speech perception is unclear, accurate identification of both vowels and consonants is a crucial component of models of word recognition (Luce & Pisoni, 1998; McClelland & Elman, 1986; Norris, 1994).

Briefly, models of word recognition (e.g., Trace, Shortlist, and Neighborhood Activation Model) describe this process as occurring in two phases, *activation* and *competition* of lexical candidates. First, a pool of lexical candidates is *activated* in response to incoming acoustic-phonetic information. The activated lexical candidates subsequently *compete.* The candidate that most resembles the acoustic-phonetic input “wins” the competition (i.e., is perceived by the listener). Thus, poor production or misperception of the vowel /ɪ/ in the word *ship* results in a pool of activated lexical candidates that may or may not include the intended target, thereby, decreasing the likelihood that the word *ship* will win the subsequent lexical competition.

The effects of vowel misperception extend beyond that of word recognition, as information gleaned from vowels can be used to facilitate speech segmentation (Cutler & Buttlerfield, 1992; Cutler & Carter, 1987; Liss, Spitzer, Caviness & Adler, 2000; Mattys, Melhorne, & White, 2005; Spitzer, Liss & Mattys, 2007). Mattys, Melhorne and White (2005) describe a hierarchical model that specifies the use of linguistic, segmental and suprasegmental information in speech segmentation is dependent on the quality of the listening condition. In optimal listening conditions, listeners rely upon linguistic, specifically lexical, information to segment the speech stream. Thus, speech segmentation occurs as a consequence of word recognition. However, in suboptimal listening conditions, speech segmentation strategies adapt to incorporate segmental and suprasegmental information to facilitate deciphering of connected speech. Specifically, stress information contained in strong syllables (e.g., presence of unreduced vowel, increased duration and amplitude) has the potential to cue word onsets in English, as the first syllable in most English words is strong (Culter & Carter, 1987). Thus, distorted/degraded vowel production and/or hindered perception of information contained in vowels may have deleterious effects on overall speech perception resulting in decreased intelligibility of the speech signal.

Despite the observed consequences of vowel misperception to speech perception, much work is needed to delineate the link between vowel production and the resulting percept. Dysarthria, a motor speech disorder arising from neurological impairment, is an ideal context for the study of the interface between vowel production and perception, as vowel production in dysarthria is commonly distorted (Darley, Aronson & Brown, 1969a, b, 1975; Duffy, 2005). The ways in which this production deficit is related to overall intelligibility has been widely investigated, albeit with dramatically varied results. The relationship between vowel production and vowel identification in dysarthria, however, has received less attention. The purpose of this review of the literature is to detail findings from classic and recent investigations of vowel perception in both non-disordered and dysarthric populations in order to identify areas that require greater attention.

Models of Vowel Perception

It has been demonstrated that the identification of vowels requires sufficient spectral and temporal cues such that perceptual distinctions can be made (Peterson & Barney, 1952; Hillenbrand, Getty, Clark & Wheeler, 1995). Early investigations of vowel perception revealed the importance of the formant frequencies, particularly F1 and F2, to perceptual identification (Delattre, Liberman, Cooper & Gertsman, 1952) and categorization (Peterson & Barney, 1952) of vowel tokens. Briefly, acoustic-articulatory coupling of vowels can be summarized by inverse relationships between F1 and tongue height and F2 and tongue advancement (Fant, 1960; Ladefoged, 1975).

Simple Target Model

Perhaps the most "textbook" model of vowel perception born out of these classic findings has become known as the *simple target model*. According to this model, vowels targets are canonically represented and each can be defined acoustically by a single point in a two (or three) dimensional plane comprised of its first two (or three) formant frequencies. While the simplicity of this model is attractive, it suffers from several limitations that prevent it from being applied to the perception of vowels produced in context and/or by many talkers.

Many shortcomings of the simple target model of vowel perception were revealed by the work of Peterson & Barney (1952). In this seminal investigation, ten vowels produced by men, women and children were classified with 75% accuracy when the discriminant function analysis was privileged to F1 and F2 information sampled from each vowel's steady state. Classification accuracy improved with the addition of fundamental frequency (F0) or F3 information (85.9% and 83.6%, respectively). Thus, 14-25% of the vowel tokens were misclassified depending on the spectral information provided to the DFA. Vowel misclassifications were largely attributed to spectral overlap of neighboring vowels, introduced by both inter-speaker (e.g., formant frequency differences depending on size of vocal tract size and shape) and intra-speaker (e.g., articulatory undershoot) causes. Despite the spectral overlap of neighboring vowels, listeners classified the same vowel tokens with approximately 94% accuracy. If the simple target approach accurately models vowel perception, the perceptual error rate (roughly 6%) should mirror that of the misclassification rate derived by the DFA (14-25%). This hypothesis is not supported by the data. The Peterson and Barney study has been criticized for not including temporal measurements and for sampling vowel tokens at a single point in time; thereby, ignoring dynamic aspects of vowel production as a consequence (Hillenbrand et al., 1995). The importance of Peterson and Barney's study should not be minimized by these limitations, however, as the results of this investigation contributed greatly to formation of subsequent models of vowel perception.

Elaborated Target Models

Peterson and Barney (1952) were among the first researchers to reveal differences in formant frequencies of vowel tokens depending on the length and shape of the speaker's vocal tract. Despite the vastly different formant frequency averages revealed for vowel tokens produced by male, female and children speakers, listeners demonstrated very little difficulty with perceptual discrimination of the vowel tokens. Speaker normalization, a process whereby the perceptual system of the listener recalibrates to accommodate individual speakers, is proposed to account for the ease with which we understand spoken language produced by multiple speakers with different sized and shaped vocal tracts. *Elaborated target models* address this shortcoming of the simple target model by incorporating speaker normalization in their accounts of vowel perception. Formant ratios (e.g., comparison of F1 and F2 to F0 and/or F3) and psychophysically motivated transformations (e.g., log, mel, bark, and Koenig) are common normalization procedures (e.g., Hillenbrand & Gayvert, 1987; Miller, 1989; Monahan & Idsardi, 2010; Syrdal & Gopal, 1986).

In a recent article, Flynn (2011) compares 20 methods of vowel normalization with respect to their ability to eliminate inter-speaker 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 & Fabricus, 2002) eliminated variability arising from inter-speaker 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. While elaborated target models account well for variability in vowel production arising from differences vocal tract shapes and sizes of the speakers, intra-speaker variability in vowel production, such as articulatory undershoot in connected speech, is not addressed by these models.

Dynamic Specification Models

Rarely do vowel tokens produced in context reach their canonical values (Lindblom, 1963). This phenomenon, known as articulatory undershoot (i.e., target undershoot or vowel reduction), largely is attributed to coarticulation. The effects of coarticulation on a vowel target's formant frequencies depend on the consonantal context (Stevens & House, 1963), speaking style (e.g., casual vs. clear; Lindblom, 1983) and rate of speech (Gay, 1978). Despite such articulatory undershoot, listeners perceive vowels produced in context with ease (Macchi, 1980). Perceptual "overshoot" on the part of the listener is one mechanism that has been proposed to cope with articulatory undershoot (Divenyi, 2009; Lindblom & Studdert-Kennedy, 1967). A number of theoretical accounts of perceptual overshoot have been proposed. For example, articulatory/gestural theories of speech perception (e.g., motor theory or direct realism) propose the mere existence of perceptual overshoot provides evidence that listeners perceive the intended target gestures associated with vowel production from the reduced acoustic signal (Fowler, 1994). Alternatively, general auditory theories of speech perception consider perceptual overshoot a consequence of context effects and have demonstrated perceptual overshoot even when primed with non-speech stimuli (Holt, Lotto & Kluender, 2000; Lotto & Holt 2006). Regardless of the theoretical explanation of perceptual overshoot, it is essential to identify the acoustic cues associated with reduced vowel production that facilitate perceptual overshoot. Evidence suggests the information contained in the dynamic aspects of vowel production is responsible for perceptual overshoot (Strange, 1989b).

The simple target model is criticized for its failure to incorporate the dynamic and temporal aspects of vowel production to the process of vowel perception (Strange, 1989a). Dynamic vowel metrics that capture spectral change over time have been revealed to improve vowel discrimination (Hillenbrand, Clark & Nearey, 2001; Hillenbrand et al., 1995; Strange, 1989b). Hillenbrand and his colleagues (1995) replicated Peterson and Barney's work in an attempt to address its limitations and demonstrated slightly less accurate classification by DFA, accuracy ranging from 68-84% depending on the composition of the static spectral variables included in the classification models (e.g., F1, F2, F3, F0 measured from the vowel's midpoint). With the inclusion of vowel duration and spectral measurements taken at three time points (20%, 50% and 80% of vowel duration), the ability of the DFA to reliably classify the vowels reached as high as 94.8%. Thus, inclusion of acoustic metrics that capture the dynamic nature of vowel production improved discrimination.

Some monophthongs are inherently more dynamic (e.g., /æ/, /^/ and /ʊ/) than others (e.g., /ɪ/, /ɛ/ and /u/; Neel, 2008). The acoustic underpinnings of more or less dynamic vowels may very well serve as acoustic cues to vowel identification, particularly in connected speech. However, as previously stated, a primary cause of articulatory undershoot is coarticulation. Acoustic metrics that capture these effects are logical starting points for investigating the acoustic underpinnings of perceptual overshoot.

Indeed, evidence of articulatory undershoot often is observed in the formant transitions into and out of the vowel nucleus (Hillenbrand et al., 2001). Perceptually, formant transitions have been demonstrated to be just as important, if not more important than the steady state vowel segments, for vowel discrimination (Strange, 1989b; Strange, Jenkins & Johnson, 1983; Fox, 1989; Jenkins, Strange & Trent, 1999). In addition, formant transitions are fairly stable across speakers (Hillenbrand et al., 2001), indicating this acoustic feature of vowel production may facilitate speaker normalization as well.

Relationship Between Vowel Production and Perception

With these approaches to relating vowel acoustics to perception serving as a backdrop, it is important to ask how acoustic degradation of vowels influences the resulting percept. A variety of vowel metrics, spectral and temporal, static and dynamic, have been established to study this interface more closely. One context in which the relationship between vowel production and perception has been closely examined is in clear (hyper-articulated) versus conversational (citation-style) speech. Acoustic analyses of clear and conversational vowels revealed a number of important distinctions including longer vowel durations, larger vowel spaces, greater vocal intensity of vowels, increased high-low vowel contrastivity and greater formant movement in hyper-articulated vowels (Ferguson & Kewley-Port, 2002; Moon & Lindblom, 1994; Picheny, Durlach & Braida, 1986). Clear speech has been shown to yield greater intelligibility scores, particularly for non-native listeners (Bradlow & Bent, 2002) and the hearing-impaired (Payton, Uchanski, & Braida, 1994; Picheny, Durlach & Braida, 1985; Uchanski, Choi, Braida & Durlach, 1996). While the exact underpinnings of this clear-speech intelligibility benefit are unknown, vowel space expansion and increased vowel duration have been demonstrated to account for some of the intelligibility gains offered by clear speech (Ferguson & Kewley-Port, 2007).

To better understand the relationship between vowel acoustics and subsequent identification, Neel (2008) regressed a variety of derived vowel space measurements against the vowel identification scores from the Hillenbrand database and found that subsets of these metrics accounted for only 9-12% of the variance. However, well-identified vowels were found to be distinctive in F1 and F2, duration and formant movement over time as compared to poorly identified vowel tokens. Neel concluded that measurements of vowel distinctiveness among neighboring vowels, rather than vowel space area, might prove more useful in predicting vowel identification accuracy.

The weak relationship between traditional vowel space area metrics and vowel identification accuracy measures observed in Neel's study may be due to reduced variability in the perceptual data, as overall vowel identification accuracy was greater than 95% for both male and female speakers. The ceiling effect observed in these data is likely secondary to the uses of a highly constrained listening task (e.g., forced-choice, hVd paradigm) and speech stimuli obtained from neurologically healthy speakers. Of interest would be investigating the relationships between acoustic vowel metrics and vowel accuracy and intelligibility with acoustic and perceptual datasets with greater variability (e.g., with disordered speakers using a less constrained task). Despite the limitations of this study, it is important to note that Neel's results, along with those observed in clear vs. conversational speech studies, appear to provide support to the use of acoustic vowel metrics in the prediction, and potentially modeling of intelligibility of degraded and disordered speech (e.g., dysarthria).

Vowel Production in Dysarthria

Distorted vowel production is a hallmark characteristic of dysarthric speech, irrespective of the underlying neurological condition (Darley, Aronson & Brown, 1969a, b, 1975; Duffy, 2005). Thus, studying the effects of degraded vowel production on listeners' perception in this population is an ecological choice; in that the outcomes have the potential to not only inform speech perception theory but also to guide clinical practice.

Kinematic Data

In general, dysarthric vowel production is characterized by articulatory undershoot resulting in a compressed or reduced working vowel space (Kent & Kim, 2003). Such acoustic consequences of production deficits caused by motor speech disorders have been investigated widely. The articulatory underpinnings, however, have received less attention. Until recently, evidence detailing articulatory kinematics in dysarthria has been limited to case studies or to a small pool of subjects (Ackermann, Grone, Hoch & Schonle, 1993; Forrest & Weismer, 1995; Kent & Netsell, 1975, 1978; Kent, Netsell & Bauer, 1975). However, a series of studies investigating vowel production in patients with dysarthria secondary to either amyotrophic lateral sclerosis (ALS) or Parkinson's disease (PD) using x-ray microbeam technology have made important contributions to this growing body of literature. For example, Weismer, Yunusova and Westbury (2003) found tongue retraction and elevation and increased lip closure in speakers with ALS produces a lowering of F2 in /u/. Additional findings from x-ray microbeam studies include reduced excursion of tongue movements and reduced speed of lower lip and tongue (but not jaw) movements during vowel production in patients with dysarthria secondary to ALS relative to control speakers (Yunusova, Weismer, Westbury & Lindstrom, 2008). This finding was not revealed in dysarthric patients diagnosed with PD (Yunusova et al., 2008). Interarticulator coordination during vowel production for both patients with ALS and PD has not been found to differ significantly from control vowel production (Weismer et al., 2003; and Yunusova et al., 2008). However, Yunusova et al. (2008) found incoordination of the articulators in a handful of severely involved patients and noted that such incoordination may be a sign of disease progression. Evidence delineating the perceptual consequences of abnormal articulator kinematics in patients with ALS is emerging. Specifically, overall intelligibility has been found to decrease as a function of reduced speed of articulator movements during vowel production (Yunusova, Green, Lindstrom, Ball, Pattee, & Zinman, 2009).

It has been suggested that the articulatory differences found for dysarthric speakers relative to neurologically healthy speakers may be secondary to reduced scaling of movement (Yunusova, et al., 2008). Yunusova, Weismer and Lindstrom (2011) address this question with a linear discriminant analysis (LDA). Dysarthric vowels (ALS and PD vowel productions) were classified with a constellation of time-varying kinematic measures derived from a model that reliably classified vowel productions of neurologically healthy (i.e., control) individuals. PD vowel productions were reliably classified with the control-based model, albeit not with the same degree of accuracy, but the misclassification errors were in the same direction as the control errors. However, ALS vowel productions were not classified reliably with the control-based model. An alternate constellation of articulator movement derived from the ALS data demonstrated greater success with vowel classification. In sum, these results suggest any differences in articulator movement between neurologically healthy speakers and PD patients are likely due to reduced scaling of movement, but vowel production in patients with ALS is categorically different than that of neurologically healthy participants. Much of the kinematic work completed to understand the articulatory underpinnings of distorted vowel production has been limited to the PD and ALS populations. Of interest would be expansion of this line of research to include motor speech disorders arising from other neurological impairments.

Acoustic Data

As previously mentioned, the acoustic consequences of dysarthria on vowel production have been widely investigated (e.g., Kim, Weismer, Kent & Duffy, 2009; Rosen, Goozee & Murdoch, 2008; Turner, Tjaden & Weismer, 1995; Ziegler & von Cramon, 1983a, 1983b, 1986; Watanabe, Arasaki, Nagata & Shouiki, 1994; Weismer, Jeng, Laures, Kent & Kent, 2001; Weismer, Martin, Kent & Kent, 1992). Kent, Weismer, Kent, Vorperian and Duffy (1999) summarize the most commonly reported vowel production abnormalities as centralization of formant frequencies, reduction of vowel space area (quadrilateral or triangular), and abnormal formant frequencies for both high and front vowels. Other acoustic findings detailed are vowel formant pattern instability and reduced F2 slopes.

Evidence demonstrating the acoustic properties of dysarthric vowel production are distinguishable from control production is mixed. Relative to control speakers, movement of the second formant during vowel production, captured in a variety of contexts (e.g., CV transitions, diphthongs, and monophthongs), is reduced in some dysarthric speakers (Kim et al., 2009; Rosen et al., 2008; Weismer et al., 1992, 2001). Weismer and his colleagues (1992, 2001) found shallower F2 trajectories in male speakers with dysarthria secondary to ALS relative to age/gender-matched controls. Similar results have been revealed for speakers with dysarthria secondary to PD, stroke (Kim et al., 2009) and multiple sclerosis (Rosen et al., 2008).

Measures capturing overall vowel space area (quadrilateral or triangular) have demonstrated less reliable discriminability. Weismer et al. (2001) found vowel space area (VSA), as calculated as the area within the irregular quadrilateral formed by the first and second formants of the corner vowels, /i/, /æ/, /a/, and /u/, was reduced relative to control speakers in male speakers with ALS. No group differences were revealed for ALS female speakers or for dysarthric speakers with PD relative to control speakers. Somewhat contradictory to the findings of Weismer et al., quadrilateral VSA group differences were revealed for speakers with PD relative to control, but not for speakers with MS (Tjaden & Wilding, 2004). Also noteworthy, the vowel space areas of patients with PD and MS did not differ significantly (Tjaden & Wilding, 2004). Sapir, Spielman, Ramig, Story and Fox (2007) also failed to reveal a significant VSA (triangular) difference between control and PD speakers. However, between group differences were revealed for the following metrics, F2 of the vowel /u/ and the ratio of F2i/F2u.

Tjaden, Rivera, Wilding and Turner (2005) derived the vowel space area encompassed by the lax vowels /ɪ/, /ɛ/ and /ʊ/ to investigate the proposal that lax vowel production may be unaffected by motor speech disorders due to their reduced articulatory production demands (Turner et al., 1995). This hypothesis was partially supported by the data, as lax vowel space for speakers with PD could not be differentiated from that of control. Conversely, lax vowel space was robust to differences between ALS and control vowel productions. The authors speculate that the differential effects found for lax vowel spaces of PD and ALS patients may be attributed to differences in underlying pathophysiology or to overall severity differences found for the two groups (ALS more severe than PD).

Similar findings of failure to differentiate between dysarthric (specifically hypokinetic) vowel spaces from control with traditional measurements of vowel space area have led to the proposal of alternative methods of capturing centralization of formant frequencies (Sapir, Ramig, Spielman, & Fox, 2010; and Skodda, Visser & Schlegel, 2011). Sapir and his colleagues (2010) propose the formant centralization ratio (FCR) as a vowel space metric that maximizes sensitivity to vowel centralization while minimizing interspeaker variability in formant frequencies (i.e., normalizing the vowel space). 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. Sapir et al. demonstrated that the FCR, unlike the triangular VSA metric, reliably distinguished hypokinetic vowel spaces from those of neurologically healthy speakers. Skodda et al. (2011) propose the vowel articulation index (VAI), the exact inverse of the FCR, to discriminate hypokinetic from control vowel spaces. Similar justification is provided for use of the VAI, as it is an index of vowel centralization that minimized interspeaker variability. The VAI was compared with triangular vowel space with respect to its ability to discriminate the vowel spaces of 68 speakers with hypokinetic dysarthria from those of 32 neurologically healthy speakers. Triangular VSA demonstrated between group differences for male hypokinetic and non-disordered speakers only. However, the VAI values were significantly reduced for both hypokinetic male and female speakers relative to the non-disordered speakers. The authors conclude metrics that minimize interspeaker variability while maximizing vowel centralization may be more sensitive to mild dysarthria than traditional VSA metrics.

To fully understand how dysarthric and control vowel production are distinctive, greater attention must be paid not only to the effects of underlying neurological impairment, but also to those of overall severity of the speech disorder and other production deficits that hinder accurate perception of the intended vowel (e.g., hypernasality and articulation rate). One method of revealing the acoustic differences between control and dysarthric vowel production is via investigation of the perceptual challenges associated with distorted vowel production in dysarthria.

Dysarthric Vowel Perception

The effects of dysarthric vowel production on perceptual outcome measures vary widely depending on the dysarthric population being studied, the severity of the speakers and the acoustic and perceptual measures used to evaluate the relationship. As previously mentioned, dynamic 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). As summarized, the production deficits characteristic of dysarthria may have deleterious effects on acoustic metrics that capture dynamic aspects of vowel production. Thus, the investigation of the effects of disordered formant movement on intelligibility is well motivated. Kent et al. (1989) found f2 transitions correlated significantly with single word intelligibility in dysarthric patients. Weismer et al. (2001) corroborated and extended this relationship by demonstrating impressive correlations between f2 slopes of /aɪ/, /ɔ/, and /ju/ (*r* = .794, -.967 and .942 respectively) and scaled sentence intelligibility estimates in patients with dysarthria secondary to ALS and PD. In addition, ALS patients with overall scaled intelligibility estimates less than 70% had distinctly shallower F2 slopes than those with intelligibility estimates greater than 70% (Weismer, Martin, Kent & Kent, 1992). However, Kim et al. (2009) revealed a less robust, albeit significant, predictive relationship between F2 slope (measured in the words *shoot* and *wax* only) and scaled estimates of intelligibility in 40 speakers with dysarthria secondary to either PD or stroke (*n*=20). F2 slopes from *shoot* and *wax* accounted for 14.3% and 13.9% of the variance in intelligibility ratings.

The relationship between acoustic metrics approximating vowel space area (both triangular and quadrilateral) and overall intelligibility is not clear, largely due to widely variable findings. Turner et al. (1995) found VSA derived from the vowel quadrilateral accounted for 46% of the variance in scaled intelligibility ratings in patients with ALS. The same was revealed 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), vowel space area accounted for 64% of the variance in single word intelligibility scores. Similarly, Liu, Tsao and Kuhl (2005) revealed a significant correlation (*r* = .684) between vowel space area and single word intelligibility scores in Mandarin speakers with CP. However, Tjaden and Wilding (2004) demonstrated less impressive predictive power of vowel space area metrics in women with dysarthria secondary to MS or PD. 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 did not include VSA (but did include F2 slope of /aɪ/ and /eɪ/), predicted 12-21% of the variance in intelligibility scores (Tjaden & Wilding, 2004). In speakers diagnosed with PD, VSA accounted for only 12% of the variance in scaled severity scores (McRae, Tjaden & Schoonings, 2002).

Kim, Hasegawa-Johnson and Perlman (2011) use the varied VSA findings reported above as the impetus for their investigation of vowel contrast and speech intelligibility in three control speakers and nine speakers with dysarthria secondary to CP. In addition to traditional vowel space area (triangular), Kim and colleagues evaluated the ability of alternate vowel space metrics including lax vowel space area, mean Euclidean distance between the vowels, F1 and F2 variability, and overlap degree among the vowels (more on these metrics to follow) to predict intelligibility scores from a single-word transcription task. Significant regression functions were found for VSA (R^2 = .69), mean distance between the vowels (R^2 = .69), variability of F1 (R^2 = .74), and overlap degree (R^2 = .96). Interestingly, regression functions for F2 variability and lax vowel space failed to reach significance. Overlap degree was derived by the results of a per speaker classification analysis of vowel tokens into their vowel categories. Vowel misclassification rates were interpreted to reflect the degree of spectral/temporal overlap amongst the vowels. The authors concluded vowel overlap might be a more appropriate indicator of intelligibility deficits in dysarthria. However, it is important to note that the regressions reported included three control speakers, one of whom had a fairly compressed vowel space relative to the other two control speakers. When this speaker was removed from the analysis the regression function for triangular vowel space area increased from .69 to .90.

A limitation of the work detailed thus far in explaining the perceptual consequences of disordered vowel acoustics, is that overall intelligibility, not vowel identification accuracy, has been the dependent measure of interest. Fewer studies have investigated the relationship between vowel acoustics and vowel perception in dysarthria. Liu and colleagues (2005) also explored the relationship between VSA and vowel identification accuracy and found a significant correlation ($r = .63$). Whitehill, Ciocca, Chan and Samman (2006) found a significant correlation ($r = .32$) between VSA and vowel intelligibility in Cantonese speakers with partial glossectomy. While this relationship has not been directly addressed in English speakers with dysarthria, Bunton and Weismer (2001) evaluated the acoustic differences between correctly and misperceived (tongue-height errors) vowel tokens and found that they could not be reliably distinguished.

The varied results relating vowel acoustics to intelligibility have led some to question the nature of this relationship. Weismer et al. (2001) notably speculated aberrant acoustic metrics might not be an “integral component” of the intelligibility deficit. Rather, they may be an index of overall severity of the impairment, with no direct bearing on intelligibility. Yunusova, Weismer, Kent and Rusche (2005) attempted to address this possibility by relating within-speaker variability in acoustic and perceptual metrics derived from each breath group. A breath group is defined as the segment of connected speech that is measured between each breath produced by a speaker. Thus, the number of words within each breath group was not well controlled. The acoustic and perceptual metrics selected to evaluate this relationship within each breath group are a global measure of F2 variability (F2 interquartile range) and scaled intelligibility, respectively. Subjects included 10 dysarthric speakers (equal number of speakers diagnosed with PD and ALS) and 10 control speakers. Traditional regression analyses were completed predicting overall intelligibility (sentence and word) from F2 variability across-speakers and R^2 values ranged from .57 to .61. However, the ability of F2 variability to predict sentence and word intelligibility within each breath group failed to reach significance in the 6 dysarthric speakers selected for this analysis. Thus, these results support the hypothesis suggested by Weismer et al. (2001) that degraded vowel acoustics may not be an integral component of intelligibility deficits associated with dysarthria. However, the results should be interpreted with caution due to several limitations of the study, including a small sample size of speakers evaluated in the within-speakers analysis, less than optimal reliability of scaled intelligibility estimates, poorly controlled stimuli, and use of an unprecedented acoustic metric in dysarthric studies. In addition, within speaker variability in both acoustic and perceptual metrics may be fairly restricted, making it difficult to accurately assess this relationship.

Conclusions

Distorted vowel production in dysarthria is characterized by spectral and temporal degradation; flattening of spectral change formants; and vowel space distortions that may differentially affect high versus low, or front versus back contrasts. A variety of acoustic metrics have been used to study the nature of vowel production deficits in dysarthria. However, not all metrics demonstrate sensitivity to the exhibited deficits in dysarthria. Further, far less attention has been paid to quantifying the vowel production deficits associated with the specific dysarthrias.

To date, attempts to characterize the relationship between naturally degraded vowel production in dysarthria with overall intelligibility have met with mixed results. The effects of dysarthric vowel production on perceptual outcome measures vary widely depending on the dysarthric population being studied, the severity of the speakers and the acoustic and perceptual measures used to evaluate the relationship. The varied results relating vowel acoustics to intelligibility have led some to question the nature of this relationship. It has been suggested that aberrant acoustic metrics might not be an “integral component” of the intelligibility deficit. Rather, degraded vowel acoustics may be an index of overall severity of the impairment, with no direct bearing on intelligibility. A limitation of previous work detailing perceptual consequences of disordered vowel acoustics is that overall intelligibility, not vowel identification accuracy, has been the dependent measure of interest. Fewer studies have considered the relationship between vowel acoustics and vowel perception in dysarthria.

References

Ackerman, H., Grone, B.F., Hoch, G., & Schonle, P.W. (1993). Speech freezing in Parkinson's disease: a kinematic analysis of orofacial movements by means of electromagnetic articulography. *Folia Phoniutrica, 45*, 84-89.

Bigham, D. (2008). Dialect contact and accommodation among emerging adults in a university setting. Ph.D. thesis, The University of Texas at Austin.

Bradlow, A., & Bent, T. (2002). The clear speech effect for non-native listeners. *Journal of the Acoustical Society of America, 112*(1), 272-284.

Bradlow, A., Torretta, G.M. & Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication, 20*, 255-272.

Boersma, P. & Weenink, D. (2006). Praat: doing phonetics by computer (Version 4.4.24) [Computer program]. Retrieved June 19, 2006, from http://www.praat.org/

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.

Cole, R., Yan, Y., Mak, B., Fanty, M., & Bailey, T. (1996). "The contribution of consonants versus vowels to word recognition in fluent speech," in Proceedings of the ICASSP'96, pp. 853–856.

Cooper, F., Delattre, P., Liberman, A., Borst, J., & Gerstman, L. (1952). Some experiments on the perception of synthetic speech sounds. *Journal of the Acoustical Society of America, 24*, 597–606. doi: 10.1121/1.1906940

Cutler, A. & Butterfield, S. (1992). Rhythmic cues to speech segmentation: evidence from juncture misperception. *Journal of Memory and Language, 31*, 218-236.

Cutler, A. & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language*, *2,* 133-142.

Delattre, P. C., Liberman, A. M., Cooper, F. S., & Gerstman, L. J. (1952). An experimental study of the acoustic determinants of vowel color; observations on one- and two-formant vowels synthesized from spectrographic patterns. *Word 8*, 195-210.

Darley, F., Aronson, A., & Brown, J. (1969). Differential diagnostic patterns of dysarthria*. Journal of Speech and Hearing Research, 12*, 246–269.

Darley, F., Aronson, A., & Brown, J. (1975). Motor Speech Disorders. Philadelphia: W. B. Saunders Inc.

Divenyi, P. (2009). Perception of complete and incomplete formant transitions in vowels. *Journal of the Acoustical Society of America, 126*, 1427-1439. doi: 10.1121/1.3167482

Duffy, J. R. (2005). *Motor speech disorders: Substrates, differential diagnosis, and managemen*t (2nd Ed.) St. Louis, MO: Elsevier Mosby.

Fant, G. (1960). *Acoustic theory of speech production*. Mouton, the Hague.

Ferguson, S., & Kewley-Port, D. (2002). Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America, 112*(1), 259-271.

Ferguson, S. H., & Kewley-Port, D. (2007). Talker differences in clear and conversational speech: Acoustic characteristics of vowels. *Journal of Speech, Language, and Hearing Research, 50,* 1241–1255.

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

Fogerty, D. & Humes, L.E. (2010). Perceptual contributions to monosyllabic word intelligibility: Segmental, lexical, and noise replacement factors. *Journal of the Acoustical Society of America, 128*, 3114-3125.

Fogerty, D. & Kewley, Port, D. (2009). Perceptual contributions of the consonant-vowel boundary to sentence intelligibility*. Journal of the Acoustical Society of America, 126*(2), 847-857. doi: 10.1121/1.3159302

Forrest, K., & Weismer, G. (1995). Dynamic aspects of lower lip movement in Parkinsonian and neurologically normal geriatric speakers' production of stress. *Journal of Speech and Hearing Research, 38*, 260–272.

Fowler, C.A. (1994). Speech perception: Direct realist theory. In R.E. Asher (Ed.), *Encyclopedia of Language and Linguistics* (pp.4199-4203). Oxford: Pergamon.

Fox, R. (1989). Dynamic information in identification and discrimination of vowels. *Phonetica, 46*, 97–116.

Gay, T. (1978). Effect of speaking rate on vowel formant movements. *Journal of the Acoustical Society of America, 63*, 223-230.

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., Clark, M. J., & Nearey, T. N. (2001). Effect of consonant environment on vowel formant patterns. *Journal of the Acoustical Society of America, 109*, 748–763. doi:10.1121/1.1337959

Hillenbrand, J.M., & Gayvert R. (1987). Speaker-independent vowel classification based on fundamental frequency and formant frequencies. *Journal of the Acoustical Society of America, 81*(Suppl. 1), S93.

Hillenbrand, J.M., Getty, L.A., Clark, M.J. & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, *97*, 3099–31.

Holt, L. L., Lotto, A. J., & Kluender, K. R. (2000). Neighboring spectral content influences vowel identification. *Journal of the Acoustical Society of America*, *108,* 710-722.

Jenkins, J.J., Strange, W., & Trent, S.A. (1999). Context-independent dynamic information for the perception of coarticulated vowels. *Journal of the Acoustical Society of America, 106* (1), 438-448.

Kent, R. & Kim, Y. (2003). Toward an acoustic typology of motor speech disorders. *Clinical Linguistics and Phonetics, 17*(6), 427-445.

Kent, R., & Netsell, R. (1975). A case study of and ataxic dysarthric: Cineradiographic and spectrographic observations. *Journal of Speech and Hearing Disorders, 40,* 115–134.

Kent, R., & Netsell, R. (1978). Articulatory abnormalities in athetoid cerebral palsy. *Journal of Speech and Hearing Disorders, 43*, 353–373.

Kent, R. D., Netsell, R., & Bauer, L. L. (1975). Cineradiographic assessment of articulatory mobility in the dysarthrias. *Journal of Speech and Hearing Disorders, 40,* 467–480.

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.

Kewley-Port, D., Burkle, T. Z., & Lee, J. H. (2007). Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners. *J. Acoust. Soc. Am.* 122, 2365–2375. doi: 10.1121/1.2773986

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*(6), 329-335.

Labanov, B.M. (1971). Classification of Russian vowels spoken by different speakers. *JASA*49(2B): 606-8.

Ladefoged, P. (1975). *A Course in Phonetics*. (1st edition) Orlando: Harcourt Brace.

Lee, H.W., Rayner, K. & Pollatsek, A. (2001). The relative contribution of consonants and vowels to Word Identification during Reading. *Journal of Memory and Language. 44*(2). 189-205. doi:10.1006/jmla.2000.2725

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychology Review,* 74, 431–461. doi: 10.1037/h0020279

Lindblom, B. (1963). Spectrographic study of vowel reduction. *J. Acoust. Soc. Am.* 35, 1773-1781. Reprinted in Kent, R.D., Miller, J.L. and Atal, B.S. (editors), *Papers in Speech Communication: Speech Perception*, 517-525. New York: Acoustical Society of America.

Lindblom, B., & Studdert-Kennedy, M. (1967). On the role of formant transitions in vowel recognition. *Journal of the Acoustical Society of America, 42*, 830–843.

Liss, J. M., Spitzer, S. M., Caviness, J. N., & Adler, C. (2000). LBE analysis in hypokinetic and ataxic dysarthria*. Journal of the Acoustical Society of America, 107*, 3415–3424.

Liss, J.M., White, L., Mattys, S.L., Lansford, K., Spitzer, S, Lotto, A.J., and Caviness, J.N. (2009). Quantifying speech rhythm deficits in the dysarthrias. *Journal of Speech, Language, and Hearing Research, 52*(5), 1334-1352.

Liu, H.M., Tsao, F.M., and 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*(6), 3879–3889.

Lotto, A. J. & Holt, L. L. (2006). Putting phonetic context effects into context: A commentary on Fowler (2006). *Perception & Psychophysics*, *68*, 178-183.

Luce, P.A. & Pisoni, D.B. (1998) Recognizing spoken words: the neighborhood activation model. *Ear and Hearing*, *19*, 1–36.

Macchi, M.J. (1980). Identification of vowel spoken in isoloation versus vowels spoken in consonantal context. *Journal of the Acoustical Society of America, 68*, 1636-1642.

Mattys, S. L., White, L., & Melhorn, J. F (2005). Integration of multiple segmentation cues: A hierarchical framework*, Journal of Experimental Psychology General, 134*, 477–500.

McClelland, J., & Elman, J. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18,* 1-86.

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.

Milenkovic, P.H. (2004). TF32 [Computer software]. Madison: University of Wisconsin, Department of Electrical and Computer Engineering.

Miller, J.D. (1989). Auditory-perceptual interpretation of the vowel. *Journal of the Acoustical Society of America, 85*(5), 2114-2134.

Monahan, P.J. & Idsardi, W.J. (2010). Auditory sensitivity to formant ratios: toward an account of vowel normalisation. *Language and Cognitive Processes, 25(6),* 808-839.

Moon, S. Y., & Lindblom, B. (1994). Interaction between duration, context, and speaking style in English stressed vowels. *Journal of the Acoustical Society of America, 96*, 40-55.

Nearey, T.M. (1989). Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America,* 85 (5), 2088-2112.

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

Norris, D. (1994) Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52***,** 189–234.

Owens, E., Talbot, C.B., & Schubert, E.D. (1968). Vowel discrimination of hearing-impaired listeners. *Journal of Speech and Hearing Research, 11*, 648–655.

Owren, M. J., & Cardillo, G. C. (2006). The relative roles of vowels and consonants in discriminating talker identity versus word meaning. *Journal of the Acoustical Society of America, 119*, 1727–1739. doi: 10.1121/1.2161431

Payton, K., Uchanshki, R., & Braida, L. (1994). Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing. *Journal of the Acoustical Society of America, 95*(3), 1581-1592.

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

Peterson, G.E. & Lehiste, I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America, 32, 693-703.*

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.

Picheny, M., Durlach, N., & Braida, L. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research, 29*, 434-446.

Rosen, K.M, Goozee, J.V., & Murdoch, B.E. (2008). Examining the effects of Multiple Sclerosis on speech production: Does phonetic structure matter?. *Journal of Communication Disorders, 41*, 49-69.

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.

Shimron, J. (1993). The role of vowels in reading: A review of studies in Hebrew and English. *Psychological Bulletin, 114*, 52-67.

Skodda, S., Visser W., & Schlegel, U. (2011). Vowel articulation in Parkinson's disease. *Journal of Voice, 25*(4), 467-472. doi: 10.1016/j.voice.2010.01.009

Spitzer, S., Liss, J.M., & Mattys, S.L. (2007). Acoustic cues to lexical segmentation: A study of resynthesized speech. *Journal of the Acoustical Society of America, 122*(6), 3678- 3687. doi: 10.1121/1.2801545

Stevens, K.N. & House, A.S. (1963). Perturbations of vowel articulations by consonantal context: An acoustical study. *Journal of Speech and Hearing Research, 6,* 111-128.

Stilp, C.E., & Kluender, K.R. (2010). Cochlea-scaled spectral entropy, not consonants, vowels, or time, best predicts speech intelligibility. *Proceedings of the National Academy of Science, 107*(27), 12387-12392.

Strange, W. (1989a). Dynamic specification of coarticulated vowels spoken in sentence context. *Journal of the Acoustical Society of America, 85 (5), 2135-2153.*

Strange, W. (1989b). Evolving theories of vowel perception. *Journal of the Acoustical Society of America, 85*(5), 2081-2087.

Strange, W., Jenkins, J. J., & Johnson, T. L. (1983). Dynamic specification of coarticulated vowels. *Journal of the Acoustical Society of America, 74*, 695–705. doi: 10.1121/1.389855

Syrdal, A.K. & Gopal H.S. (1985). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America, 79*(4), 1086-1100.

Tjaden, K., Rivera, D., Wilding, G., & Turner, G.S. (2005). Characteristics of the lax vowel space in dysarthria. *Journal of Speech, Language, and Hearing Research, 48*(3), 554–566.

Tjaden, K., and 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.

Watanabe, S., Arasaki, K., Nagata, H., & Shouji, S. (1994). Analysis of dysarthria in amyotrophic lateral sclerosis--MRI of the tongue and formant analysis of vowels. *Rinsho Shinkeigaku, 34*(3), 217-23.

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: John Benjamin.

Weismer, G., Yunusova, Y., & Westbury, J. R. (2003). Interarticulator coordination in dysarthria: An X-ray microbeam study. *Journal of Speech, Language, and Hearing Research, 46*, 1247–1261.

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

Yunusova, Y., Green, J., Ball, L., Lindstrom, M., Pattee, G. & Zinman, L. (2010). Kinematics of disease progression in bulbar ALS. *Journal of Communication Disorders*, *43*, 6-20.

Yunusova, Y., Weismer, G., Kent, R. D., & Rusche, N. M. (2005). Breath-group intelligibility in dysarthria: Characteristics and underlying correlates. *Journal of Speech, Language, & Hearing Research, 48,* 1294-1310.

Yunusova, Y., Weismer, G., & Lindstrom, M. (2011). Classifications of vocalic segments from articulatory kinematics: healthy controls and speakers with dysarthria. *Journal of Speech, Language and Hearing Research, 54*(5), 1302-1311.

Yunusova, Y., Weismer, G., Westbury, J.R. & Lindstrom, M. (2008). Articulatory movements during vowels in speakers with dysarthria and in normal controls. *Journal of Speech, Language, and Hearing Research, 51*, 596-611.

Ziegler, W., & Cramon, D. von (1983a). Vowel distortion in traumatic dysarthria: A formant study. *Phonetica, 40,* 63-78.

Ziegler, W., & Cramon, D. von (1983b). Vowel distortion in traumatic dysarthria: Lip rounding versus tongue advancement. *Phonetica, 40*, 312-322.

Ziegler, W., & von Cramon, D. (1986). Spastic dysarthria after acquired brain injury: An acoustic study. *British Journal of Disorders of Communication, 21*, 173–187.

DEGRADED VOWEL ACOUSTICS AND THE PERCEPTUAL CONSEQUENCES IN DYSARTHRIA

Introduction

It has been demonstrated that the identification of vowels requires sufficient spectral and temporal cues such that perceptual distinctions can be made (Peterson & Barney, 1952; Hillenbrand, Getty, Clark & Wheeler, 1995). In a seminal study by Peterson and Barney (1952), vowels embedded in an /hVd/ were categorized by a discriminant function analysis (DFA) on the basis of static spectral measurements taken at each vowel's steady state. The DFA, when privileged to f0, F1, F2 and F3 information classified the vowels with roughly 86% accuracy. Hillenbrand and colleagues (1995) replicated the work of Peterson and Barney, and demonstrated slightly less accurate classification by DFA (84%). However, the ability of the DFA to reliably classify the vowels reached as high as 94.8% with the inclusion of vowel duration and spectral measurements taken at three time points (20%, 50% and 80% of vowel duration). Thus, inclusion of metrics that capture the dynamic nature of vowel production improved discrimination. The acoustic measurements derived from these works have become crucial to the development and testing of theories of vowel perception, and in defining the ways in which vowel acoustics influence speech intelligibility.

Production-Perception Relationship in Vowels

The relative potency of acoustic information conveyed by vowels—as compared to consonants—in speech perception has been widely demonstrated

(Cole et al., 1996; Fogerty & Kewley-Port, 2009; Kewley-Port, Burkle & Lee, 2007; see Owren & Cardillo, 2006, for an opposite account). Kewley-Port et al. (2007) replaced either the vocalic or consonantal segments of sentences with noise, rendering each sentence as containing only consonant or vowel information, respectively. The authors analyzed listener transcripts collected from both healthy young adults and elderly adults with hearing loss and found a 2:1 advantage to intelligibility for the vowel-only sentences for both groups of listeners. These findings, which replicated the results of Cole et al. (1996) and were supported by a subsequent study (Fogerty & Kewley-Port, 2009), suggest the absence of vowels from a speech signal is more detrimental to recovering the intended message than the absence of consonants.

Because acoustic information critical to accurate speech perception is contained in vowels, it is important to ask how degradation of vowels influences the resulting percept. A variety of vowel metrics, spectral and temporal, static and dynamic, have been established to study this interface more closely. In an investigation of speech intelligibility of sentences produced by normal speakers in quiet, Bradlow, Torrenta and Pisoni (1996) found that speakers with a larger vowel space and more variable f0 range were more intelligible than speakers with reduced vowel spaces and less variable f0. Another context in which the relationship between vowel production and perception has been closely examined is in clear (hyper-articulated) versus conversational (citation-style) speech. Acoustic analyses of clear and conversational vowels revealed a number of important distinctions including longer vowel durations, larger vowel spaces, greater vocal intensity of vowels, increased high-low vowel contrastivity and greater formant movement in hyper-articulated vowels (Ferguson & Kewley-Port, 2002; Moon & Lindblom, 1994; Picheny, Durlach & Braida, 1986). Clear speech has been shown to yield greater intelligibility scores, particularly for non-native listeners (Bradlow & Bent, 2002) and the hearing-impaired (Payton, Uchanski, & Braida, 1994; Picheny, Durlach & Braida, 1985; Uchanski, Choi, Braida & Durlach, 1996). While the exact underpinnings of this clear-speech intelligibility benefit are unknown, vowel space expansion and increased vowel duration have been demonstrated to account for some of the intelligibility gains offered by clear speech (Ferguson & Kewley-Port, 2007).

To better understand the relationship between vowel acoustics and subsequent identification, Neel (2008) regressed a variety of derived vowel space measurements against the vowel identification scores from the Hillenbrand database and found that subsets of these metrics accounted for only 9-12% of the variance. However, well-identified vowels were found to be distinctive in F1 and F2, duration and formant movement over time as compared to poorly identified vowel tokens. Neel concluded that measurements of vowel distinctiveness among neighboring vowels, rather than VSA, might prove more useful in predicting vowel identification accuracy.

The weak relationship between traditional vowel space area metrics and vowel identification accuracy measures observed in Neel's study may be due to reduced variability in the perceptual data, as overall vowel identification accuracy was greater than 95% for both male and female speakers. The ceiling effect observed in these data is likely secondary to the uses of a highly constrained listening task (e.g., forced-choice, hVd paradigm) and speech stimuli obtained from neurologically healthy speakers. Of interest would be investigating the relationships between acoustic vowel metrics and vowel accuracy and intelligibility with acoustic and perceptual datasets with greater variability (e.g., with disordered speakers using a less constrained task). Despite the limitations of this study, it is important to note that Neel's results, along with those observed in clear vs. conversational speech studies, support the use of acoustic vowel metrics in the prediction, and potential modeling of intelligibility of degraded and disordered speech (e.g., dysarthria).

Vowel Production in Dysarthria

Distorted vowel production is a hallmark characteristic of dysarthric speech, irrespective of the underlying neurological condition (Darley, Aronson & Brown, 1969a, b, 1975; Duffy, 2005). Thus, studying the effects of degraded vowel production on listeners' perception in this population is an ecological choice; in that the outcomes have the potential to not only inform speech perception theory but also to guide clinical practice. Kent, Weismer, Kent, Vorperian and Duffy (1999) summarize the most commonly reported vowel production abnormalities as centralization of formant frequencies, reduction of vowel space area (quadrilateral or triangular), and abnormal formant frequencies for both high and front vowels. Other acoustic findings detailed are vowel formant pattern instability and reduced F2 slopes.

Evidence demonstrating these acoustic properties of dysarthric vowel production are distinguishable from control production is mixed. Relative to control speakers, movement of the second formant during vowel production, captured in a variety of contexts (e.g., CV transitions, diphthongs, and monophthongs) is reduced in some dysarthric speakers (Kim, Weismer, Kent & Duffy, 2009; Rosen, Goozee & Murdoch, 2008; Weismer, Jeng, Laures, Kent & Kent, 2001; Weismer, Martin, Kent & Kent, 1992). Measures capturing overall vowel space area (quadrilateral or triangular) have demonstrated less reliable discriminability. Weismer et al. (2001) found vowel space area (VSA), as calculated as the area within the irregular quadrilateral formed by the first and second formants of the corner vowels, /i/, /æ/, /a/, and /u/, was reduced relative to control speakers in male speakers with ALS. No group differences were revealed for ALS female speakers or for dysarthric speakers with PD relative to control speakers. Somewhat contradictory to the findings of Weismer et al., quadrilateral VSA group differences were revealed for speakers with PD relative to control, but not for speakers with MS (Tjaden & Wilding, 2004). Also noteworthy, the vowel space areas of patients with PD and MS did not differ significantly (Tjaden & Wilding, 2004). Sapir, Spielman, Ramig, Story and Fox (2007) also failed to reveal a significant VSA (triangular) difference between control and PD speakers.

Similar findings of failure to differentiate between dysarthric (specifically hypokinetic) vowel spaces from control with traditional measurements of vowel space area have led to the proposal of alternative methods of capturing centralization of formant frequencies (Sapir, Ramig, Spielman, & Fox, 2010; and

Skodda, Visser & Schlegel, 2011). Sapir and his colleagues (2010) propose the formant centralization ratio (FCR) as a vowel space metric that maximizes sensitivity to vowel centralization while minimizing interspeaker variability in formant frequencies (i.e., normalizing the vowel space). Sapir et al. demonstrated that the FCR, unlike the triangular VSA metric, reliably distinguished between vowel productions of control and hypokinetic speakers, and concluded metrics that minimize interspeaker variability while maximizing vowel centralization might be more sensitive to mild dysarthria than traditional VSA metrics.

To fully understand how dysarthric and control vowel production are distinctive, greater attention must be paid to not only the effects of underlying neurological impairment, but also to those of overall severity of the speech disorder and other production deficits that hinder accurate perception of the intended vowel (e.g., hypernasality and articulation rate). One method of revealing the acoustic differences between control and dysarthric vowel production is via investigation of the perceptual challenges associated with distorted vowel production in dysarthria.

Dysarthric Vowel Perception

The effects of dysarthric vowel production on perceptual outcome measures vary widely depending on the dysarthric population being studied, the severity of the speakers and the acoustic and perceptual measures used to evaluate the relationship. Dynamic metrics that capture formant movement (specifically F2 movement) during vowel production have contributed greatly to current theories of vowel perception (Nearey, 1989, and Strange, 1989a, 1989b). The production deficits associated with dysarthria may have deleterious effects on acoustic metrics that capture dynamic aspects of vowel production. Thus, the investigation of the effects of disordered formant movement on intelligibility is well motivated. Indeed, Kent, Weismer, Kent and Rosenbeck (1989) found f2 transitions correlated significantly with single word intelligibility in dysarthric patients. Weismer et al. (2001) corroborated and extended this relationship by demonstrating impressive correlations between F2 slopes of /aɪ/, /ɔ/, and /ju/ (r = .794, -.967 and .942 respectively) and scaled sentence intelligibility estimates in patients with dysarthria secondary to ALS and PD. Kim et al. (2009) revealed a less robust, albeit significant, predictive relationship between F2 slopes and scaled estimates of intelligibility in speakers with dysarthria secondary to PD and stroke.

The relationship between acoustic metrics approximating vowel space area (VSA; triangular and quadrilateral) and overall intelligibility is not clear, largely due to widely variable findings. Such VSA measurements have demonstrated varying degrees of predictability, accounting for anywhere from 6 to 64% of the variance (Higgins & Hodge, 2002; McRae, Tjaden & Schoonings, 2002; Tjaden & Wilding, 2004; Turner, Tjaden & Weismer, 1995; Weismer et al., 2001). The extent to which VSA measures predicted intelligibility appears to be dependent on a number factors, including gender of the speaker, nature of the underlying disease and type of stimuli used in the investigation.

Kim, Hasegawa-Johnson and Perlman (2011) use the varied VSA findings reported above as the impetus for their investigation of vowel contrast and speech intelligibility in three control speakers and nine speakers with dysarthria secondary to CP. In addition to traditional vowel space area (triangular), Kim and colleagues evaluated the ability of alternate vowel space metrics including lax vowel space area, mean Euclidean distance between the vowels, F1 and F2 variability, and overlap degree among the vowels (more on these metrics to follow) to predict intelligibility scores from a single-word transcription task. Significant regression functions were found for VSA ($R^2 = .69$), mean distance between the vowels ($R^2 = .69$), variability of F1 ($R^2 = .74$), and overlap degree ($R^2 = .96$). Overlap degree was derived by the results of a per speaker classification analysis of vowel tokens into their vowel categories. Vowel misclassification rates were interpreted to reflect the degree of spectral/temporal overlap amongst the vowels. The authors concluded vowel overlap might be a more appropriate indicator of intelligibility deficits in dysarthria.

A limitation of the work detailed thus far in explaining the perceptual consequences of disordered vowel acoustics, is that overall intelligibility, not vowel identification accuracy, has been the dependent measure of interest. Fewer studies have investigated the relationship between vowel acoustics and vowel perception in dysarthria. In addition to relating VSA to word intelligibility in Mandarin patients with CP, Liu and colleagues (2005) also explored the relationship between VSA and vowel identification accuracy and found a significant correlation ($r = .63$). Whitehill, Ciocca, Chan and Samman (2006) found a significant correlation ($r = .32$) between VSA and vowel intelligibility in Cantonese speakers with partial glossectomy. While this relationship has not been directly addressed in English speakers with dysarthria, Bunton and Weismer (2001) evaluated the acoustic differences between correctly and misperceived (tongue-height errors) vowel tokens and found that were not reliably distinguishable.

The varied results relating vowel acoustics to intelligibility have led some to question the nature of this relationship. Weismer et al. (2001) notably speculated aberrant acoustic metrics might not be an "integral component" of the intelligibility deficit. Rather, they may be an index of overall severity of the impairment, with no direct bearing on intelligibility. Yunusova, Weismer, Kent & Rusche (2005) addressed this hypothesis by relating within-speaker variability in acoustic and perceptual metrics derived from each breath group in control and dysarthric speakers. The acoustic and perceptual metrics selected to evaluate this relationship within each breath group are a global measure of F2 variability (F2 interquartile range) and scaled intelligibility, respectively. Regression analysis revealed that F2 variability predicted overall intelligibility (not contained in a breath group) across-speakers and R^2 values ranged from .57 to .61. However, the ability of F2 variability to predict sentence and word intelligibility within each breath group failed to reach significance in the subset of dysarthric speakers selected for this part of the analysis. The results appear to support the hypothesis suggested by Weismer et al. (2001), although, they should be interpreted with caution due to several limitations of the study, including a small sample size of speakers evaluated in the within-speakers analysis, less than optimal reliability of scaled intelligibility estimates, poorly controlled stimuli, and use of an unprecedented acoustic metric in dysarthric studies. In addition, within speaker variability in both acoustic and perceptual metrics may be fairly restricted, making it difficult to accurately assess this relationship.

Summary and Purpose of the Present Investigation

Distorted vowel production in dysarthria is characterized by spectral and temporal degradation; flattening of spectral change formants; and vowel space distortions that may differentially affect high versus low, or front versus back contrasts. A variety of acoustic metrics have been used to study the nature of vowel production deficits in dysarthria. However, not all metrics demonstrate sensitivity to the exhibited deficits in dysarthria. Further, far less attention has been paid to quantifying the vowel production deficits associated with the specific dysarthrias. Thus, one goal of the present investigation is to identify subsets of vowel metrics that may be used to 1) reliably distinguish speakers with dysarthria from non-disordered speakers, and 2) reliably differentiate the dysarthria subtypes (**Experiment 1**).

To date, attempts to characterize the relationship between naturally degraded vowel production in dysarthria with overall intelligibility have met with mixed results. The effects of dysarthric vowel production on perceptual outcome measures vary widely depending on the dysarthric population being studied, the severity of the speakers and the acoustic and perceptual measures used to evaluate the relationship. The varied results relating vowel acoustics to intelligibility have led some to question the nature of this relationship. It has been suggested that aberrant acoustic metrics might not be an “integral component” of the intelligibility deficit. Rather, degraded vowel acoustics may be an index of overall severity of the impairment, with no direct bearing on intelligibility. A limitation of previous work detailing perceptual consequences of disordered vowel acoustics is that overall intelligibility, not vowel identification accuracy, has been the dependent measure of interest. Fewer studies have considered the relationship between vowel acoustics and vowel perception in dysarthria. The present investigation aims to add to this growing body of literature by assessing a correlative and then predictive relationship between a variety of established and novel vowel metrics and two perceptual outcome measures, overall intelligibility and vowel identification accuracy (**Experiment 2**).

Experiment 2 considers the relationship between degraded vowel acoustics and vowel perception macroscopically via correlation and regression analyses of acoustic and perceptual metrics that capture each speaker’s overall severity of impairment (e.g., vowel space area, vowel identification accuracy). This relationship is evaluated at a microscopic level in **Experiment 3** by relating the acoustic and perceptual metrics associated with each vowel token in a series of analyses.

Experiment 1

Study Overview

The goal of the first experiment is to identify vowel metrics that differentiate 1) disordered from non-disordered speakers, and 2) the dysarthria subtypes. Towards this end, means testing (e.g., t-tests and analyses of variance) and stepwise discriminant function analysis (DFA) were conducted.

Method

Speakers. Speech samples from 57 speakers (29 male), collected as part of a larger study, were used in the present analysis. Of the 57 speakers, 45 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). The remaining 12 speakers had no history of neurological impairment and served as the control group. 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. Speaker age, gender and severity of impairment are provided in Table 1.

Stimuli. All speech stimuli, recorded as part of the larger investigation, were obtained during one session (on a speaker-by-speaker basis). Participants were fitted with a head-mounted microphone (Plantronics DSP-100), seated in a sound-attenuating booth, and instructed to read stimuli from visual prompts presented on the computer screen. Recordings were made using a custom script in TF32 (Milenkovic, 2004; 16-bit, 44kHz) and were saved directly to disc for subsequent editing using commercially available software (SoundForge; Sony

Corporation, Palo Alto, CA) to remove any noise or extraneous articulations before or after target utterances. The speakers read 80 short phrases aloud in a "normal, conversational voice." The phrases all contained 6 syllables and were composed of 3-5 mono- or disyllabic words, with low semantic transitional probability. The phrases alternated between strong and weak syllables, where strong syllables were defined as those carrying lexical stress in citation form. The acoustic features and listeners' perceptions of vowels produced within the *strong* syllables were the targets of analysis.

Of the 80 phrases, 36 were selected for the present analysis (see Appendix A). The phrases were divided into two stimulus lists, each produced by half of the speakers. The productions of 18 phrases per speaker were analyzed. The lists were balanced for presence of vowels, such that each of the ten vowels (/i/, /ɪ/, /e/, /ɛ/, /æ/, /u/, /ʊ/, /o/, /a/ and /^/) was represented equally. In addition, the speaker composition of each stimulus set was balanced for severity of the speech impairment (based on clinical judgment; see Table 1). Within each stimulus set, a vowel was produced a minimum of four times, thus the acoustic and perceptual analyses were limited to 4 tokens per vowel per speaker (with the exception of /ʊ/). The vowel /ʊ/ is represented in only three of the 80 experimental phrases. Because many of the vowel space area acoustic metrics require measurements from all ten vowels, measurements of /ʊ/ were derived from all three phrases per speaker, irrespective of their assigned stimulus set.

Acoustic metrics. All speech samples were analyzed using Praat (Boersma & Weenik, 2006). Vowels were identified and segmented by two trained members of the Motor Speech Disorders Lab at Arizona State University via visual inspection of the waveform and spectrogram according to standard segmentation criteria (Petersen & Lehiste, 1960; see Liss et al., 2009 for a detailed description of the vowel segmentation strategies used).

Static formant measurements. The first and second formants were measured in Hz at each vowel's onset (20% of vowel duration), midpoint (50% of vowel duration) and offset (80% of vowel duration). F0 measurements were made at the vowel's midpoint. In addition, total vowel duration (ms) was measured. To determine inter- and intra-rater reliability of the formant measurements, 10% of all vowel tokens were re-measured by same and different judges. Inter- and intra-rater reliability (Cronbach's alpha) was demonstrated to be .889 and .886 for F1 and .884 and .819 for F2 measurements, respectively.

Dynamic formant measurements. Measures that capture the dynamic nature of vowel production were calculated for each vowel token. The dynamic measures include slope of the second formant from onset to offset and formant movement (Euclidean distance) in F1 X F2 perceptual space captured in four ways: 1) from vowel onset to midpoint, 2) from midpoint to offset, 3) from onset to offset, and 4) sum of movement obtained from onset to midpoint and from midpoint to offset.

Global and fine-grained vowel space metrics. As described by Neel (2008), vowel metrics derived from static and dynamic formant measurements generally are designed to capture either 1) the mean characteristics of the entire vowel set or 2) the distinctiveness of each speaker's vowels. Vowel metrics representing the mean characteristics of the entire vowel set, also known as global vowel space metrics, typically include the following: mean F0, F1 and F2, and mean duration (Bradlow et al., 1996; and Neel, 2008). In the present analysis, mean fundamental and formant frequency metrics were derived by averaging the respective midpoint measurements (in Hz) across the ten vowels. Likewise, mean duration was calculated via averaging duration across the ten vowels. Vowel metrics that capture vowel distinctiveness, known as fine-grained vowel space metrics, include the following: vowel space area, mean distance (or dispersion) among the vowels, range of F0, F1 and F2, ratio of most dynamic to least dynamic vowels (dynamic ratio) and ratio of longest to shortest vowels (duration ratio; see Table 2 for the calculations used to derive each global and fine-grained metric).

Alternate vowel space area metrics. Recent evidence supports the use of alternate vowel space area metrics to explore vowel production deficits associated with dysarthria (Sapir et al., 2010 and Skodda et al., 2011). Specifically, the formant centralization ratio (FCR), an alternative to traditional vowel space area, is touted to maximize the effects of vowel centralization while minimizing inter-speaker effects. Sapir and colleagues (2010) revealed the FCRs derived for patients with hypokinetic dysarthria and non-disordered speakers were significantly different. To evaluate the ability of the FCR to capture vowel space reduction in a diverse sample of speakers with dysarthria, the FCR was calculated for all speakers and included in the present analysis. Similarly, Skodda et al. (2011) propose the vowel articulation index (VAI), the exact inverse of the FCR, to discriminate hypokinetic from control vowel spaces. Similar justification is provided for use of the VAI, as it is an index of vowel centralization that minimized interspeaker variability. The authors speculate metrics that minimize interspeaker variability while maximizing vowel centralization may be more sensitive to mild dysarthria than traditional VSA metrics. Considering the VAI is the inverse of the FCR, only the FCR was derived for each speaker.

Dispersion/distance vowel space metrics. Several established and novel dispersion and distance metrics were calculated in order to capture the many ways the vowel space might be warped. For example, depending on the nature of the vowel production deficit, the vowel space associated with front and/or back vowels may be differentially compressed. In order to capture front vowel space compression, the Euclidean distance in F1 x F2 space between /i/ and /æ/ and mean dispersion of the front vowels was derived for each speaker. The Euclidean distances between high vowels /i/ and /u/ and low vowels /æ/ and /a/ were also calculated as an index of high and low vowel compression. Dispersion metrics have the potential to capture vowel reduction and degree of spectral overlap among neighboring vowels. Thus, the following metrics were calculated for each speaker to be included in the analysis: mean dispersion of the corner vowels to /^/, mean dispersion of all vowels to the global formant means, and mean dispersion between neighboring vowel pairs. Liu and colleagues (2011) introduced another metric proposed to capture the degree of spectral overlap of neighboring vowels within a speaker. Briefly, this metric is the vowel misclassification rate revealed by discriminant function analysis conducted for each speaker.

F2 slope metrics. Finally, reduced F2 slope is reportedly related to perceptual decrements associated with dysarthria (e.g., Kent et al., 1989, Kim et al., 2009; Weismer et al., 2001). Accordingly, the absolute values of the F2 slopes from vowel onset to offset were averaged across the entire vowel set. Additionally, the absolute values of F2 slopes associated with the most dynamic vowels were averaged and included in this analysis. (For more information regarding the global, fine-grained and alternate vowel space metrics described, see Table 2).

In the present analysis, global, fine-grained, alternate, dispersion/distance and F2 slope vowel space metrics were derived from the obtained static and dynamic vowel measurements to assess their abilities to 1) differentiate control and disordered speakers and 2) discriminate among the dysarthria subtypes.

Results

Dysarthric versus non-disordered. In order to identify metrics sensitive to vowel production deficits associated with dysarthric speech, a series of t-tests was conducted comparing the mean scores of 12 non-disordered and 45 dysarthric speakers. Despite the unequal sample sizes, parametric treatment was appropriate for all but five variables. For these five variables, Mann-Whitney U tests were conducted to evaluate the between group differences. (See Tables 3 and 4 for group means and t-test results, respectively). Briefly, mean vowel duration was the only global vowel space metrics that demonstrated significant between group differences. Mean vowel duration in the disordered speaker group was significantly longer than that observed in the non-disordered group. Overall, the fine-grained vowel space metrics demonstrated greater sensitivity to the acoustic differences associated with disordered and non-disordered speech than global vowel space metrics. Specifically, significant differences were revealed for vowel space area, mean dispersion, F1 and F2 range and the ratio of long to short vowels. Of the 13 alternate measures, only two failed to demonstrate between group differences (Euclidean distances between high vowels, /i/ and /u/, and low vowels /ae/ and /a/).

Vowel space metrics that demonstrated significant between group differences were included in a stepwise discriminant function analysis (DFA) to determine which were best suited to differentiate disordered from control speakers. At each step of the DFA, the variable that minimizes Wilks' lambda is entered into the DFA, provided its F-statistic is significant ($p < .05$). This process continues until none of the remaining variables' F-statistics reaches significance. At any point during the stepwise DFA, a variable can be removed from the classification function should its F statistic no longer be significant ($p > .10$). Canonical variables, representing linear combinations of the selected predictors, were established to create the classification rules for group membership. The ability of the stepwise DFA to classify speakers into their appropriate groups was supported by a cross-validation procedure. This method constructs the classification rule using all of the observations with the exception of one. The excluded observation is then classified based on the established rule. The following variables were selected by the stepwise DFA: Euclidean distance between front vowels, /i/ and /æ/, in F1 X F2 space, Euclidean distance between back vowels, /u/ and /a/, in F1 X F2 space, spectral overlap degree, mean vowel duration and average F2 slope. Speakers were classified as dysarthric or non-disordered with 96.5% accuracy (94.7% accuracy on cross-validation). All non-disordered speakers were classified accordingly. Two dysarthric speakers were misclassified.

Dysarthria subtypes. The vowel metrics calculated for the 45 speakers with dysarthria were subjected to one-way analyses of variance (ANOVAs) to identify those sensitive to dysarthria-specific effects. Significant between group differences were revealed for 3 of the vowel metrics, average F2 slope, F2 slope of the most dynamic vowels, and mean vowel duration (see Table 5 for ANOVA results and Table 6 for group means of metrics with significant between group differences). To explore the between group differences in average F2 slope, F2 slope of the most dynamic vowels, and mean vowel duration, multiple comparison analysis were conducted. Briefly, mean vowel duration was shorter and average F2 slope and F2 slope of the most dynamic vowels was greater for speakers diagnosed with hypokinetic dysarthria than those with ataxic or mixed flaccid-spastic dysarthrias. Additionally, mean vowel duration was shorter and average F2 slope and F2 slope of the most dynamic vowels was greater for hyperkinetic speakers than for mixed flaccid-spastic speakers.

The variables that demonstrated significant between group differences were included in the subsequent stepwise DFA. Mean vowel duration was the sole variable selected by the DFA and classified the dysarthric speakers by subtype with 62.2% accuracy (same upon cross validation). Evaluation of the output (see Table 7) revealed reliable classification of speakers with PD (roughly 92% accuracy), yet classification of the other three subtypes ranged from 40-58.3%.

Discussion

Dysarthric versus non-disordered. Overall, fine-grained, alternative, distance/dispersion and F2 slope metrics demonstrated greater sensitivity to the acoustic differences associated with dysarthric and non-disordered vowel production than global vowel space metrics.

Dysarthric speakers exhibited longer vowel duration compared to non-disordered speakers. This finding is not surprising given the reduction in overall speaking rate for most speakers with dysarthria. Relatedly, the duration ratio of long to short vowels (a fine-grain measure) was reduced for dysarthric speakers relative to non-disordered, indicating a reduced contrast between long and short vowels. Prolonged vowel duration (together with prosodic differences not discussed in this paper) associated with dysarthria is likely the cause of the duration ratio reduction.

As expected, reductions in VSA and mean vowel space dispersion were revealed for speakers with dysarthria. Similarly, the FCR, an alternative to VSA, associated with dysarthric vowel production was significantly higher than that of non-disordered speakers, suggesting the presence of vowel centralization in dysarthric speakers. This conclusion is further supported by findings that revealed reductions in mean dispersion between the corner vowels and /^/ and mean dispersion between spectral neighbors and an increase in spectral overlap of vowels in dysarthric speakers relative to non-disordered.

The ranges of the first and second formants (fine-grained metrics) were reduced for dysarthric relative to non-disordered speakers, indicating a potential for reductions in both high-low and front-back vowel contrasts. A closer look at the formant minima and maxima revealed no differences in F2 minima between non-disordered and dysarthric speakers. Relatedly, the Euclidean distance measured in F1 x F2 perceptual space between the high-low corner vowel pairs /i, æ/ and /u, a/ in speakers with dysarthria was significantly shorter than that of non-disordered speakers. Mean front and back vowel space dispersion (along the high-low dimension) was significantly less for dysarthric than non-disordered speakers. Distance reduction was not revealed, however, for front-back corner vowel pairs, /æ, a/ and /i, u/, suggesting the contrast between front-back vowel pairs, but not high-low vowels, is preserved in dysarthric speakers. Based on these findings, it is not surprising that two of the three variables entered into the DFA to differentiate dysarthric from non-disordered speakers were the distance measures between the high-low corner vowel pairs /i, æ/ and /u, a/. These acoustic findings track to previously reported perceptual data that revealed a frequent occurrence of tongue-height vowel errors in dysarthria (Bunton & Weismer, 2001).

Dysarthria subtypes. Overall, only mean vowel duration and the F2 slope metrics demonstrated sensitivity to the acoustic differences associated with the dysarthria subtypes. Results of the multiple comparison analyses revealed that speakers with hypokinetic dysarthria are differentiated from those with ataxic or mixed flaccid-spastic dysarthrias by mean vowel duration and the F2 slope metrics. A post-hoc analysis comparing mean vowel duration, mean F2 slope of all vowels and mean F2 slope of the most dynamic vowels associated with non-disordered and hypokinetic vowel productions failed to reveal significant between-group differences. Thus, acoustic metrics that differentiate hypokinetic from other dysarthric speakers cannot be used to discriminate hypokinetic from non-disordered speakers.

Experiment 2

Study Overview

Experiment 2 was conducted to evaluate the varied relationships between the vowel metrics and overall intelligibility (words correct) and vowel identification accuracy. These relationships were evaluated via correlation and regression analyses.

Method

Speakers. All disordered speakers described in Experiment 1 were included.

Stimuli. Same as in Experiment 1.

Acoustic metrics. The vowel metrics derived in Experiment 1 were used.

Perceptual task

Listeners. Listeners were 120 undergraduate and graduate students (115 female) recruited from the Arizona State University population. Listeners' ages ranged from 18-54 with a mean age of 24, 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, six listening blocks per dysarthria group were created. In each listening block, listeners heard three different phrases produced by the twelve speakers. The speaker/phrase composition of each listening block was counterbalanced such that perceptual data for each speaker's production of the 18 phrases were collected.

Procedures. Five listeners were randomly assigned to each of the six listening blocks per speaker group. Thus the perceptual dataset 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 free of 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 while the phrases were comprised of English words, the words were strung together in a manner that rendered the phrase meaningless. 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 by two trained members of the motor speech disorders lab for 1) words correctly identified and 2) vowel identification accuracy. Vowel tokens were identified correctly 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. Misidentified tokens were coded as 0's, and the erroneously perceived vowel was noted for a subsequent analysis (e.g., if *meet* was transcribed as *met*, vowel identification accuracy was coded as a 0, and the misidentification was coded as an /ɛ/). Vowel identification accuracy was averaged in two ways for subsequent analyses. First, token accuracy was computed by averaging the binary token identification scores across the 5 listeners. Thus, for each speaker, a total of 36 token accuracy scores (4 tokens per 9 vowels) were calculated. Next, *vowel identification accuracy* was computed by averaging the token accuracy scores for all of the vowels per speaker.

Results

Perceptual data. T-tests were conducted to ensure the speakers assigned to sets 1 and 2 did not differ significantly on the perceptual measurements. Neither vowel identification accuracy nor intelligibility scores (% words correct) obtained from the speakers assigned to the two stimuli lists differed significantly.

Mean vowel identification accuracy for set 1 and 2 speakers were 69% (*SD* = .20) and 71% (*SD* = .17), respectively and intelligibility scores for set 1 and 2 speakers were 49% (*SD* = .21) and 50% (*SD* = .20), respectively. Thus, the perceptual data obtained for sets 1 and 2 were analyzed together.

Overall intelligibility and vowel accuracy scores obtained from the listeners of each dysarthric speaker may be found in Table 8. Two one-way ANOVAs were conducted to evaluate the effect of dysarthria group on intelligibility scores and vowel identification accuracy. The main effect of dysarthria group was not significant for intelligibility scores [$F(3, 41) = .825$, $p = .488$] or for vowel identification 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 identification accuracy.

Correlation analysis. To evaluate the relationships between the global, fine-grained, alternate, dispersion/distance and F2 slope vowel metrics and the perceptual outcome measures (intelligibility and vowel accuracy) Pearson correlation analysis was conducted. Correlations between the global vowel space metrics and the perceptual outcome measures revealed only a moderate inverse relationship between mean vowel duration and vowel identification accuracy ($r = -.318$; see Table 9). A number of moderate positive relationships were revealed between the fine-grained vowel space metrics and the perceptual outcome measures (see Table 10). Notably, negligible relationships were revealed between the fine-grained metrics, F0 range, the ratio of the most to least dynamic vowels and the ratio of the longest to shortest vowels, and both perceptual outcome measures, intelligibility and vowel accuracy. Finally, a number of moderate relationships were revealed between the perceptual metrics and the alternate, dispersion and F2 slope metrics (see Table 11).

Regression analysis. The interdependency of the vowel metrics was investigated and as expected many moderate to strong correlations between vowel space metrics exist (see Appendix B). A benefit to using stepwise regression methods to identify subsets of variables predictive of intelligibility and vowel accuracy is that effects of multicollinearity generally are circumvented. Due to the large set of acoustic variables, forward stepwise regression was conducted in order to construct predictive models of intelligibility and vowel accuracy.

The acoustic data were not normalized for this experiment in order to preserve the ability of the various vowel space metrics to capture the acoustic degradations. Due to the known spectral differences in vowels produced by male and female speakers (Hillenbrand et al., 1995; 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.

Intelligibility. All vowel metrics were included in the stepwise multiple regression. The regression entered the following metrics into the predictive model of intelligibility: mean dispersion of the corner vowels to /^/, mean F1, spectral overlap and mean F2 slope (adjusted $R^2 = .423$, $p < .001$; see Table 12 for regression details). Deleterious effects of multicollinearity are not present in this model, as the variance inflation factor (VIF) was less than 2 for all variables entered into the model (VIF < 5 indicates an issue with multicollinearity). In summary, greater distance between the corner vowels and /^/, lower mean F1, reduced spectral overlap, and greater excursion of the F2 slope are associated with better overall intelligibility.

For female dysarthric speakers, the subset of variables containing mean slope of the most dynamic vowels, mean dispersion of the corner vowels to /^/, and spectral overlap was best predictive of intelligibility (adjusted $R^2 = .749$, $p < .001$; see Table 12 for regression details). Thus, greater excursion of the F2 slope in dynamic vowels, greater distance between the corner vowels and /^/ and reduced spectral overlap were associated with greater intelligibility scores. For the male dysarthric speakers, only mean dispersion of the corner vowels to /^/ was selected by the stepwise regression (adjusted $R^2 = .182$, $p < .05$; see Table 12 for regression details). Increased distance between /^/ and the corner vowels was associated with increased intelligibility scores.

Vowel accuracy. All vowel metrics were included in this analysis. Formant centralization ratio, mean F2 slope, and range of F2 were selected by the stepwise regression to be included in the predictive model of vowel identification accuracy (adjusted $R^2 = .473$, $p < .001$; see Table 13 for regression details). Thus, reduced formant centralization, greater excursion of the F2 slope and restricted F2 range were associated with increased vowel identification 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 identification accuracy (adjusted $R^2 = .794$, $p < .001$; see Table 13 for regression details). Formant centralization ratio, VSA and mean F2 slope were best predictive of vowel identification scores in male speakers (adjusted $R^2 = .495$, $p < .001$; see Table 13 for regression details). Interestingly, and not predicted, vowel space area reduction, reduced formant centralization, and increased F2 slope were associated with increased vowel identification accuracy.

Discussion

Acoustic metrics capturing reduced working vowel space (e.g., VSA, FCR and various distance/dispersion metrics) were most predictive of both overall intelligibility and vowel identification accuracy. In general, vowel space area decrements, irrespective of the measurement method, are associated with reduced intelligibility and vowel identification accuracy. The intelligibility findings revealed in this experiment are in line with the results of previous studies conducted in dysarthria. Crucially, however, the results of this analysis extend such previous findings to include vowel identification accuracy as an affected perceptual outcome measure of degraded vowel acoustics. In fact, the regression analyses predicting vowel identification accuracy from subsets of acoustic variables accounted for more variance than models predicting intelligibility.

The degree of variance accounted for by these acoustic metrics is impressive given the top-down influences provided to listeners by the stimuli (e.g. lexical and syntactic) and the fact that all vowel metrics were derived from vowel tokens embedded in connected speech. The results of this experiment provide strong evidence relating degraded vowel acoustics to vowel perception; however, conclusions suggesting degraded vowel acoustics are an integral component of the intelligibility disorder caused by dysarthria are premature at this point.

Experiment 3

Study Overview

Experiment 3 was conducted to consider the relationship between vowel acoustics and perception at a microscopic level. Towards this end, the acoustic and perceptual data collected per token are treated in a variety of ways. First, in order to test the hypothesis that vowel tokens with distinctive spectral and temporal acoustics are more accurately perceived, perceptual token accuracy scores (collected via listeners) of correctly classified and misclassified vowel tokens (via DFA) were compared. Next, to validate and extend the findings of the first analysis, tokens identified with 100% accuracy and tokens identified with 0-60% accuracy were compared with respect to their ability to be classified via discriminant function analysis. It is expected that well-identified vowel tokens will be classified with greater accuracy than those vowel tokens that present perceptual challenges to the listener. Finally, in order to address the concern that degraded vowel acoustics are merely indices of severity and not integral components of the intelligibility disorder in dysarthria (Weismer et al., 2001), a point-by-point analysis comparing misclassified vowel tokens to listeners' misperceptions was conducted.

Method

Speakers. All disordered speakers described in Experiment 1 were included.

Stimuli. Same as in Experiment 1.

Acoustic metrics. The static and dynamic formant and temporal measurements associated with each vowel token (obtained in Experiment 1) were the acoustic units of interest in this experiment. 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 sampled at 50% duration, total vowel duration, slope of the second formant from onset to offset and formant movement (Euclidean distance) in F1 X F2 perceptual space captured in four ways: 1) from vowel onset to midpoint, 2) from midpoint to offset, 3) from onset to offset, and 4) 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 inter-speaker variation[1]. The

---

[1] Flynn (2011) compares 20 methods of vowel normalization with respect to their ability to eliminate inter-speaker 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

data were normalized for this experiment in order to improve classification accuracy of the discriminant function analysis.

Perceptual metrics. The token accuracy scores, calculated from listener transcripts and described in Experiment 1, were used in this experiment. In addition to overall scores, correct token identifications and misidentifications for each speaker were coded and assembled into confusion matrices (see Table 14). Overall, vowel tokens were perceived with 71% accuracy.

Results

Analysis 1. The static and dynamic formant metrics associated with each vowel token (as described in Experiment 1) produced by all 45 dysarthric speakers were used to classify the tokens as one of the ten vowels via stepwise discriminant function analysis. The following variables were selected by the stepwise DFA to classify the 1749 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 15 for classification summary).

---

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; and Watt and Fabricus, 2002) eliminated variability arising from inter-speaker 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.

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

Analysis 2. To validate and extend the results from the first analysis, vowel tokens perceived with 100% accuracy (*n* = 768) and those with 60% and less accuracy (*n* = 638) were subjected to separate stepwise classification analyses, in which the static and dynamic formant and temporal measurements were used to classify well-perceived and poorly perceived 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 with 71.2% accuracy (69% upon cross validation; see Table 16 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 17 for detailed classification results).

In an effort to identify classification models of well- and poorly identified vowel tokens with greater parsimony, a second set of DFAs that limited entry of variables to the first four variables entered into 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). The spectral differences associated with well- and poorly identified tokens are depicted in Figures 1 and 2, respectively.

Analysis 3. In this descriptive analysis, only those tokens misclassified by the DFA and misidentified by listeners are considered to evaluate the degree to which degraded vowel acoustics influence the resulting percept. This subset of the data is evaluated exclusively in an attempt to avoid introduction of lexical influence (of the target word) vowel perception. Thus, accurate perceptions of vowel identity despite token misclassifications are excluded from this analysis. Due to the nature of this analysis, the data are not treated statistically. Nevertheless, agreement between misclassification and perceptual errors may be interpreted as evidence suggesting degraded vowel acoustics are a component of the intelligibility disorder caused by dysarthria and not merely an index of severity.

A confusion matrix of misclassified to misperceived vowel tokens is found in Table 18. It is important to note that the classification results of the DFA are constrained, in that errors are limited to one of nine other vowels. However, the perceptual data were collected from an unconstrained transcription task, thus perceptual errors are not limited to the ten vowels studied here. Examples of other perceptual errors are diphthong or schwar substitutions or vowel omissions. To constrain the perceptual data in a similar manner as the acoustic data, other perceptual errors were excluded from the calculations of percent agreement between misclassified tokens and misperceptions. Greater than 10% agreement between misclassified tokens and misperceptions indicates an above chance-level agreement. Agreement percentages varied from 23 - 48% depending on the vowel.

Discussion

Vowel tokens embedded in strong syllables of phrases produced by dysarthric speakers were normalized and classified via DFA with approximately 65% accuracy. Listeners, benefitting from lexical and syntactic top-down information, identified the vowel tokens with 71% accuracy. Spectrally and temporally distinctive vowel tokens (i.e., tokens correctly classified via discriminant function analysis) were identified with significantly greater accuracy than misclassified tokens. This finding is strengthened by the results of the second analysis, which revealed that tokens identified with 100% accuracy were classified via DFA with nearly 20% greater accuracy than those tokens that presented perceptual challenges to listeners (perceived with 0-60% accuracy). Finally, an above-chance level agreement between the nature of misclassification and misperception errors was revealed for all vowels in the third analysis. The results of the three analyses provide compelling evidence in support of the view that degraded vowel acoustics are not merely an index of severity in dysarthria, but rather are an integral component of the resultant intelligibility disorder.

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 view has not been universally demonstrated (e.g., see Sapir et al., 2007; Weismer et al., 2001). In the first experiment, dysarthric speakers are reliably differentiated from non-disordered speakers by most vowel space metrics. VSA, the most commonly reported metric capturing vowel space compression, was considered in a subsequent post-hoc analysis that evaluated the effect of speaker group (non-disordered, ataxic, mixed flaccid-spastic, hyperkinetic and hypokinetic dysarthria) on VSA measurements. The effect of speaker group was significant [$F(4, 52) = 6.43$, $p < .0001$] and multiple comparisons revealed the VSAs associated with each of the dysarthrias were significantly compressed relative to non-disordered VSA; however no significant differences were revealed between the dysarthria subtypes. Similarly, most vowel metrics failed to demonstrate acoustic differences specifically associated with each dysarthria subtype.

These results support a taxonomical approach to studying the perceptual challenges associated with the dysarthrias suggested by Weismer and Kim (2010). This approach is motivated by the substantial overlap of perceptual characteristics associated with the dysarthria subtypes and the notion that characteristics of a given dysarthria vary with severity. The overarching goal of this approach is to identify a core set of deficits (i.e., perceptual similarities) common to most, if not all, speakers with dysarthria. Identification of such similarities would permit the detection of differences that reliably distinguish different types of motor speech disorders irrespective of etiology. Towards this end, Kim, Kent and Weismer (2011) used a variety of acoustic metrics, including VSA and F2 slope, to classify a large cohort of speakers with dysarthria arising from traumatic brain injury, stroke, multiple systems atrophy and Parkinson's disease according to 1) underlying medical etiology, 2) dysarthria diagnosis, and 3) severity of the speech disorder. The vowel metrics, VSA and F2 slope, demonstrated significant relationships with scaled severity ratings, and, as such, were included by the model constructed to classify speakers according to overall severity of their impairment. In line with the results presented here, the vowel space metrics failed to demonstrate utility in classifying dysarthric speakers according to their underlying medical etiology or speech diagnosis. Thus, the notion that vowel space compression represents a "perceptual similarity" uniting most, if not all, speakers with dysarthria, as suggested by Weismer and Kim, is supported by the results reported herein. Further investigation of the specific effects of severity of impairment on degradation of vowel acoustics is warranted.

A major 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 % words correct), not vowel identification accuracy, have been the perceptual units of interest. This practice has prevented causative interpretation of the findings. Specifically, conclusions implicating degraded vowel acoustics as contributory factors to the intelligibility disorder associated with dysarthria are premature due to the inability to rule out the possibility that degraded vowel acoustics are merely an index of overall severity of the disorder (Weismer et al., 2001). Thus, the perceptual consequences of degraded vowel acoustics was studied in the context of vowel identification accuracy, in addition to overall intelligibility (% words correct), in this investigation.

As revealed by the correlation and regression analyses, vowel space metrics that capture vowel centralization tendencies and reduced working vowel space (e.g., distance/dispersion metrics) demonstrated the strongest relationships with both vowel identification accuracy and intelligibility. Specifically, reduced working vowel space was associated with reduced vowel identification accuracy and intelligibility. In addition, metrics capturing reduced F2 slope excursion associated with dysarthric vowel production were also moderately related to overall intelligibility and vowel identification. These findings not only were demonstrated with established metrics, such as VSA and mean dispersion, but also were extended to recently introduced and novel metrics. In fact, many novel and recently introduced metrics demonstrate some of the strongest relationships with these perceptual outcome measures. One such metric, the formant centralization ratio (FCR), which is touted to minimize variability arising from inter-speaker differences while maximizing sensitivity to vowel centralization, has been demonstrated to differentiate between the vowel spaces produced by 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 the

FCR is related to both intelligibility and vowel identification accuracy. Corner vowel to /^/ dispersion, a novel metric capturing vowel centralization, also is correlated with both perceptual outcome measures (see Table 11). Non-redundant information is offered by this dispersion metric, despite being moderately correlated ($r = -.677$) with the FCR. The FCR considers only the formant information of three corner vowels. Construction of the FCR is highly dependent on the formant information associated with /u/ (represented twice in the numerator). As is evidenced in Figures 1 and 2, /u/ tokens are fairly disparate, particularly along the F2 dimension, and /a/ along the F1 dimension. It is possible that the instability of these tokens may be unduly inflating the FCR. This possibility warrants further investigation.

Kim et al. (2010) introduced a metric referred to as overlap degree that when compared to VSA and other vowel metrics accounted for the greatest amount of variability in intelligibility scores in 9 speakers with CP. As reported by Kim and her colleagues, overlap degree is simply the misclassification rate of vowel tokens (/i/, /ɪ/, /ɛ/, /a/, /ʊ/ and /u/), categorized via DFA for each speaker. In the larger and more diverse population of dysarthric speakers studied here, this metric failed to reach the values from the Kim study ($R^2 = .96$), but it was moderately correlated with intelligibility and vowel accuracy. The discrepancy is likely due to differences in perceptual task, stimuli, and subsets of vowels studied. Nevertheless, the results of the present investigation provide compelling evidence supporting the use of recently introduced and novel vowel metrics that capture centralization and vowel distinctiveness to study dysarthric vowel perception.

Based on the results of the present investigation, subsets of vowels metrics recommended to 1) detect acoustic consequences of dysarthric vowel production, 2) predict overall intelligibility (perhaps an index of severity), and 3) predict vowel identification accuracy are summarized in Table 19.

The results of Experiment 2 link degraded vowel acoustics to reduced perceptual outcome measures, including vowel identification accuracy. However, the direct implications of such degradations on the resulting percept are evaluated specifically in Experiment 3. Results of the first analysis revealed that tokens that are more distinctive (i.e., correctly classified via DFA) were better identified. The second analysis validated and extended these findings as 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). Thus, the results of the first two analyses suggest that distinctive vowel tokens are better identified and, likewise, better- identified tokens are more distinctive.

Finally, an above-chance level agreement between the nature of the misclassification and misidentification errors was demonstrated for all vowels. The level of agreement, however, was stronger for some vowels than for others. Specifically, misclassification-misidentification agreement was stronger for front vowels that vary along the tongue-height (F1) dimension. As revealed in Experiment 1, these vowels possess a tight articulatory working space, raising the propensity to elicit perceptual errors. Thus it follows that the acoustic features that led to misclassification of vowels in such a tight working space similarly guide perceptual errors.

While the relative potency of the segmental information offered by vowels to speech perception remains unclear, it is certain that accurate identification of vowels, and consonants alike, is a crucial component of models of word recognition (Luce & Pisoni, 1998; McClelland & Elman, 1986; Norris, 1994). Briefly, models of word recognition (e.g., Trace, Shortlist, and Neighborhood Activation Model) describe this process as occurring in two phases, *activation* and *competition* of lexical candidates. First, a pool of lexical candidates is *activated* in response to incoming acoustic-phonetic information. The activated lexical candidates subsequently *compete.* The candidate that most resembles the acoustic-phonetic input "wins" the competition (i.e., is perceived by the listener). Thus, poor production or misperception of the vowel /ɪ/ in the word *ship* results in a pool of activated lexical candidates that may not include the intended target, thereby, decreasing the likelihood that the word *ship* will win the subsequent lexical competition. The results of the present work are well accounted for by the conceptual framework provided by word recognition models, as the nature of acoustic degradations associated with non-distinctive vowel tokens (i.e., vowel tokens misclassified via DFA) played a role in guiding perception.

The effects of vowel misperception extend beyond that of word recognition, as information gleaned from vowels can be used to facilitate speech segmentation (Cutler & Buttlerfield, 1992; Cutler & Carter, 1987; Liss, Spitzer, Caviness & Adler, 2000, Mattys, Melhorne & White, 2005; Spitzer, Liss &

Mattys, 2007). Mattys, Melhorne and White (2005) describe a hierarchical model that specifies the use of linguistic, segmental and suprasegmental information in speech segmentation is dependent on the quality of the listening condition. In optimal listening conditions, listeners rely upon linguistic, specifically lexical, information to segment the speech stream. Thus, speech segmentation occurs as a consequence of word recognition. However, in suboptimal listening conditions, speech segmentation strategies adapt to incorporate segmental and suprasegmental information to facilitate deciphering of connected speech. Specifically, stress information contained in strong syllables (e.g., presence of unreduced vowel, increased duration and amplitude) has the potential to cue word onsets in English, as the first syllable in most English words is strong (Culter & Carter, 1987). Thus, distorted/degraded vowel production and/or hindered perception of information contained in vowels may have deleterious effects on overall speech perception resulting in decreased intelligibility of the speech signal. Investigation of the effects of degraded vowel acoustics of speech segmentation strategies was beyond the scope of the present investigation. However, future studies focusing of this aspect of dysarthric vowel perception are well motivated by the results presented herein linking vowel production and perception.

The clinical implications of the present work should not be minimized. By establishing the link between vowel production errors and the nature of perceptual errors, therapeutic interventions that aim to improve vowel production on the part of the speaker or vowel perception on the part of the listener should result in increases to vowel identification accuracy, and ultimately intelligibility. For example, reduced high-low vowel contrast (i.e. reduced distance or dispersion of front and/or back vowels) in a speaker with dysarthria will likely produce perceptual errors along the same dimension. Thus, a goal of speaker-directed therapy should be to increase spectral distinctiveness of neighboring vowel tokens along the affected dimension. In cases where speaker-directed therapy is not feasible, as is the case for many patients diagnosed with progressive neurodegenerative disorders, caregivers may undergo perceptual training aimed to retune their perceptual boundaries for specific vowels tokens to accommodate less distinctive vowel tokens. Benefits to intelligibility following therapy or perceptual training are predicted by the outcomes of this investigation.

Conclusions

Results of the present set of experiments contribute substantially to the growing body of literature in the area of dysarthric vowel perception. Not only are a variety of acoustic vowel space metrics (e.g., global, fine-grained, and distance/dispersion) considered with regard to their abilities to 1) differentiate dysarthric from non-disordered vowel production and 2) predict perceptual outcomes, but their contributions also are evaluated within the context of a broad cohort of dysarthric speakers. Equipped with fairly equivalent groups of speakers diagnosed with the various dysarthria subtypes, exploration of dysarthria-specific effects on vowel production (represented acoustically) was possible. Another significant contribution of the present study is that vowel identification accuracy, in addition to overall intelligibility (% words correct), was included as a perceptual outcome measure. Finally, results of this experiment directly inform the justifiably questionable nature of the relationship between degraded vowel production and the resulting percept in dysarthria.

References

Bradlow, A., & Bent, T. (2002). The clear speech effect for non-native listeners. *Journal of the Acoustical Society of America, 112*(1), 272-284.

Bradlow, A., Torretta, G.M. & Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication, 20*, 255-272.

Boersma, P. & Weenink, D. (2006). Praat: doing phonetics by computer (Version 4.4.24) [Computer program]. Retrieved June 19, 2006, from http://www.praat.org/

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.

Cole, R., Yan, Y., Mak, B., Fanty, M., & Bailey, T. (1996). “The contribution of consonants versus vowels to word recognition in fluent speech,” in Proceedings of the ICASSP’96, pp. 853–856.

Cutler, A. & Butterfield, S. (1992). Rhythmic cues to speech segmentation: evidence from juncture misperception. *Journal of Memory and Language*, *31*, 218-236.

Cutler, A. & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language*, *2,* 133-142.

Darley, F., Aronson, A., & Brown, J. (1969). Differential diagnostic patterns of dysarthria*. Journal of Speech and Hearing Research, 12*, 246–269.

Darley, F., Aronson, A., & Brown, J. (1975). Motor Speech Disorders. Philadelphia: W. B. Saunders Inc.

Duffy, J. R. (2005). *Motor speech disorders: Substrates, differential diagnosis, and managemen*t (2nd Ed.) St. Louis, MO: Elsevier Mosby.

Ferguson, S., & Kewley-Port, D. (2002). Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America, 112*(1), 259-271.

Fogerty, D. & Kewley, Port, D. (2009). Perceptual contributions of the consonant-vowel boundary to sentence intelligibility. *Journal of the Acoustical Society of America, 126*(2), 847-857. doi: 10.1121/1.3159302.

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. *Journal of the Acoustical Society of America*, ***97*,** 3099–31.

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.

Kewley-Port, D., Burkle, T. Z., & Lee, J. H. (2007). Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners. *J. Acoust. Soc. Am.* 122, 2365–2375. doi: 10.1121/1.2773986.

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., Kent, R.D., and Weismer, G. (2011). An acoustic study of the relationships among neurologic disease, dysarthria type and severity of dysarthria. *Journal of Speech, Language, and Hearing Research, 54*, 417-429.

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*(6), 329-335.

Liss, J.M., White, L., Mattys, S.L., Lansford, K., Spitzer, S, Lotto, A.J., and Caviness, J.N. (2009). Quantifying speech rhythm deficits in the dysarthrias. *Journal of Speech, Language, and Hearing Research, 52*(5), 1334-1352.

Liss, J. M., Spitzer, S. M., Caviness, J. N., & Adler, C. (2000). LBE analysis in hypokinetic and ataxic dysarthria*. Journal of the Acoustical Society of America, 107*, 3415–3424.

Liu, H.M., Tsao, F.M., and 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*(6), 3879–3889.

Luce, P.A. & Pisoni, D.B. (1998) Recognizing spoken words: the neighborhood activation model. *Ear and Hearing*, *19*, 1–36.

Mattys, S. L., White, L., & Melhorn, J. F (2005). Integration of multiple segmentation cues: A hierarchical framework*, Journal of Experimental Psychology General, 134*, 477–500.

McClelland, J., & Elman, J. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18,* 1-86.

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.

Milenkovic, P.H. (2004). TF32 [Computer software]. Madison: University of Wisconsin, Department of Electrical and Computer Engineering.

Moon, S. Y., & Lindblom, B. (1994). Interaction between duration, context, and speaking style in English stressed vowels. *Journal of the Acoustical Society of America, 96*, 40-55.

Nearey, T.M. (1989). Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America,* 85 (5), 2088-2112.

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

Norris, D. (1994) Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52***,** 189–234.

Owren, M. J., & Cardillo, G. C. (2006). The relative roles of vowels and consonants in discriminating talker identity versus word meaning. *Journal of the Acoustical Society of America, 119*, 1727–1739. doi: 10.1121/1.2161431

Payton, K., Uchanshki, R., & Braida, L. (1994). Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing. *Journal of the Acoustical Society of America, 95*(3), 1581-1592.

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

Peterson, G.E. & Lehiste, I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America, 32, 693-703.*

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.

Picheny, M., Durlach, N., & Braida, L. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research, 29*, 434-446.

Rosen, K.M, Goozee, J.V., & Murdoch, B.E. (2008). Examining the effects of Multiple Sclerosis on speech production: Does phonetic structure matter?. *Journal of Communication Disorders, 41*, 49-69.

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.

Skodda, S., Visser W., & Schlegel, U. (2011). Vowel articulation in Parkinson's disease. *Journal of Voice, 25*(4), 467-472. doi: 10.1016/j.voice.2010.01.009

Spitzer, S., Liss, J.M., & Mattys, S.L. (2007). Acoustic cues to lexical segmentation: A study of resynthesized speech. *Journal of the Acoustical Society of America, 122*(6), 3678- 3687. doi: 10.1121/1.2801545

Strange, W. (1989a). Dynamic specification of coarticulated vowels spoken in sentence context. *Journal of the Acoustical Society of America, 85 (5), 2135-2153.*

Strange, W. (1989b). Evolving theories of vowel perception. *Journal of the Acoustical Society of America, 85*(5), 2081-2087.

Tjaden, K., and 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.

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. & Kim, Y-J. (2010). Classification and taxonomy of motor speech disorders: What are the issues? In B. Maassen and P.H.H.M. van Lieshout (Eds.), *Speech Motor Control: New developments in basic and applied research* (pp. 229-241). Oxford University Press.

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: John Benjamin.

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

Yunusova, Y., Weismer, G., Kent, R. D., & Rusche, N. M. (2005). Breath-group intelligibility in dysarthria: Characteristics and underlying correlates. *Journal of Speech, Language, & Hearing Research, 48,* 1294-1310.

Table 1

Dysarthric speaker demographic information per stimulus set

Set	Speakers	Sex	Age	Medical Etiology	Severity of Speech Disorder
1	ALSF2	F	75	ALS	Severe
	ALSF8	F	63	ALS	Moderate
	ALSM1	M	56	ALS	Moderate
	ALSM5	M	50	ALS	Mild
	ALSM7	M	60	ALS	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
	HDF4	F	67	Huntington's disease	Severe
	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 disease	Mild
	PDF7	F	58	Parkinson disease	Moderate
	PDF9	F	71	Parkinson disease	Mild
	PDM8	M	77	Parkinson disease	Moderate
	PDM9	M	76	Parkinson disease	Moderate
	PDM15	M	57	Parkinson disease	Moderate
2	ALSF5	F	73	ALS	Severe
	ALSF7	F	54	ALS	Moderate
	ALSF9	F	86	ALS	Severe
	ALSM3	M	41	ALS	Mild
	ALSM4	M	64	ALS	Moderate
	ALSM8	M	46	ALS	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

Set	Speakers	Sex	Age	Medical Etiology	Severity of Speech Disorder
	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 disease	Mild
	PDF5	F	54	Parkinson disease	Moderate
	PDF6	F	65	Parkinson disease	Mild
	PDM1	M	69	Parkinson disease	Severe
	PDM10	M	80	Parkinson disease	Moderate
	PDM12	M	66	Parkinson disease	Severe

Note. ALS = amyotrophic lateral sclerosis.

Table 2

Derived vowel metrics

Type	Vowel Metric	Description
Global	Mean F0	Mean F0 of the entire vowel set, derived by averaging the midpoint measurements (in Hz) across the ten vowels.
	Mean F1	Mean F1 of the entire vowel set, derived by averaging the midpoint measurements (in Hz) across the ten vowels.
	Mean F2	Mean F2 of the entire vowel set, derived by averaging the midpoint measurements (in Hz) across the ten vowels.
	Mean dur	Mean vowel duration of the entire vowel set, derived by averaging vowel durations across the ten vowels.
Fine-grained	F0 range	F0 range was calculated by subtracting the lowest f0 (Hz) value across the 10 vowels from the highest value.
	F1 range	F1 range was calculated by subtracting the lowest F1 (Hz) value across the 10 vowels from the highest value.
	F2 range	F2 range was calculated by subtracting the lowest F2 (Hz) value across the 10 vowels from the highest value.
	VSA	Vowel space area. Heron’s formula was used to calculate the area of the irregular quadrilateral formed by the corner vowels in F1 X F2 space.
	Mean disp	This metric captures the overall dispersion (or distance) of each pair of the ten vowels, as indexed by the Euclidean distance between each pair in the F1 X F2 space.
	Dyn ratio	Mean EDs from vowel onset to midpoint to offset in F1 × F2 space for each vowel were averaged. The average EDs of the most dynamic (æ, ^, ʊ) was divided by the average EDs of the least dynamic (i, ε, u) vowels. Larger values are interpreted to reflect greater distinctiveness in vowels with dynamic and static trajectories.
	Dur ratio	Ratio of longest (a, o, e, æ) to shortest vowels (ɪ, ʊ, ε, ^). The average value of the longest vowels was divided by the average value of the shortest vowels. Larger values are interpreted to reflect

Type	Vowel Metric	Description
		greater distinctiveness in vowel length.
Alternative	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.
Distance/ dispersion	ED /i/ - /æ/	Euclidean distance in F1 X F2 space from /i/ to /æ/ (front vowels)
	ED /u/ - /a/	Euclidean distance in F1 X F2 space from /u/ to /a/ (back vowels)
	ED /i/ - /u/	Euclidean distance in F1 X F2 space from /i/ to /u/ (high vowels)
	ED /æ/ - /a/	Euclidean distance in F1 X F2 space from /a/ to / æ / (low vowels)
	Front disp	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 X F2 space.
	Back disp	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 backvowels in F1 X F2 space
	Corner disp	This metric is expressed by the average Euclidean distance of each of the corner vowels, /i/, /æ/, /a/, and /u/, to the center vowel /^/.
	Global disp	Mean dispersion of all vowels to the global formant means (ED in F1 X F2 space).
	Neighbor disp	Average Euclidean distance of the following spectral neighbors were used to compute this dispersion metric: (/i/- /e/, /e/- /ɪ/, /ɪ/-/ɛ/, /ɛ/-/æ/, /æ/-/a/, /a/-/o/, /o/-/ʊ/, /ʊ/-/u/, and /u/-/i/)
	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 (ED in F1 X F2 space) from vowel onset to midpoint to offset.
F2 slope metrics	Mean F2 slope	The absolute values of the F2 slopes from vowel onset to offset were averaged across the entire

Type	Vowel Metric	Description
		vowel set.
	Dynamic F2 slope	The absolute values of F2 slopes associated with the most dynamic vowels (æ, ^, ʊ) were averaged.

Note. ED = Euclidean distance

Table 3

Non-disordered and dysarthric group means

	Vowel Metric	Group	*n*	*M*	*SD*
Global	Mean F0	ND	12	150.84	33.47
		D	45	160.30	36.54
	Mean F1	ND	12	532.04	50.25
		D	45	528.21	75.35
	Mean F2	ND	12	1705.82	125.78
		D	45	1630.20	189.84
	Mean dur	ND	12	87.93	11.66
		D	45	150.33	54.03
Fine-grained	VSA	ND	12	286213.07	71217.41
		D	45	174822.17	66928.04
	Mean disp	ND	12	400.54	69.31
		D	45	330.46	64.76
	Range F0	ND	12	43.35	25.67
		D	45	53.45	47.27
	Range F1	ND	12	468.79	62.66
		D	45	362.53	80.46
	Range F2	ND	12	1396.65	225.27
		D	45	1145.49	229.20
	Dyn ratio	ND	12	1.41	0.51
		D	45	1.45	0.36
	Dur ratio	ND	12	1.43	0.09
		D	45	1.31	0.17
Alternate	FCR	ND	12	1.07	0.05
		D	45	1.19	0.12
Dispersion/ Distance	ED /i/ - /ae/	ND	12	851.07	118.43
		D	45	591.63	179.12
	ED /i/ - /u/	ND	12	906.64	142.18
		D	45	848.76	264.97
	ED /u/ - /a/	ND	12	576.08	105.59
		D	45	364.43	97.78
	ED /æ/ - /a/	ND	12	563.50	185.73
		D	45	460.26	165.26
	Front disp	ND	12	503.32	83.38
		D	45	345.65	89.34
	Back disp	ND	12	368.45	75.32
		D	45	276.13	71.86
	Corner disp	ND	12	563.45	120.48

	Vowel Metric	Group	*n*	*M*	*SD*
		D	45	432.14	93.89
	Global disp	ND	12	597.56	101.37
		D	45	484.11	90.76
	Neighbor disp	ND	12	350.44	72.38
		D	45	279.39	57.61
	Spectral overlap	ND	12	0.38	0.11
		D	45	0.56	0.13
F2 slope metrics	Mean F2 slope	ND	12	2.08	0.29
		D	45	1.55	0.61
	Dynamic F2 slope	ND	12	3.21	0.70
		D	45	2.32	0.99

Note. ND = non-disordered; D = dysarthric.

Table 4

Independent samples t-test results comparing the acoustic metrics derived from dysarthric and non-disordered speakers

	Vowel Metric	*t*	*df*	*p*
Global	Mean F0	-.810	55	.421
	Mean F1	.166	55	.869
	Mean F2	1.301	55	.199
	Mean dur*	-7.147	54.110	.000
Fine-grained	VSA	5.056	55	.000
	Mean disp	3.283	55	.002
	Range F0	-.710	55	.481
	Range F1	4.235	55	.000
	Range F2	3.384	55	.001
	Dyn ratio*	-.258	14.008	.800
	Dur ratio*	2.299	55	.025
		3.344	37.368	.002
Alternative	FCR	-5.098	43.981	.000
Dispersion/ distance	ED /i/ - /ae/	4.733	55	.000
	ED /i/ - /u/	.726	55	.471
	ED /u/ - /a/	6.555	55	.000
	ED /æ/ - /a/	1.874	55	.066
	Front disp	5.503	55	.000
	Back disp	3.916	55	.000
	Corner disp	4.051	55	.000
	Global disp	3.756	55	.000
	Neigh disp	3.594	55	.001
	Spectral overlap	-4.559	55	.000
F2 Slope metrics	Mean F2 slope	4.271	39.742	.000
	Dyn slope	2.927	55	.005

Note. *denotes equality of variance is not assumed.

Table 5

Results of one-way analysis of variance (ANOVA) testing equality of means for dysarthria subtypes.

	Vowel Metric	*F*(3, 41)	*p*
Global	Mean F0	1.063	.375
	Mean F1	2.238	.098
	Mean F2	.731	.539
	Mean dur.	16.443	.000*
Fine-grained	Range F0	.337	.798
	Range F1	1.018	.395
	Range F2	1.388	.260
	VSA	.358	.783
	Mean disp.	.436	.728
	Dyn. ratio	1.605	.203
	Dur. ratio	.817	.492
Alternative	FCR	.672	.574
Dispersion/distance	ED /i/ - /æ/	1.706	.181
	ED /u/ - /i/	.778	.513
	ED /u/ - /a/	.453	.716
	ED /a/ - /æ/	.637	.595
	Neighbor disp.	1.243	.306
	Corner disp.	.974	.414
	Front disp.	1.634	.196
	Back disp.	.614	.610
	Global disp.	.669	.576
	Spectral overlap	1.239	.308
F2 slope	Mean F2 slope	14.327	.000*
	Dynamic F2 slope	12.270	.000*

*denotes significant between group differences

Table 6

Group means of significant variables

					95% CI	
		n	*M*	*SD*	*LL*	*UL*
Mean Duration	Ataxic	12	163.64	26.65	146.71	180.58
	ALS	11	206.47	57.66	167.73	245.21
	HD	10	132.92	39.61	104.58	161.26
	PD	12	100.06	16.85	89.36	110.76
	Total	45	150.33	54.03	134.10	166.56
Average F2 slope	Ataxic	12	1.32	0.34	1.10	1.54
	ALS	11	1.01	0.45	0.71	1.31
	HD	10	1.70	0.32	1.47	1.93
	PD	12	2.16	0.59	1.78	2.54
	Total	45	1.55	0.61	1.37	1.74
Dynamic F2 slope	Ataxic	12	1.90	0.80	1.40	2.41
	ALS	11	1.51	0.81	0.97	2.05
	HD	10	2.59	0.32	2.37	2.82
	PD	12	3.25	0.87	2.69	3.81
	Total	45	2.32	0.99	2.02	2.62

Note. CI = confidence interval; LL = lower limit, UL = upper limit.

Table 7

Classification summary by dysarthria-subtype

		Predicted Group Membership				
	Group	Ataxic	ALS	HD	PD	Total
Count	Ataxic	5	3	4	0	12
	ALS	5	6	0	0	11
	HD	0	1	6	3	10
	PD	0	0	1	11	12
%	Ataxic	41.7	25.0	33.3	.0	100.0
	ALS	45.5	54.5	.0	.0	100.0
	HD	.0	10.0	60.0	30.0	100.0
	PD	.0	.0	8.3	91.7	100.0

Note. 62.2% of originally grouped speakers were correctly classified (same upon cross-validation).

Table 8

Proportion of words and vowels correct per speaker

Group	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

Group	Speaker	Words correct	Vowel accuracy
	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)

Table 9

Pearson correlations between perceptual outcome measures and global vowel space metrics

	Mean F0	Mean F1	Mean F2	Mean Dur
Intelligibility	-.039	-.161	.084	-.225
Vowel Accuracy	-.045	-.235	.116	-.318*

* $p < 0.05$.

Table 10

Pearson correlations between perceptual outcome measures and fine-grained vowel space metrics

	VSA	Disp Mean	Range F0	Range F1	Range F2	Dynamic Ratio	Duration Ratio
Intelligibility	.401**	.317*	.059	.306*	.310*	.106	.239
VA	.412**	.364*	.096	.275	.395**	.149	.260

Note. VA = vowel accuracy.

* $p < .05$

** $p < .001$

Table 11

Pearson correlations between perceptual outcome measures and FCR, dispersion and F2 slope metrics

	Vowel Space Metric	Intelligibility	VA
Alternate	FCR	$-.442^{**}$	$-.526^{**}$
Dispersion/Distance	ED /i/ - /ae/	.246	$.318^{*}$
	ED /u/ - /i/	.234	$.333^{*}$
	ED /u/ - /a/	$.323^{*}$	.264
	ED /a/ - /ae/	.292	.226
	Front Disp	.237	$.308^{*}$
	Back Disp	.204	.218
	Corner Disp	$.458^{**}$	$.447^{**}$
	Global Disp	$.335^{*}$	$.392^{**}$
	Neighbor Disp	.218	.246
	Spectral overlap	$-.415^{*}$	$-.421^{**}$
F2 Slope	Mean F2 slope	$.401^{**}$	$.461^{**}$
	Dynamic F2 slope	$.422^{**}$	$.478^{**}$

Note. VA = vowel accuracy.
* $p < 0.05$.
** $p < 0.01$.

Table 12

Results of stepwise regressions in which the acoustic variables predict overall intelligibility in all, female and male speakers

Regression	Variable Entered	Beta	t	p
All Speakers	Corner Disp	.433	3.381	.002
	Mean F1	-.339	-2.757	.009
	Spectral overlap	-.322	-2.733	.009
	Mean F2 slope	.249	2.079	.044
Female speakers	Dynamic slope	.579	5.041	.000
	Corner Disp	.378	3.320	.004
	Spectral overlap	-.319	-2.879	.010
Male speakers	Corner Disp	.468	2.425	.024

Table 13

Results of stepwise regressions in which the acoustic variables predict vowel accuracy in all, female and male speakers

Regression	Variable Entered	Beta	t	p
All Speakers	FCR	-.791	-4.599	.000
	Mean F2 slope	.584	4.406	.000
	F2 range	-.446	-2.319	.025
Female speakers	Dynamic slope	.441	3.915	.001
	Corner Disp	.329	3.087	.007
	Spectral overlap	-.463	-3.964	.001
	Front Disp	.331	2.679	.016
Male speakers	FCR	-1.169	-4.034	.001
	VSA	-.756	-2.608	.017
	Mean F2 slope	.337	2.215	.039

Table 14

Confusion matrix of correctly identified vowels tokens and perceptual errors

		Perceived vowel								
		i	ɪ	e	ɛ	æ	a	o	u	^
Target vowel (Count)	i	663	54	29	24	5	3	4	3	10
	ɪ	23	590	53	62	23	4	8	15	15
	e	22	39	716	23	9	4	3		9
	ɛ		50	5	667	28	8	2	3	38
	æ	1	25	19	136	581	15	2	2	17
	a		6	1	10	23	653	11	2	52
	o	5	10	3	5	10	45	623	27	52
	u	20	27	7	24	3	11	43	556	22
	^	1	10	2	29	18	41	27	4	657
Target vowel (%)	i	74	6	3	3	1				1
	ɪ	3	66	6	7	3		1	2	2
	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
	u	2	3	1	3		1	5	62	2
	^		1		3	2	5	3		73

Note. Vowel tokens were perceived with 71% accuracy.

Table 15

Classification summary of all vowel tokens

		Predicted Vowel									
		i	ɪ	e	ɛ	æ	a	o	u	^	ʊ
Count	i	158	3	11	3				3		2
	ɪ	11	97	16	25	3		5	11	9	2
	e	17	14	142	2	4					1
	ɛ		42	1	88	28	1	1	4	14	1
	æ		11	5	31	113	12	4	1	3	
	a		1		2	12	124	15	4	17	5
	o		1			1	8	126	11	27	6
	u	14	22	2	2			27	103	2	6
	^		12		11	8	24	28	3	78	14
	ʊ		5	4			2	7	6	1	109
%	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
	^		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).

Table 16

Classification summary of well-identified vowel tokens

		Predicted Vowel									
	Vowel	i	ɪ	e	ɛ	æ	a	o	u	^	ʊ
Count	i	89		4							
	ɪ		53	1	10			1	1	5	
	e	7	4	96	3						
	ɛ		21	1	40	19	1	1		6	
	æ		5	2	12	46	1			4	
	a					4	62	10	2	8	1
	o		1				5	52	2	10	
	u	7	10					6	47	1	1
	^		5		4	7	10	8	1	49	6
	ʊ		1	1				1			13
%	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
	^		6		4	8	11	9	1	54	7
	ʊ		6	6				6			81

Note. 71.2% of originally grouped vowels were correctly classified (69% upon cross-validation).

Table 17

DFA classification results of poorly perceived tokens

		Predicted Vowel									
	Vowel	i	ɪ	e	ɛ	æ	a	o	u	^	ʊ
Count	i	45	4	6	2	1			1		1
	ɪ	10	28	13	10	6		3	9	2	2
	e	8	3	32	2	2					1
	ɛ		19		27	11		2	2	1	
	æ		4	1	14	48	10	3	1	3	
	a				2	1	52	6	2	6	
	o		1			1	3	48	8	6	6
	u	5	12	2	2			13	41		6
	^		3		3	2	10	12	3	22	5
	ʊ	1	1			1			3		12
%	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
	^		5		5	3	17	20	5	37	8
	ʊ	6	6			6			17		67

Note. 55.6% of originally grouped vowels were correctly classified (51.6% upon cross-validation).

Table 18

Misclassified to misidentified vowel agreement

		Identification Error									
		i	ɪ	e	ɛ	æ	a	o	u	^	ʊ
Class. error (count)	i	17	9	12	3	1		3	2	6	
	ɪ	11	46	9	6	7	7	1	5	17	1
	e	4	6	28	13	4	2		1	2	
	ɛ	1	15	13	54	10	7		3	6	3
	æ	1	4	1	5	18	3	1		13	2
	a	1	3	1	18		12	3		6	
	o	1	4	1	5		16	43		9	32
	u	5	4		6	3	4	4	14	18	2
	^		9	2	7	5	8	4	3	22	4
	ʊ	3	5	1	7	2	8	6	3	3	12
Class. error (%)	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
	^		14	3	11	8	13	6	5	34	6
	ʊ	6	10	2	14	4	16	12	6	6	24

Note. Classification error percentages were derived by dividing the counts by the total excluding other errors

Table 19

Vowel metrics recommended for the study of dysarthric vowel production and perception

Analysis type	Speakers	Recommended vowel metrics	Results
DFA	Non-disordered vs. dysarthric	ED /i/-/æ/, ED /u/-/a/, spectral overlap, mean duration, and average F2 slope	96.5% classification accuracy
Regression (Intell)	All dysarthric speakers	Corner disp, mean F1, spectral overlap, average F2 slope	Adjusted R^2 = .423**
	Female	Dynamic F2 slope, corner disp, and spectral overlap	Adjusted R^2 = .749**
	Male	Corner disp	Adjusted R^2 = .182*
Regression (VA)	All dysarthric speakers	FCR, mean F2 slope, and F2 range	Adjusted R^2 = .473**
	Female	Dynamic F2 slope, corner disp, spectral overlap, and front disp	Adjusted R^2 = .794**
	Male	FCR, VSA, and mean F2 slope	Adjusted R^2 = .495**

* $p < .05$
**$p < .001$

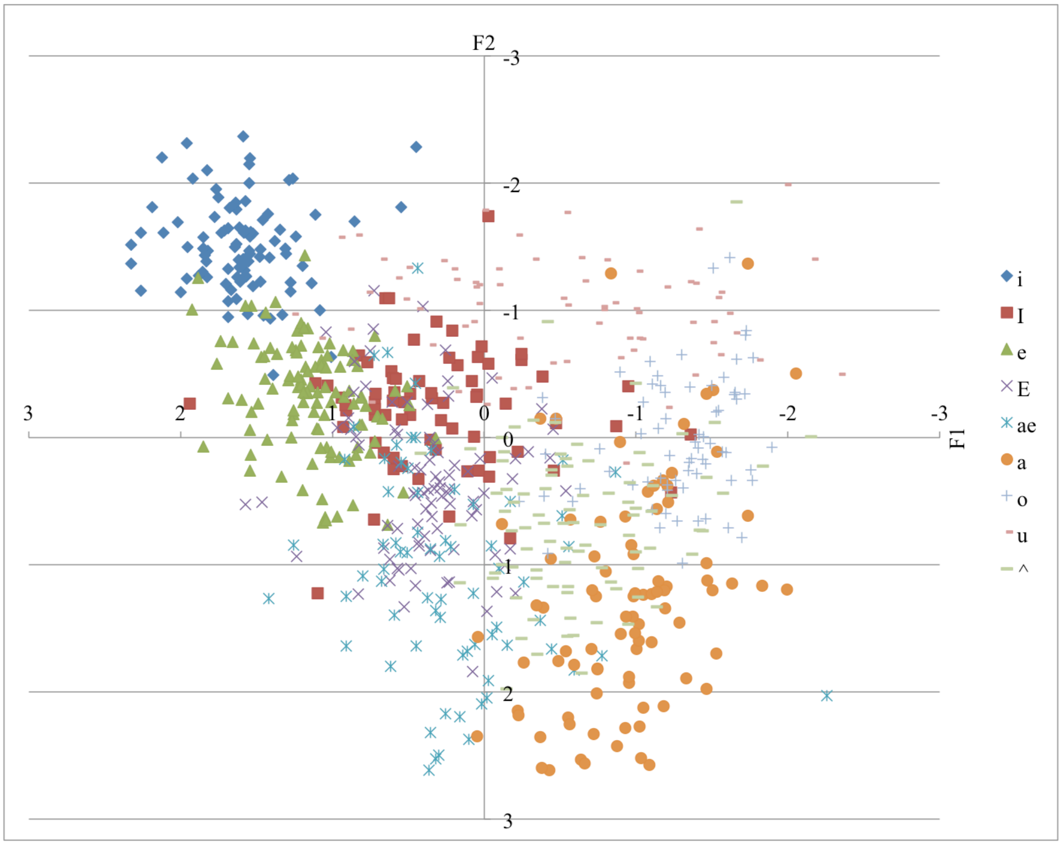

Figure 1. Normalized (Labonov's method) dysarthric vowel tokens, identified with 100% accuracy, represented in F1 x F2 perceptual space.

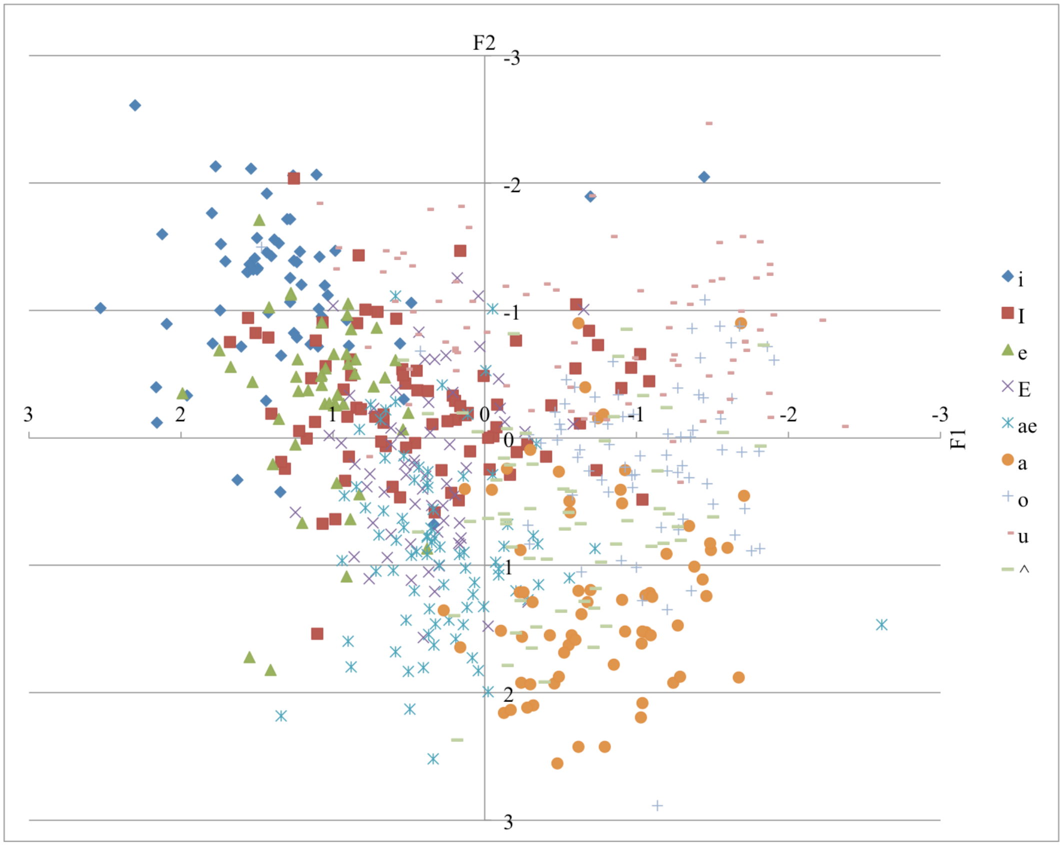

Figure 2. Normalized (Labonov's method) dysarthric vowel tokens, identified with 0-60% accuracy, represented in F1 x F2 perceptual space.

APPENDIX A

STIMULUS SETS

Set 1	*Set 1*
account for who could knock	admit the gear beyond
balance clamp and bottle	assume to catch control
beside a sunken bat	attend the trend success
commit such used advice	butcher in the middle
constant willing walker	confused but roared again
embark or take her sheet	cool the jar in private
listen final station	done with finest handle
may the same pursued it	had eaten junk and train
mode campaign for budget	indeed a tax ascent
narrow seated member	kick a tad above them
her owners arm the phone	mate denotes a judgment
pooling pill or cattle	mistake delight for heat
push her equal culture	model sad and local
rode the lamp for teasing	rampant boasting captain
or spent sincere aside	remove and name for stake
technique but sent result	rocking modern poster
transcend almost betrayed	support with dock and cheer
unseen machines agree	vital seats with wonder

APPENDIX B

INTERCORRELATIONS OF DYSARTHRIC ACOUSTIC AND PERCEPTUAL VOWEL METRICS

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Intell													
2. VA	.950**												
3. Mean F0	-.039	-.045											
4. Mean F1	-.161	-.235	.373*										
5. Mean F2	.084	.116	.227	.472**									
6. Mean Dur	-.225	-.318*	-.142	.251	-.281								
7. Range F0	.059	.096	.184	.263	.287	.049							
8. Range F1	.306*	.275	.193	.471**	.379*	-.132	.155						
9. Range F2	.310*	.395**	.340*	.250	.431**	-.284	.227	.446**					
10. VSA	.401**	.412**	.213	.377*	.394**	-.116	.230	.829**	.643**				
11. Mean Disp	.317*	.364*	.377*	.269	.427**	-.110	.142	.502**	.858**	.707**			
12. Dyn Ratio	.106	.149	-.037	-.129	-.121	-.282	-.232	.094	-.039	-.023	-.137		
13. Dur Ratio	.239	.260	-.275	-.127	-.251	.013	-.205	.185	.096	.151	.053	.341*	

* $p < .05$

** $p < .01$

	1	2	3	4	5	6	7	8	9	10	11	12	13
14. FCR	-.442**	-.526**	-.097	-.047	.004	-.026	-.053	-.479**	-.727**	-.658**	-.777**	-.033	-.293
15. ED i - æ	.246	.318*	.091	.315*	.385**	-.087	.213	.346*	.519**	.352*	.455**	.075	.077
16. ED u - a	.323*	.264	.086	.385**	.406**	-.170	.161	.627**	.322*	.575**	.267	-.070	.153
17. ED u - i	.234	.333*	.188	.228	.279	.057	.122	.348*	.793**	.572**	.858**	-.041	.179
18. ED a - æ	.292	.226	.303*	.019	.215	-.166	.060	.338*	.369*	.478**	.393**	-.143	-.054
19. Front Disp	.237	.308*	-.001	.234	.339*	-.220	.235	.441**	.599**	.434**	.458**	.062	.148
20. Back Disp	.204	.218	.163	.348*	.383**	-.318*	.175	.588**	.462**	.507**	.289	.094	.165
21. Corner Disp	.458**	.447**	.320*	.322*	.256	-.167	.108	.623**	.671**	.731**	.718**	.038	.232
22. Global Disp	.335*	.392**	.348*	.292	.447**	-.165	.177	.555**	.913**	.744**	.986**	-.108	.087
23. Neighb Disp	.218	.246	.150	.210	.350*	-.359*	.246	.572**	.701**	.587**	.482**	.014	.115
24. Overlap	-.379*	-.405**	.027	-.157	-.056	-.211	-.155	-.303*	-.256	-.453**	-.245	.128	-.112
25. Mean F2 Slope	.401**	.461**	.105	-.050	.510**	-.754**	.203	.251	.454**	.319*	.282	.032	.089
26. Dyn Slope	.422**	.478**	.260	.023	.549**	-.671**	.228	.329*	.530**	.446**	.366*	.190	.178

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

	1	2	3	4	5	6	7	8	9	10	11	12	13
14. FCR													
15. ED i - æ	-.438**												
16. ED u - a	-.192	.078											
17. ED u - i	-.846**	.512**	.034										
18. ED a - æ	-.269	-.299*	.565**	.101									
19. Front Disp	-.473**	.846**	.228	.504**	-.141								
20. Back Disp	-.155	.119	.633**	.103	.282	.221							
21. Corner Disp	-.660**	.377*	.532**	.574**	.501**	.508**	.388**						
22. Global Disp	-.794**	.511**	.339*	.857**	.385**	.554**	.380**	.760**					
23. Neighb Disp	-.400**	.282	.528**	.372*	.391**	.651**	.659**	.563**	.591**				
24. Overlap	.437**	-.130	-.256	-.288	-.269	.002	-.167	-.215	-.253	-.031			
25. Mean F2 Slope	-.100	.337*	.275	.140	.215	.402**	.382**	.260	.347*	.472**	-.069		
26. Dyn Slope	-.191	.338*	.336*	.240	.289	.380**	.479**	.360*	.428**	.508**	-.165	.904**	

* $p < .05$

** $p < .01$

APPENDIX C

IRB APPROVAL

ASU Knowledge Enterprise Development

Office of Research Integrity and Assurance

To:	Julie Liss COOR
From:	for Mark Roosa, Chair Soc Beh IRB
Date:	12/10/2010
Committee Action:	**Renewal**
Renewal Date:	12/10/2010
Review Type:	Expedited F7
IRB Protocol #:	0310001421
Study Title:	PERCEPTION OF DYSARTHRIC SPEECH
Expiration Date:	12/09/2011

The above-referenced protocol was given renewed approval following Expedited Review by the Institutional Review Board.

It is the Principal Investigator's responsibility to obtain review and continued approval of ongoing research before the expiration noted above. Please allow sufficient time for reapproval. Research activity of any sort may not continue beyond the expiration date without committee approval. Failure to receive approval for continuation before the expiration date will result in the automatic suspension of the approval of this protocol on the expiration date. Information collected following suspension is unapproved research and cannot be reported or published as research data. If you do not wish continued approval, please notify the Committee of the study termination.

This approval by the Soc Beh IRB does not replace or supersede any departmental or oversight committee review that may be required by institutional policy.

Adverse Reactions: If any untoward incidents or severe reactions should develop as a result of this study, you are required to notify the Soc Beh IRB immediately. If necessary a member of the IRB will be assigned to look into the matter. If the problem is serious, approval may be withdrawn pending IRB review.

Amendments: If you wish to change any aspect of this study, such as the procedures, the consent forms, or the investigators, please communicate your requested changes to the Soc Beh IRB. The new procedure is not to be initiated until the IRB approval has been given.