This is the *Accepted Manuscript* of an article published by the American Speech-Language-Hearing Association (ASHA) in the *Journal of Speech, Language, and Hearing Research, 64*, 4736-4753 © 2021. The manuscript is reprinted here with permission from ASHA and is further available online https://doi.org/10.1044/2021_JSLHR-21-00201

Validation of an Acoustic-Based Framework of Speech Motor Control: Assessing Criterion and Construct Validity using Kinematic and Perceptual Measures

Rowe, Hannah P., M.A., CCC-SLP[1], hrowe@mghihp.edu

Stipancic, Kaila L., Ph.D., CCC-SLP[1,2], klstip@buffalo.edu

Lammert, Adam C., Ph.D.[3], alammert@wpi.edu

Green, Jordan R., Ph.D., CCC-SLP[1], jgreen2@mghihp.edu

[1]*MGH Institute of Health Professions, Boston, MA, USA*

[2]*Department of Communicative Disorders and Sciences, The State University of New York at Buffalo, Buffalo, NY, USA*

[3]*Department of Biomedical Engineering, Worcester Polytechnic Institute, Worcester, MA, USA*

Corresponding Author:
Jordan Green
jgreen2@mghihp.edu
Department of Rehabilitation Sciences
Massachusetts General Hospital Institute of Health Professions
Charlestown, MA, United States

ABSTRACT

Purpose: This study investigated the criterion (analytical and clinical) and construct (divergent) validity of a novel, acoustic-based framework comprised of five key components of motor control: *Coordination*, *Consistency*, *Speed*, *Precision*, and *Rate*.

Method: Acoustic and kinematic analyses were performed on audio recordings from 22 subjects with amyotrophic lateral sclerosis during a sequential motion rate task. Perceptual analyses were completed by two licensed speech-language pathologists, who rated each subject's speech on the five framework components and their overall severity. Analytical and clinical validity were assessed by comparing performance on the acoustic features to their kinematic correlates and to clinician ratings of the five components, respectively. Divergent validity of the acoustic-based framework was then assessed by comparing performance on each pair of acoustic features to determine whether the features represent distinct articulatory constructs. Bivariate correlations and partial correlations with severity as a covariate were conducted for each comparison.

Results: Results revealed moderate to strong analytical validity for every acoustic feature, both with and without controlling for severity, and moderate to strong clinical validity for all acoustic features except *Coordination* without controlling for severity. When severity was included as a covariate, the strong associations for *Speed* and *Precision* became weak. Divergent validity was supported by the existence of multiple distinct articulatory constructs in the framework, as evidenced by weak to moderate pairwise associations between all acoustic features except *Speed (F2Slope)* and *Precision (ConVar.F2Slope)*.

Conclusions: This study demonstrated that the acoustic-based framework has potential as an objective, valid, and clinically useful tool for profiling articulatory deficits in individuals with speech motor disorders. The findings also suggest that compared to clinician ratings, instrumental measures are more sensitive to subtle differences in articulatory function. With further research, this framework could provide more accurate and reliable characterizations of articulatory impairment, which may eventually increase clinical confidence in diagnosis and treatment of patients with different articulatory phenotypes.

INTRODUCTION

Despite the growing number of diagnostic applications for speech testing (Parkinson's disease [PD], Hlavnicka et al., 2017; depression, Williamson et al., 2019; COVID-19, Quatieri et al., 2020), there is little consensus regarding how to characterize the diversity of articulatory impairments across individuals with speech motor disorders. Although articulation is only one of five subsystems (i.e., along with phonation, resonation, respiration, and prosody) necessary to produce speech, articulatory impairments have a disproportionate effect on overall speech intelligibility (Lee et al., 2014; Rong et al., 2015; Sidtis et al., 2011).

The most widely used paradigm for characterizing speech impairments was developed over fifty years ago by Darley, Aronson, and Brown (DAB) (1969a and 1969b) and relies on identifying auditory-perceptual clusters of speech abnormalities. During the clinical assessment of articulatory function, clinicians—guided by the DAB paradigm—listen for signs of articulatory abnormalities, such as imprecise consonants. The detection of articulatory abnormalities triggers a more comprehensive probe into identifying anatomic and physiologic mechanisms. This assessment typically includes grading oral muscle strength and tone as well as movement range, speed, steadiness, and consistency (Darley et al., 1969a and 1969b; Enderby & Palmer, 2008). The reliability and validity of clinician-based assessments, however, have increasingly come into question because of listener bias and disagreement among clinicians (Borrie et al., 2012; Bunton et al., 2007; Kent, 1996). Consequently, researchers have been developing quantitative measures of impaired articulation (Green et al., 2013; Gupta et al., 2016). Identifying *acoustic* features has been of particular interest because good quality speech recordings can now be easily obtained from a multitude of personal devices (e.g., smart phones, laptops, etc.).

In a recent study, we introduced a novel acoustic-based framework for profiling articulatory motor impairments composed of the following five components: *Coordination*, *Consistency*, *Speed*, *Precision,* and *Rate* (Rowe & Green, 2019). The definition of each component is provided in Table 2. These components are based on articulatory features derived from the DAB paradigm (i.e., incoordination, irregular articulatory breakdowns, reduced range of movement, and imprecise consonants) and previous literature investigating limb and speech motor control (Adams et al., 1993; Fletcher, 1972; Gracco & Abs, 1986; Green et al., 2000;

Ketcham & Stelmach, 2003). Each component in our framework is represented by an acoustic feature that best reflects the corresponding articulatory construct (e.g., the number of syllables produced per second represents *Rate*). However, a critical step toward the successful adoption of new speech measures in the clinical and research settings is determining their validity. Despite the abundance of previous research investigating articulatory function and the promise of acoustic-based technology, little attention has been directed towards establishing the analytical, clinical, or divergent validity of acoustic-based features (Allison et al., 2020).

Table 2. Definitions of five articulatory components in framework.

Articulatory Component	Definition
Coordination	❖appropriate temporal alignment of articulatory movements to meet the task demands
Consistency	❖similarity of consecutive repetitions of speech sounds
Speed	❖quickness of articulatory movement during each speech sound
Precision	❖clearness and distinctiveness of speech sounds
Rate	❖quickness of completion of repeated speech sounds

Research on improving clinical speech diagnostics ultimately strives to identify features with high clinical utility (i.e., how impactful the measure is to a patient's functional communication). Clinical utility, however, is predicated on prior evidence of *analytical validity* (i.e., the measure's ability to accurately and reliably assess the quantifiable articulatory characteristic) and *clinical validity* (i.e., the measure's ability to detect the clinical status of the articulatory characteristic) (Baehner, 2016), both of which are within the scope of *criterion validity*. In addition to criterion validity, establishing *construct validity* (specifically *divergent validity*) is crucial for determining that the measures of interest are not redundant and truly represent distinct constructs.

Assessing Criterion Validity of Acoustic-Based Articulatory Features: Analytical Validity

The analytical validity of an acoustic articulatory feature is supported when it is highly correlated with another reliable measure representing the same articulatory construct. For example, the analytical validity of vowel space area (VSA) (measured acoustically) as a construct of *Precision*, which is defined as articulatory distinctiveness, is supported by a strong correlation with tongue displacement (measured kinematically).

Testing for these associations is particularly essential in the speech domain because mappings between articulatory movements and acoustic features are many-to-one; the vocal tract can be configured in various ways to achieve the same acoustic outcome (Stevens, 1972). Features derived from speech biomechanics are increasingly viewed as direct representations of articulatory patterns and, therefore, provide a good reference standard for less direct methods, such as acoustic analyses (Green, 2015). Nevertheless, evidence for the analytical validity of existing acoustic measures of articulation, particularly in patients with disordered speech, is sparse. Below we review recent findings organized by the five key components identified in our framework (i.e., *Coordination*, *Consistency*, *Speed*, *Precision*, and *Rate*).

Speed: For features representing *Speed*, one study investigated speakers with amyotrophic lateral sclerosis (ALS) and found moderate correlations between the slope of the second formant (F2) and the movement speed of the articulators, which was registered using three-dimensional (3D) electromagnetic articulography (Yunusova et al., 2012). F2 slope has been shown to be a viable correlate of articulatory movement *Speed*, as it represents the rate of change in the vocal tract configuration during speech (Kent & Kim, 2003; Kim et al., 2009; Yunusova et al., 2012). ***Precision:*** Yunusova et al. (2012) also investigated the relationship between F2 range and articulator displacement but found only a weak association (Yunusova et al., 2012). This latter finding was in contrast to a prior study that demonstrated strong correlations between acoustic specification or distinctiveness (as indexed by the Euclidean distance in the first formant (F1)/F2 planar space between the high vowel /i/ and the low vowel /a/) and articulatory specification (as indexed by the extent of tongue displacement between the high vowel /i/ and the low vowel /a/) in healthy speakers (Mefferd & Green, 2010). ***Consistency:*** In addition to measures of *Precision*, Mefferd and Green (2010) examined movement variability using the spatiotemporal index (STI) both kinematically and acoustically, revealing a weak association between the two measures of STI (Mefferd & Green, 2010). However, a recently developed acoustic measure of *Consistency*, cycle-to-cycle temporal variability of envelope fluctuations, was shown to be highly correlated with the kinematic correlate of tongue movement jitter in speakers with ALS (Rong, 2020). ***Rate:*** Aside from consistency, Rong (2020) investigated acoustic and kinematic measures of articulatory rate, revealing a strong relationship between syllable repetition *Rate* (derived acoustically) and alternating tongue movement *Rate* (derived kinematically) during an alternating motion rate task (Rong, 2020).

Coordination: Lastly, the validity of acoustic features representing *Coordination* has been minimally investigated. One notable study found strong correlations between the timing of F2 movement and that of the lingual and labial movements during the production of /u/ in healthy speakers and speakers with ALS (Weismer et al., 2003). Taken together, although the literature supporting analytical validation is growing, (1) findings have been inconsistent to date and (2) there is a critical need for additional validation testing of acoustic-based articulatory features.

Assessing Criterion Validity of Acoustic-Based Articulatory Features: Clinical Validity

The aforementioned findings provide preliminary support for the analytic validity of each measure – specifically with regard to their ability to quantitate a target articulatory characteristic, such as *Coordination*. However, despite the uniquity of articulatory testing in clinical speech exams, research examining the *clinical validity* of acoustic-based articulatory features is limited (Enderby & Palmer, 2008). Most clinical validity research has focused on acoustic measures of vocal function rather than articulatory function. While still emerging, this literature is supported by a critical mass of research establishing links between different acoustic-based voice features and their clinical perceptual correlates (e.g., cepstral peak prominence and breathiness, Heman-Ackah et al., 2002; jitter and roughness, Rabinov et al., 1995; harmonics-to-noise ratio and hoarseness, Yumoto et al., 1984). Similar attention is needed to establish the clinical validity of acoustic measures of articulatory motor impairments.

The use of assessments, such as the Frenchay Dysarthria Assessment (FDA) (Enderby & Palmer, 2008), and established frameworks, such as the DAB paradigm (Darley et al., 1969a and 1969b), requires clinicians to routinely identify and rate the severity of perceived articulatory impairments, such as imprecision, to guide the diagnosis of speech motor subtypes. Of the articulatory features derived from the DAB paradigm, the clinical validity of *Rate* has been the most thoroughly investigated, with previous research demonstrating moderate to strong associations between acoustic features representing *Rate* and their perceptual correlates (Grosjean & Lane, 1981; Tjaden, 2000; Turner & Weismer, 1993; Waito et al., 2021).

Although recent work has examined perceptual ratings of *Precision* (Eklund et al., 2015; Folker et al., 2010; Jiao et al., 2016; Sidtis et al., 2011; Tjaden et al., 2014; Waito et al., 2021), only one study, to our

knowledge, used perceptual ratings to examine the clinical validity of an acoustic-based measure of *Precision*. The study by Jiao et al. (2016) found a moderate correlation between their novel measure of articulatory precision (i.e., articulation entropy, which characterizes a speaker's phonemic inventory through a non-parametric estimate of the entropy of the distribution of sounds) and perceptions of precision (Jiao et al., 2016). Previous studies have examined associations between perceptual measures and various acoustic-based articulatory features, such as F2 slope or fricative duration (Kent et al., 1989; Kim et al., 2011; Feenaughty et al., 2014; Lansford & Liss, 2014; Sapir et al., 2007; Weismer & Martin, 1992; Whitfield & Goberman, 2014); however, the perceptual measures consisted of global ratings of severity, intelligibility, or speech clarity rather than direct correlates of the articulatory construct of interest (e.g., *Speed*).

One study used a free classification approach, in which listeners were instructed to group similar-sounding speakers (Lansford et al., 2014). As a separate part of the experiment, five clinicians rated the speech samples on articulatory imprecision. The authors also analyzed articulatory *Precision* using acoustic VSA, formant centralization ratio, and vowel dispersion. This study, however, examined whether the *listeners*, who were not instructed to rate any specific articulatory feature, created clusters that corresponded with the perceptual ratings and acoustic features. Thus, the authors did not assess the association between the clinician ratings and acoustic correlates of *Precision*. Overall, there is a paucity of studies investigating *direct* perceptual correlates of specific acoustic-based articulatory features beyond *Rate* (e.g., F2 slope and *Speed*; syllable length variability and *Consistency*). This gap in the literature translates to a significant need for more research assessing the clinical validity of articulatory features, with the ultimate goal of promoting the use of validated acoustic features in the clinical setting.

Assessing Construct Validity of Acoustic-Based Articulatory Features: Divergent Validity

Construct validity, which determines whether a specific measure is assessing the intended construct, is supported if (1) it is strongly correlated with another measure representing a similar articulatory construct (known as *convergent validity*) and/or (2) it is weakly correlated with another measure representing a distinct articulatory construct (known as *divergent validity*). Research on the construct validity—particularly divergent validity—of acoustic-based articulatory features is scarce at best. This scarcity may be due, in part, to the lack

of studies that simultaneously examine multiple components of articulation. Recent work investigated speech phenotypes in multiple sclerosis (MS) (Rusz et al., 2018) and progressive supranuclear palsy (PSP) (Skodda et al., 2011), but these studies only used one or two features to represent the articulatory subsystem. Indeed, much of the clinical acoustic literature has been based on a unidimensional description of articulation, focusing primarily on global measures such as *Rate*, which reduces the need or ability to examine construct divergence.

Validating an Acoustic-Based Framework of Articulatory Motor Control

As aforementioned, our novel acoustic-based framework for characterizing articulatory impairments (Rowe & Green, 2019) was developed based on key components of motor control identified from the limb and speech literature (i.e., *Coordination*, *Consistency*, *Speed*, *Precision*, and *Rate*). Because impairments in *Coordination*, *Consistency*, *Speed*, *Precision*, and *Rate* are common to many speech motor disorders, these features, individually, are sensitive but not specific. Recent work has demonstrated that individual features, such as *Rate*, do not provide sufficient information to distinguish speech phenotypes within diseases such as ALS (Stipancic et al., under review). Instead, identifying *profiles* of articulatory characteristics, including characteristics that are spared, may be integral for distinguishing speech subtypes.

Articulatory phenotypes are critical not only for diagnostic purposes but also for providing information regarding the potential interaction between components. Prior work, for example, has identified speed-accuracy tradeoffs consistent with Fitts' law (Fitts, 1954; Lammert et al., 2018). Thus, speakers may reduce their speaking rate as a compensatory mechanism to increase accuracy and improve their intelligibility. From a clinical standpoint, articulatory tradeoffs may be essential for understanding why some treatments are more beneficial for one population or patient than another (Yorkston et al., 2007). For example, reduced *Speed* during Slow Speech treatment may result in reduced *Consistency* (Kleinow et al., 2001; Mefferd et al., 2014), which may pose issues for patients with *Consistency* deficits that impact intelligibility. A unifying framework of articulation will advance our understanding of the effects of certain pathophysiologies on articulatory function and the phenotypic profiles that may influence response to therapy.

In two previous studies, we tested the “known groups” validity of the framework (Rowe & Green, 2019; Rowe et al., 2020). Known groups validity assesses whether a measure (e.g., syllables per second) can

differentiate groups known to differ on a variable of interest (e.g., *Rate*). Our studies revealed divergent articulatory profiles for patients with ALS and PD. Individuals with ALS exhibited deficits across all components except *Consistency*, consistent with the increased weakness and effort caused by gradual motor neuron degeneration (Darley et al., 1969a and 1969b; Strand, 2013). In contrast, individuals with PD only demonstrated impaired *Speed*, consistent with the reduced movement resulting from basal ganglia hypofunction (Darley et al., 1969a and 1969b; Strand, 2013). The findings of this research, therefore, provided strong preliminary support for the known groups validity of the framework.

In the current study, we used speech samples from patients with mixed flaccid-spastic dysarthria secondary to ALS because the neuronal degeneration characteristic of this population has broad impacts on articulatory performance, which allows for the representation of deficits in nearly all components of interest. We first assessed the *criterion (analytical and clinical) validity* of the acoustic-based framework by examining associations between: (1) performance on the five acoustic features and performance on their kinematic correlates (*analytical validity*) and (2) performance on the five acoustics features and clinician ratings of the five framework components (*clinical validity*). We then assessed the *construct (divergent) validity* of the five acoustic features by comparing them to one another. That is, we know from prior work that the features selected for the framework assess articulatory performance. However, to determine whether the features represent *distinct* articulatory constructs, we examined correlations between each pair of acoustic features, with weaker associations interpreted as greater divergence. Our research questions were as follows:

1. ***Criterion (Analytical and Clinical) Validity:*** How does performance on the acoustic features relate to performance on their kinematic correlates *(analytical validity)* and to clinician ratings of the five components *(clinical validity)?*
2. ***Construct (Divergent) Validity:*** Do the acoustic features in our framework represent distinct articulatory constructs?

Controlling for Severity

Establishing criterion and construct validity can be challenging because of the potentially confounding effects of overall severity. Researchers must continually ask, *How can I be sure that a measure of "precision"*

actually measures precision and is not solely a proxy for overall speech motor severity? The overall severity of a speech motor disorder is likely to have a global impact on speech with a concurrent effect on many, if not all, measures of articulatory function. This severity-driven covariation among measures of articulation may challenge statistical approaches designed to test for unique articulatory motor deficits. As a result, strong correlations between measures may be due to a relationship with a third variable, such as severity, rather than to excellent analytical or clinical validity, or poor divergent validity.

Nevertheless, including speakers across the whole range of the severity spectrum allows for greater generalizability of a study's findings and provides the variability needed to adequately model the data. Indeed, severity often drives changes in articulation, particularly in a population with impaired speech. Thus, removing the variability related to severity may also remove important variability in articulatory function needed to assess differences in the measures. To account for the different roles of severity in this study, we first examined clinical, analytical, and divergent validity *without* controlling for severity. Then, for the measures with moderate to strong associations, we conducted partial correlations—with severity as a covariate—to provide a better understanding of its contribution to the relationship.

METHOD

Subjects

Subjects for the study included 22 individuals with ALS (14 males, 8 females; age in years: 39-71, M = 52.89, SD = 9.65) selected from two databases. The demographic and speech characteristics of the subjects are outlined in Table 1. Seven ALS speakers were selected from the X-ray Microbeam (XRMB) Speech Production Database (R01DC00820; Westbury, 1994), and 15 ALS speakers were selected from the database of an ongoing longitudinal study conducted at the MGH Institute of Health Professions (IHP) (R01DC0135470; Green, 2013). Participants in the MGH IHP database scored within normal limits on the ALS Cognitive Behavioral Screen (CBS; Wooley et al., 2010). The XRMB database did not include information on cognitive function. All subjects were native English speakers and passed a bilateral hearing screening at 30dB at 1000, 2000, and 4000 Hz.

Table 1. Demographic and speech characteristics of speakers with ALS.

Subject	Age	Sex	Database	SMR Severity Rating	Overall Severity Rating	Intelligibility Rating
ALS01*	56	M	MGH IHP	1.5	0	0
ALS02*	46	F	MGH IHP	4.5	3	0
ALS03*	34	M	MGH IHP	0	0	0
ALS04*	64	F	MGH IHP	76	91	76.5
ALS05*	—	M	MGH IHP	72.5	28.5	0
ALS06	62	F	MGH IHP	83.5	100	100
ALS07*	60	F	MGH IHP	72	45.5	24
ALS08*	50	M	MGH IHP	36.5	10.5	0
ALS09*	54	M	MGH IHP	5	2	0
ALS10	59	M	MGH IHP	68.5	14.5	0
ALS11*	56	M	MGH IHP	85	89.5	75
ALS12	—	M	MGH IHP	0	1.5	0
ALS13*	59	M	MGH IHP	3.5	0	0
ALS14	—	F	MGH IHP	68.5	43	14
ALS15*	—	F	MGH IHP	26	6	0
ALS16	39	F	XRMB	63.5	52.5	13
ALS17	43	M	XRMB	25.5	30	0
ALS18	55	M	XRMB	100	100	100
ALS19	44	M	XRMB	100	93	86.5
ALS20	43	M	XRMB	100	100	100
ALS21	71	F	XRMB	79.5	47	11
ALS22	57	M	XRMB	10	18	5.5

Note. Em dashes indicate missing data. Severity and intelligibility ratings were made based on a visual analog scale (VAS) with the labels "Normal" (0) on the left to "Very Impaired" (100) on the right. * = subject was included in kinematic analysis.

For the purpose of describing the speech of the subjects, two licensed speech-language pathologists (SLPs) rated perceptual severity and intelligibility of the speakers during two tasks (i.e., the sequential motion rate [SMR] task and a connected speech task) on visual analog scales (VASs); the methods for obtaining these ratings will be discussed further in the Procedure section. The subjects exhibited a wide range of SMR severity ratings (based on the SMR task) (M = 49.16, SD = 37.67), overall severity ratings (based on the connected speech task) (M = 33.86, SD = 38.72), and intelligibility ratings (based on the connected speech task) (M = 27.52, SD = 39.87). Independent samples t-tests of the perceptual ratings between the two databases revealed no significant differences for any of the severity rating types (SMR severity rating: $t(11.25) = -1.67$, $p = .12$; overall severity rating: $t(12.51) = -2.11$, $p = .06$; and intelligibility rating: $t(9.08) = -1.29$, $p = .23$, respectively).

Procedure

All subjects performed the SMR task (part of the diadochokinetic [DDK] task), in which they repeated the syllable sequence /pataka/ as quickly and accurately as possible on one breath. An experimenter was present in the room to ensure that subjects performed the task correctly. The SMR task was used because it is efficient, widely implemented in the clinical setting, and sensitive to bulbar motor involvement in individuals with ALS (Rong et al., 2018). Moreover, it is well suited to extract acoustic features that are representative of each component.

Acoustic data acquisition. For both databases, acoustic data were collected at a sampling frequency of 22 kHz. Data was collected in a quiet room to minimize any background noise that may have introduced variability into the acoustic signal. During the task, subjects wore a head-mounted, professional-quality microphone positioned approximately 5 cm from their mouth.

Kinematic data acquisition. Kinematic data for 11 subjects (all from the MGH IHP database) were analyzed for this study due to limitations in data availability. Kinematic data was collected at a sampling frequency of 40-160 Hz. Lip and tongue movements were obtained using the Wave electromagnetic articulography tracking system (Wave; Northern Digital, Inc.). Five-degree-of-freedom sensors were attached to four anatomical locations: (1) the vermillion border at the center of the upper lip; (2) the vermillion border at the center of the lower lip; (3) on the surface of the ventral tongue approximately one cm posterior to the apex (i.e., tongue tip); and (4) on the surface of the dorsal tongue approximately four cm posterior to the apex (i.e., tongue dorsum). All signals were recorded in three dimensions (i.e., anterior-posterior, superior-inferior, and medial-lateral). Lip sensors were adhered to the subject using medical tape and tongue sensors were placed using PeriAcryl Oral Tissues Adhesive (GluStitch Inc.), a non-toxic dental glue. For a reference marker, one six-degree-of-freedom sensor was secured to the center of the subject's forehead. This marker was used to re-express the lip and tongue data relative to a head-based coordinate system, thereby subtracting head movement from the analysis.

Perceptual data acquisition. Two licensed SLPs with clinical expertise in speech motor disorders rated the speech of the individuals with ALS in this study. Each clinician listened to a total of 54 randomized audio recordings of the subjects using Audacity (Audacity Team, 2020). The clinicians first listened to 27

randomized recordings of connected speech, which included five randomly selected subject recordings repeated for the calculation of intrarater reliability. To maximize the number of words heard by the listener, the 15-word sentence from the randomly generated Sentence Intelligibility Test (SIT; Yorkston et al., 2007) was selected for the subjects from the MGH IHP database. Since the SIT was not collected for the XRMB database, a segment of the hunter passage (i.e., *"This March, Tom found himself by a small stream with a gun at rest in the crook of his arm."*) was selected. This segment was chosen for its comparable length to the 15-word sentence from the MGH IHP database. While listening to each recording, the clinicians completed a survey through REDCap electronic data capture tools (Harris et al., 2019) hosted at MGH IHP. In this survey, the SLPs were asked to rate overall severity and intelligibility of the subject's speech on a 100-point VAS with endpoints labeled "Very Impaired" (100) on the left and "Normal" (0) on the right.

The clinicians then listened to 27 randomized recordings of the SMR task, which included five randomly selected subject recordings again repeated for intrarater reliability. While listening to each recording, the clinicians completed another survey through REDCap. In this survey, they were first asked to rate the severity of the SMR task. The survey then explained that the clinicians would be given five components of articulation along with their definitions on which to rate each subject's SMR performance (see Table 2). Each item was rated using a 100-point VAS with endpoints labeled "Very Impaired" (100) on the left and "Normal" (0) on the right. Recent work has demonstrated strong correlations between VAS ratings and orthographic transcription (Abur et al., 2019; Hustad, 2006; Stipancic et al., 2016) and even slightly higher interrater reliability for VAS ratings than for intelligibility derived from orthographic transcription (Hustad, 2006; Stipancic et al., 2016). For both connected speech and SMR recordings, the clinicians were blind to speaker identity and diagnosis and were allowed to re-listen to the recordings as many times as needed. These ratings served as the perceptual features used for this study (see Table 3).

Table 3. Acoustic, kinematic, and perceptual features used to assess analytical and clinical validity.

Acoustic	Kinematic	Perceptual
	Coordination	

Ratio of /p/-/t/ Duration to /t/-/k/ Duration (DurRatio): **1.** Durations from (a) release burst of /p/ to release burst of /t/ and (b) release burst of /t/ to release burst of /k/ were extracted from each repetition of /pataka/ **2.** Ratio of (a) to (b) was calculated for each repetition of /pataka/ **3.** Mean ratio of (a) to (b) was calculated across all three repetitions for each subject	**Ratio of Durations between Labial and Lingual Closures:** **1.** Durations from (a) labial opening for /pa/ to anterior lingual release for /ta/ and (b) anterior lingual release for /ta/ to posterior lingual release for /ka/ were extracted from each repetition of /pataka/ **2.** Ratio of (a) to (b) was calculated for each syllable of /pataka/ **3.** Mean ratio of (a) to (b) was calculated across all three repetitions for each subject	ratings of appropriate temporal alignment of movements to meet task demands
	Consistency	
Between-Repetition Variability in Syllable Duration (RepVar.Syll): **1.** Durations from (a) release burst of /p/ to release burst of /t/, (b) release burst of /t/ to release burst of /k/, and (c) release burst of /k/ to release burst of /p/ were extracted from each repetition of /pataka/ **2.** Standard deviation of (a), (b), and (c) was calculated across three repetitions for each syllable (e.g., /pa1/, /pa2/, /pa3/) **3.** Mean standard deviation of syllable duration was calculated across all three syllables for each subject	**Between-Repetition Variability in Durations between Closures:** **1.** Durations from (a) labial closure for /pa/ to anterior lingual elevation for /ta/, (b) anterior lingual elevation for /ta/ to posterior lingual elevation for /ka/, and (c) posterior lingual elevation for /ka/ and labial closure for /pa/) were extracted from each repetition of /pataka/ **2.** Standard deviation of (a), (b), and (c) was calculated across three repetitions for each syllable (e.g., /pa1/, /pa2/, /pa3/) **3.** Mean standard deviation of syllable duration was calculated across all three syllables for each subject	ratings of stability of speech sounds across multiple repetitions
	Speed	
Second Formant Slope in Consonant-Vowel Transition (F2Slope): **1.** F2 time series was extracted from vowel segment in /ka/, and F2 slope of consonant-vowel transition was calculated using first timepoint as onset frequency and midpoint as offset frequency **3.** Mean F2 slope of /ka/ was calculated across all three repetitions for each subject	***Lingual Velocity:*** **1.** Maximum and minimum values of first derivative of posterior lingual movement (as indexed by TD_d velocity maxima and minima) were extracted from /ka/ **2.** Mean TD_d velocity of /ka/ was calculated across all three repetitions of /ka/ for each subject	ratings of quickness of movement during each syllable repetition
	Precision	
Between-Consonant Variability in F2 Slope (ConVar.F2Slope): **1.** F2 time series was extracted from vowel segment in /pa/, /ta/, and /ka/, and F2 slope of consonant-vowel transition was calculated using first timepoint as onset frequency and midpoint as offset frequency **2.** Standard deviation of F2 slope was calculated across /pa/, /ta/, and /ka/ for each repetition of /pataka/ **2.** Mean standard deviation of F2 slope was calculated across all three repetitions for each subject	**Between-Consonant Variability in Lingual Speed:** **1.** Maximum and minimum values of first derivative of posterior lingual movement (as indexed by TD_d velocity maxima and minima) were extracted during /pa/, /ta/, and /ka/ for each repetition of /pataka/ **2.** Standard deviation of TD_d velocity was calculated across /pa/, /ta/, and /ka/ for each repetition of /pataka/ **3.** Mean standard deviation of TD_d velocity was calculated across all three repetitions for each subject	ratings of clearness and distinctiveness of consonants
	Rate	
Syllables per Second (SyllRate): **1.** Duration from release burst of /p/ in first repetition of /pataka/ to vowel offset in third repetition of /ka/ was extracted for each subject **2.** Syllables per second was calculated by dividing number of syllables produced in three repetitions by task duration	**Closures per Second:** **1.** Duration from labial closure for /pa/ in first repetition of /pataka/ to labial closure for /pa/ in fourth repetition of /pataka/) was extracted for each subject **2.** Syllables per second was calculated by dividing number of cycles of labial closure for /pa/, anterior lingual elevation for /ta/, and posterior lingual elevation for /ka/ produced in three repetitions by task duration	ratings of quickness of completion of syllable sequences

Acoustic Analysis

The acoustic recordings were analyzed using Praat (Boersma, 2001). Prior to analysis, each wideband spectrogram was manually reviewed for signal artifacts. Formant settings were set to a maximum frequency of 5500 Hz for women and 5000 Hz for men. The first repetition of /pataka/ (identified by the

release burst of /p/ in /pa/ to the final glottal pulse in /ka/) was excluded due to the variability often associated with initiating a complex speech task. Each spectrogram was parsed at the phoneme level for three repetitions of the sequence /pataka/. Each consonant was parsed from its release burst to the first glottal pulse of the subsequent vowel (i.e., the vowel onset). Each vowel was then parsed from the vowel onset, which also includes the consonant transition, to the final glottal pulse (i.e., the vowel offset). These signals were used to develop the acoustic features that make up our novel framework of articulatory motor control (see Figure 1 and Table 3). The data required for each feature was extracted using a custom Praat script and calculated using custom MATLAB (MathWorks, 2019) and R (R Core Team, 2014) scripts. The rationale for the feature selected for each framework component is as follows:

***Coordination* → Ratio of /p/-/t/ Duration to /t/-/k/ Duration (DurRatio):** Articulatory *Coordination* is a poorly defined construct and, as a result, many different acoustic features—representing arguably distinct constructs—have been used in the speech motor control literature. Seminal work in speech motor control posited that *Coordination* refers to the spatiotemporal patterning of speech articulators and has been measured using quantitative indices representing movement sequencing, interarticulator coupling, motor equivalence covariability, and movement smoothness (Caruso et al., 1988; Gracco & Abs, 1986). *DurRatio* best reflects the indices of movement sequencing and smoothness. In the case of the SMR task, sequencing from the labial closure for /p/ to the tongue tip closure for /t/ and then to the tongue back closure for /k/ with the appropriate relative timing of the three gestures will result in a smooth production of the syllable sequence. If all articulatory closures are achieved in an efficient manner, the /p/-/t/ and /t/-/k/ durations should be approximately equal, producing a ratio of 1. *DurRatio* thus reflects movement smoothness during a sequencing task, with values further from 1 indicating reduced *Coordination*.

***Consistency* → Between-Repetition Variability in Syllable Duration (RepVar.Syll):** *RepVar.Syll* consists of the standard deviation in syllable duration across repetitions of /pataka/. Variations of *RepVar.Syll* have frequently been employed in the clinical speech science literature, as syllable variability is a sensitive measure in populations who lack stability in repeated sound productions (Kent & Rosenbek, 1983; McNeil et al., 1995; Rong, 2020). *RepVar.Syll* thus reflects variability in syllable duration across multiple productions. All values were reversed (multiplied by -1) to ease interpretability of the correlations with the kinematic and

perceptual measures. Therefore, larger *negative* values indicate more variability or reduced *Consistency*.

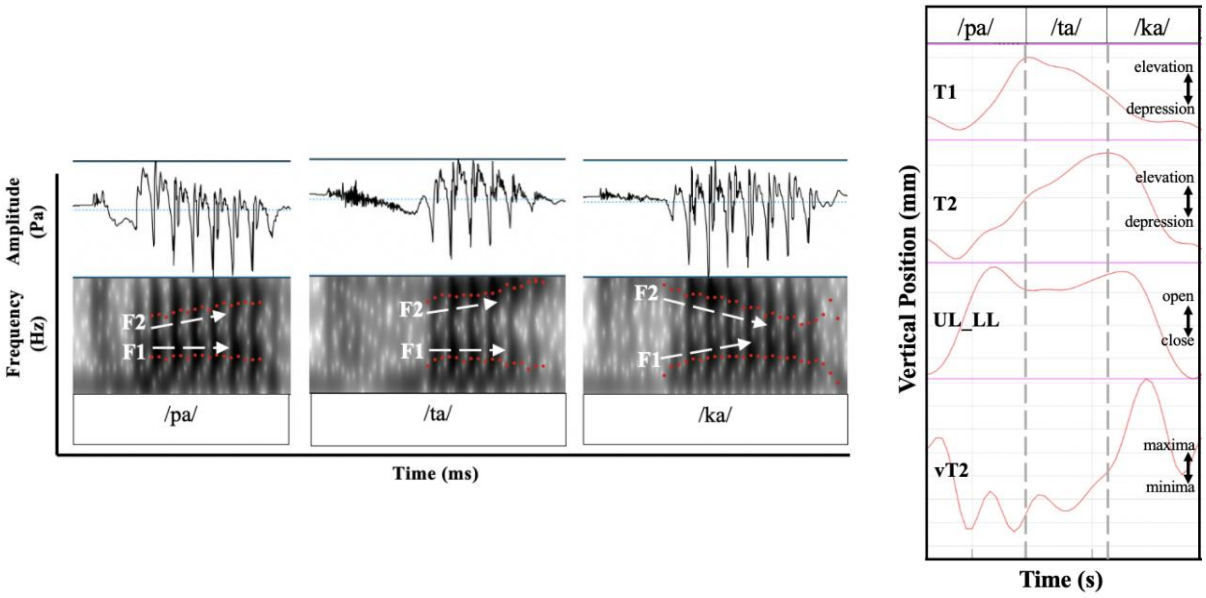
***Speed* → F2 Slope of Consonant Transition (F2Slope):** *F2Slope* consists of the slope of F2 (i.e., range/time) in the consonant-vowel transition of /ka/. Prior work has advocated for the use of stimuli that result in large changes in vocal tract configuration (Kim et al., 2009). The decreasing consonant-vowel transition of /ka/ was, therefore, used because it is the greatest and most robust of the three consonants, suggesting that the transition may be particularly sensitive to articulatory abnormalities. *F2Slope* thus reflects the speed of lingual movement, with smaller values indicating reduced *Speed*.

***Precision* → Between-Consonant Variability in F2 Slope (ConVar.F2Slope):** Much of the acoustic and kinematic literature on articulatory *Precision* has focused on vowel distinctiveness, using measures such as vowel space area, formant ranges during the production of diphthongs, or articulatory displacement during the production of corner vowels. In the current study, we focused on consonant distinctiveness because behavioral research has shown that consonants, though more challenging to measure acoustically, may have a greater impact than vowels on word-based intelligibility (Owren & Cardillo, 2006). *ConVar.F2Slope* consists of the standard deviation in F2 slope across the three distinct consonant-vowel transitions of /pa/, /ta/, and /ka/. F2 slope was selected over F2 range because of the information it provides about place *and* manner of articulation, as the slope encapsulates both formant frequency range and length (Chen & Alwan, 2000; Delattre et al., 1955; MacKay, 2014). Indeed, place of articulation can be determined by formant frequency range (e.g., compared to alveolar stops, bilabial stops produce a longer vocal tract, which results in a smaller transition range), while manner of articulation can be determined by the length of the consonant transition (e.g., compared to stops, fricatives tend to have a slower rate of vocal tract adjustment, which results in a longer, more gradual transition length) (MacKay, 2014). Using the transitions of /p/, /t/, and /k/ is particularly informative given the antagonistic places of articulation for each phoneme, which should theoretically produce highly distinct transitions. *ConVar.F2Slope* thus reflects variability in F2 slope across the three discrete consonants, with smaller values indicating less variation or reduced *Precision*.

***Rate* → Syllables Produced Per Second (SyllRate):** *SyllRate* is the number of syllables produced per second during the SMR task. The measure of *Rate* in syllables per second has been used extensively throughout acoustic literature, as reductions in this measure are a prominent sign of articulatory decline in

speech motor populations (Nishio & Nimi, 2006; Tjaden & Watling, 2003; Ziegler & Wessel, 1996). *SyllRate* thus reflects the rate of syllable sequence performance, with smaller values indicating reduced *Rate*.

Figure 1. *Left panel:* Acoustic spectrogram and waveform of the utterance /pataka/ produced by a speaker with ALS. *Right panel:* Time series of anterior lingual movement (T1), posterior lingual movement (T2), labial movement (UL_LL), and velocity of posterior lingual movement (vT2) during the utterance /pataka/ produced by a speaker with ALS.



Kinematic Analysis

All kinematic data was analyzed using a custom MATLAB (MathWorks, 2019) software program called Speech Movement Analysis for Speech and Hearing research (SMASH; Green et al., 2013). Prior to analysis, each file was manually checked for movement artifacts and missing markers. Lip and tongue movements in each signal were then smoothed using a 15-Hz low-pass filter to remove high-frequency noise and resampled at 100 Hz. To account for head movement during the task, 3D Euclidean distances were derived between the following signals: (1) the reference marker (on the forehead) and the tongue tip (TT_d); (2) the reference marker and the tongue dorsum (TD_d); and (3) the upper and lower lips (UL_LL_d). All movements were assessed along the vertical (y) axis of the frontal plane (see Figure 1). The first repetition of /pataka/ (identified by the first cycle of upper and lower lip closure) was excluded due to the movement

variability often associated with initiation of a complex speech task. Each file was parsed at the syllable level for three repetitions of /pataka/. The syllable /pa/ was parsed from the UL_LL_d minima (i.e., lip closure) to the subsequent TT_d minima (i.e., tongue tip closure), /ta/ was parsed from the TT_d minima to the subsequent TD_d minima (i.e., tongue dorsum closure), and /ka/ was parsed from the TD_d minima to the subsequent UL_LL_d minima. These signals were used to develop the kinematic measures used for this study (see Figure 1 and Table 3).

Statistical Analysis

All statistical analyses were completed in R (R Core Team, 2014).

Interrater Reliability

For both the acoustic and kinematic data, the first author parsed all the SMR samples. To obtain interrater reliability, a second trained researcher then re-measured the boundaries of a random selection of 20% of the samples from both datatypes. Because parsing was the only manual step in the data analysis, reliability was assessed based on the agreement on the measures derived from the parsed data. ICC (2,1) (i.e., two-way random single measures used for consistency/absolute agreement) was used to calculate interrater reliability for the acoustic and kinematic data as well as for SMR severity ratings, overall severity ratings, intelligibility ratings, and perceptual ratings of each framework component.

Intrarater Reliability

To obtain intrarater reliability, the first author re-measured the boundaries of a random selection of 20% of the SMR samples from both the acoustic and kinematic data. ICC (2,1) was used to assess intrarater reliability on the measures derived from the parsed acoustic and kinematic data, whereas percent agreement was used to assess intrarater reliability for SMR severity ratings, overall severity ratings, intelligibility ratings, and perceptual ratings of each framework component. For the measure of percent agreement, the ratings were considered to be in "agreement" if they were within 10 points on the 100-point VAS. Intrarater reliability was calculated for both clinicians individually, and then the mean intrarater reliability was calculated. Percent agreement was used instead of ICC (2,1) for the perceptual ratings due to the small

sample size of data points (i.e., five) used for this calculation, as each repeated speaker only had one datapoint. Although only a small number of speakers were re-measured for the acoustic and kinematic data as well, we had a sufficient amount of data to assess reliability using ICC (2,1) because measures were calculated for each repetition (i.e., first, second, and third) and syllable (i.e., /pa/, /ta/, /ka/).

Effects of Sex on Acoustic Features

Because there was an unequal ratio of males and females in the group of ALS speakers, we used independent t-tests to examine the effect of sex on each of the five acoustic features.

Criterion (Analytical and Clinical) Validation of Acoustic Features

To examine the relationship between performance on the acoustic features and clinician ratings of the five components (clinical validity), we conducted either Pearson product moment correlations or Spearman's rank order correlations (depending on the normality of the model residuals) between the acoustic measures and perceptual ratings of each articulatory component. Subsequently, for the acoustic measures and perceptual ratings that were moderate to strongly associated, we conducted partial correlations with severity as a covariate. Similarly, to examine the relationship between performance on the acoustic features and performance on their kinematic correlates (analytical validity), we conducted Pearson product moment correlations or Spearman's rank order correlations (depending on the normality of the model residuals) between the acoustic and kinematic measures for each component. Subsequently, for the acoustic and kinematic measures that were moderate to strongly associated, we conducted partial correlations with severity as a covariate.

Construct (Divergent) Validation of Acoustic Features

To examine construct validity, we assessed the pairwise associations between performance on the acoustic features. For this analysis, we first conducted either Pearson product moment correlations or Spearman's rank order correlations (depending on the normality of the model residuals) for each pair of

450 acoustic features. Subsequently, for the pairs of acoustic features that were moderate to strongly associated,
451 we conducted partial correlations with severity as a covariate.
452
453 **RESULTS**
454 Descriptive statistics of performance on the five acoustic-based framework components in ALS
455 speakers are reported in Table 5.

Table 5. Means and standard deviations of the five acoustic-based framework components in speakers with ALS.

Coordination (DurRatio)	Consistency (RepVar.Syll)	Speed (F2Slope)	Precision (ConVar.F2Slope)	Rate (SyllRate)
1.04 (.45)	-13.39 (6.97)	4.94 (4.32) kHz	5.44 (3.67)	4.92 (2.03) syll/sec

456
457 ***Interrater Reliability***
458 ICC (2,1) demonstrated moderate to excellent interrater reliability for SMR severity ratings, overall
459 severity ratings, speech intelligibility ratings, and all five components measured acoustically, kinematically,
460 and perceptually except *Consistency* (see Table 4). For the moderate ICCs (i.e., .63 for acoustic *Consistency*,
461 .68 for kinematic *Speed*, and .70 for kinematic *Precision*), scatterplots of the data revealed an outlier that
462 influenced each ICC calculation. When the outlier was temporarily removed in a post hoc analysis, all ICCs
463 increased to .83 or greater. We were, however, surprised to find poor reliability for the perceptual ratings of
464 *Consistency*. Further inspection of the scatterplots of the perceptual data revealed two distinct ranges of
465 *Consistency* values for the clinicians, with one range much more restricted than the other. Despite a strong
466 correlation between the perceptual ratings of the two clinicians, one clinician used the entire length of the
467 VAS to rate *Consistency* (i.e., between 0 [normal] to 100 [very impaired]), whereas the other clinician rated
468 *Consistency* between 0 and 20 for all subjects. This discrepancy in VAS ranges has been explored in prior
469 work on how the internal yardsticks of raters tend to differ for perceptual features (Miller, 2013). Differences
470 in what a rater regards as disordered can be influenced by many factors, such as the order of stimuli
471 presentation during the listening task or clinician experience with the population or parameters on which

they are rating. These factors may have contributed to the poor reliability in the perceptual ratings of *Consistency*, though additional research is needed to further test this potential explanation.

Table 4. Inter- and intrarater reliability with significance levels (* $p < .05$, ** $p < .01$, *** $p < .001$) for acoustic, kinematic, and perceptual measures of each component in speakers with ALS.

Measure	Interrater Reliability	Intrarater Reliability
Coordination		
Acoustic (DurRatio)	ICC (2,1) = .90**	ICC (2,1) = .98***
Kinematic (Variability in Lingual Correlation)	ICC (2,1) = .89**	ICC (2,1) = .89***
Perceptual (Coordination Rating)	ICC (2,1) = .86**	% agreement = 80
Consistency		
Acoustic (RepVar.Syll)	ICC (2,1) = .63*	ICC (2,1) = .92**
Kinematic (Variability in Durations between Closures)	ICC (2,1) = .96***	ICC (2,1) = .97***
Perceptual (Consistency Rating)	ICC (2,1) = .31*	% agreement = 90
Speed		
Acoustic (F2Slope)	ICC (2,1) = .88**	ICC (2,1) = .77*
Kinematic (Lingual Velocity)	ICC (2,1) = .68*	ICC (2,1) = .68*
Perceptual (Speed Rating)	ICC (2,1) = .92**	% agreement = 70
Precision		
Acoustic (ConVar.F2Slope)	ICC (2,1) = .83*	ICC (2,1) = .82*
Kinematic (Variability in Lingual Velocity)	ICC (2,1) = .70***	ICC (2,1) = .71**
Perceptual (Precision Rating)	ICC (2,1) = .77*	% agreement = 100
Rate		
Acoustic (SyllRate)	ICC (2,1) = .99***	ICC (2,1) = .99***
Kinematic (Closures per Second)	ICC (2,1) = .99***	ICC (2,1) = .99***
Perceptual (Rate Rating)	ICC (2,1) = .84*	% agreement = 100
Severity		
SMR Severity Rating	ICC (2,1) = .96***	% agreement = 80
Overall Severity Rating	ICC (2,1) = .92**	% agreement = 90
Speech Intelligibility Rating	ICC (2,1) = .97***	% agreement = 100

Intrarater Reliability

ICC (2,1) demonstrated moderate to excellent intrarater reliability for all five components measured acoustically and kinematically (see Table 4). Percent agreement ranged from 70 to 100 for SMR severity ratings, overall severity ratings, speech intelligibility ratings, and ratings on all five framework components (see Table 4). For the moderate ICCs (i.e., .68 for kinematic *Speed* and .71 for kinematic *Precision*), scatterplots of the data revealed an outlier that influenced the ICC calculations. Therefore, similar to results for interrater reliability, all ICCs increased to .91 or greater when the outlier was temporarily removed in a post hoc analysis.

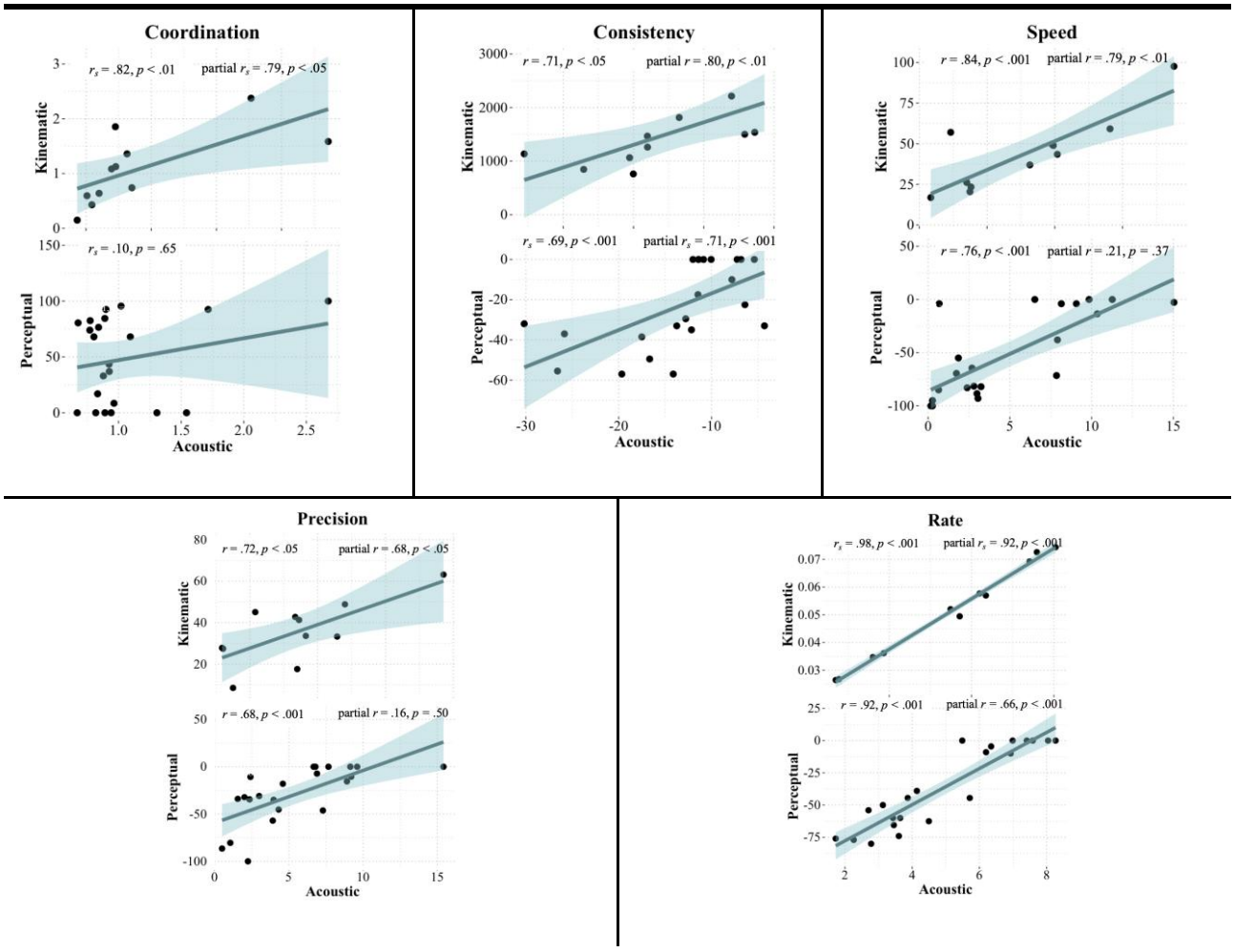
Effects of Sex on Acoustic Features

There were no significant effects of sex for any of the five acoustic features. We thus combined the data from both sexes for all correlation analyses.

Criterion (Analytical and Clinical) Validation of Acoustic Features (Research Question 1)

To determine whether to use a parametric or nonparametric correlation for each pair of measures, we conducted the Kolmogorov-Smirnov (KS) test, which assesses the normality of the distribution of the model residuals. The KS test revealed violations of the assumption of normality ($p < .05$) for four of the 10 pairs. Thus, the four aforementioned pairs were analyzed using Spearman's correlations, while the remaining six were analyzed using Pearson correlations (see Figure 2).

Figure 2. Pearson (r) or Spearman's (r_s) bivariate and partial correlation coefficients demonstrating the acoustic-kinematic and acoustic-perceptual relationships for the five components in speakers with ALS (Research Question 1).



For the acoustic-kinematic comparisons (i.e., analytical validity), all correlation coefficients revealed moderate to strong associations for the five components both with and without controlling for severity (see Figure 2). For the acoustic-perceptual comparisons (i.e., clinical validity), moderate to strong relationships were found between the acoustic features and the perceptual ratings for all components except *Coordination (DurRatio)* without controlling for severity. However, when severity was used as a covariate, moderate to strong associations were found only for *Consistency (RepVar.Syll)* and *Rate (DDKRate)* (see Figure 2).

Construct (Divergent) Validation of Acoustic Features (Research Question 2)

To determine whether to use a parametric or nonparametric correlation for each pair of measures, we again conducted the KS test. The KS test revealed violations of the assumption of normality ($p < .05$) for seven of the 10 pairs. Thus, the seven aforementioned pairs were analyzed using Spearman's correlations, while the remaining three were analyzed using Pearson correlations (see Table 6).

Table 6. Pearson (r) or Spearman's (r_s) bivariate and partial correlation coefficients with significance levels (* $p < .05$, ** $p < .01$, *** $p < .001$) demonstrating divergence in performance on the acoustic features in speakers with ALS (Research Question 2).

	Coordination	**Consistency**	**Speed**	**Precision**	**Rate**
Coordination Consistency	r_s = -.13				
Speed	r_s = .17	r_s = .49* partial r_s = .02			
Precision	r_s = .15	r_s = .51* partial r_s = -.01	r_s = .92*** partial r = .81***		
Rate	r_s = -.13	r_s = .62** partial r_s = .21	r_s = .81*** partial r = .49*	r_s = .85*** partial r = .52*	

Without controlling for severity, the correlation coefficients revealed weak relationships for four of the ten pairs of acoustic features, moderate relationships for three pairs, and strong relationships for the final three pairs (see Table 6). However, when controlling for severity in the moderate to strong correlations, all relationships decreased, with only *Speed (F2Slope)* and *Precision (ConVar.F2Slope)* still demonstrating a strong association. It should be noted that the weak associations, which best exemplify divergent validity,

were not statistically significant. However, detecting such small effects requires a larger sample size than we had in our study. A post hoc power analysis demonstrated that, while detecting the medium to large effects we found (r = .5 to r = .9) requires approximately 20 participants, detecting a small effect (r = .2) would require 266 participants. Thus, we expect these weak relationships to become statistically significant with a larger sample size, as more datapoints would increase statistical power.

DISCUSSION

The goal of this study was to validate a novel acoustic-based framework for characterizing articulatory motor impairments in speakers with ALS. The framework is composed of five unique components of articulatory control: *Coordination*, *Consistency*, *Speed*, *Precision*, and *Rate*. Our first research question examined criterion validity by comparing performance on the acoustic features to their kinematic correlates and to clinician ratings of the framework components. Our second research question examined the construct validity of the five acoustic features by comparing performance on each acoustic feature to the other four features. There were two key findings from our study: (1) All acoustic features except *Coordination (DurRatio)* exhibited moderate to strong analytical *and* clinical validity; and (2) construct divergence was demonstrated across most correlations. All acoustic, kinematic, and perceptual measures, except for the perceptual rating of *Consistency*, demonstrated moderate to excellent inter- and intrarater reliability.

Evidence for Criterion Validity of Acoustic-Based Articulatory Features: Analytical Validity

Taken together, analytical validity was supported by moderate to strong associations between the acoustic features and their kinematic correlates for all five framework components both with and without controlling for severity. These findings were largely in line with those of previous work. The strong correlation we found for *Coordination* is consistent with the associations Weismer et al. (2003) reported between the timing of F2 movement and of lingual and labial movements. Similarly, Mefferd and Green (2010) observed a moderate association between acoustic- and kinematic-based articulatory specification (*Precision*) in healthy speakers (Mefferd & Green, 2010). The strong relationships we found for *Speed*, *Rate*,

and *Consistency* are also consistent with prior research. Yunusova et al. (2012), in her work examining movement *Speed*, noted strong associations between articulator *Speed* (derived kinematically) and F2 slope in speakers with ALS. Additionally, Rong (2020) found strong associations between (1) the *Rate* of alternating tongue movements and acoustic envelope cycles per second (*Rate*) and (2) the variability in the acoustic envelope and tongue movement jitter (*Consistency*) in speakers with ALS (Rong, 2020).

Nevertheless, there were two notable distinctions between our findings and those of previous studies. First, Yunusova et al. (2012) found only a weak relationship between F2 range and articulator displacement (*Precision*) in speakers with ALS. This discrepancy may be partially due to the differences in the type of phonemes examined, as Yunusova et al. (2012) investigated vowel distinctiveness, while our study investigated consonant distinctiveness. Second, Mefferd and Green (2010) observed a weak correlation between acoustic and kinematic measures of *Consistency*. The authors, however, examined only healthy controls, who typically exhibit less variability than patients with speech motor disorders, which would yield a weaker relationship (Mefferd & Green, 2010).

Evidence for Criterion Validity of Acoustic-Based Articulatory Features: Clinical Validity

Overall, without controlling for severity, our study revealed moderate to strong clinical validity for every acoustic feature except *Coordination (DurRatio)*. With severity as a covariate, the associations between the acoustic features for *Speed (F2Slope)* and *Precision (ConVar.F2Slope)* and their perceptual ratings decreased to weak relationships. However, as aforementioned, because severity can drive articulatory impairments in speech-impaired populations, removing all the variability related to severity may also remove important differences in articulatory function. This interdependence of severity and articulatory function may be particularly present when both variables are determined by clinician judgment. Therefore, the decrease in association upon controlling for severity does not necessarily detract from the clinical validity of *Speed* and *Precision*. It may, instead, reflect the disproportionate role of perceived *Speed* and *Precision* in the clinicians' judgments of severity or of severity in determining *Speed* and *Precision*.

Of the five articulatory components, *Rate* and *Precision* were, to our knowledge, the only components for which clinical validity had been investigated prior to the current study. The moderate to

strong correlation between *Rate (SyllRate)* and its perceptual correlate is consistent with previous studies, all of which revealed strong associations between instrumental and perceptual measures of *Rate* (Grosjean & Lane, 1981; Tjaden, 2000; Turner & Weismer, 1993). Likewise, the moderate correlation observed in our study between acoustic and kinematic measures of *Precision*, prior to controlling for severity, has been reported previously (Jiao et al., 2016). Findings of strong clinical validity for *Rate* and *Precision* are not surprising given the auditory-perceptual cues that might encode slower *Rate* (e.g., increased vowel duration, decreased pause duration) and imprecise speech (e.g., consonant bursts, aspiration, voicing distinctions).

In contrast to *Rate* and *Precision*, no study, to our knowledge, has examined the association between auditory perceptions of *Coordination*, *Consistency*, or *Speed* and their quantitative correlates in speech motor impaired populations. Similar to *Rate* and *Precision*, *Speed* and *Consistency* may be encoded by acoustic cues (e.g., increased vowel duration for *Speed* and between-repetition changes in syllable length for *Consistency*) that may aid in their perception. Indeed, our study revealed moderate to strong correlations for both *Speed (F2Slope)* (prior to controlling for severity) and *Consistency (RepVar.Syll)* and their perceptual correlates.

Although no study, to our knowledge, has investigated the auditory-perceptual correlates of *Coordination*, research on chewing has identified challenges with visually perceiving coordinative movements (Simione et al., 2016). Similar to Simione et al. (2016)'s study, our study revealed weak associations between *Coordination* (DurRatio) and its perceptual rating. The visual perception of *Coordination* may be challenging because there is no information beyond the jaw movement, such as the movements of the tongue or velum. Furthermore, the variations in movement may be small in amplitude and, therefore, difficult to perceive. Perceiving *Coordination* auditorily may be even more challenging because of the added phenomena of motor equivalence and quantal effects. Motor equivalence as it relates to acoustic-based motor control is the ability to achieve the same goal through different movement configurations (Kelso et al., 1984). Similarly, quantal theory explicates the nonlinear relations between vocal tract movement and speech acoustics, where in some cases, articulator movements do not result in corresponding acoustic changes (Stevens, 1972). For example, moving the jaw laterally while speaking may be a sign of dyscoordination (Laboissiere, Ostry, and Feldman, 1996), but this movement is likely to have minimal

impact on speech output. Although little is known about the extent to which these factors influence the percept of *Coordination*, these aspects of articulatory-acoustic relations raise the possibility that the level of analysis we are using for our framework (i.e., acoustic) may not fully convey perceptual information about dyscoordination. Thus, while a weak association could indicate poor clinical validity, this finding may instead be indicative of the challenges with perceiving *Coordination*.

Evidence for Construct Validity of Acoustic-Based Articulatory Features: Divergent Validity

Strong correlations between measures from different levels of analysis (e.g., acoustic with kinematic or acoustic with perceptual) are essential for establishing criterion validity; such relationships do not, however, demonstrate that our acoustic features are representing *distinct* constructs. To investigate construct divergence, we examined correlations between each pair of acoustic features. Therefore, divergence would be evidenced by weak associations for the correlation analyses.

Interestingly, even without severity as a covariate, our findings revealed divergence between five of the ten pairs of acoustic features. This finding supports the existence of multiple distinct constructs, particularly given that we had not yet accounted for severity. Upon introducing severity as a covariate, the majority of the moderate to strong relationships were reduced to weak relationships, which provides evidence that the measures represent distinct constructs and suggests that severity had been driving at least a portion of the initial relationships. The strong association that remained between *Speed (F2Slope)* and *Precision (ConVar.F2Slope)*, regardless of severity, was expected given that both measures are derived from F2 slope. Nevertheless, a larger sample size is needed to examine this relationship further and to determine if the divergence we found among the majority of acoustic pairs is statistically significant or meaningful. With a greater number of subjects, future work could also implement factor analyses, which would further inform the presence of divergence and multidimensionality of the framework components.

Clinical Implications

Our findings support the reliability, validity, and clinical utility of the proposed articulatory motor framework. Analytical validity was supported by moderate to strong correlations between all five acoustic

features and their kinematic correlates. Clinical validity was also supported by moderate to strong associations with clinician ratings for all features except *Coordination (DurRatio)* when severity remained in the model. While the weak correlation for *Coordination (DurRatio)* may indicate that our acoustic feature is unable to quantitate incoordination, it may more likely underscore an ongoing challenge with the perceptual judgement of articulatory function. Indeed, although our findings suggest that clinician perception is consistent with performance on the five acoustic features, the clinicians in our study had a significant amount of experience with perceptually characterizing dysarthria subtypes. In real-world clinical settings, clinicians have varying levels of experience with speech motor populations, which biases their judgements of different features from the DAB model. As a result, there is often disagreement among healthcare professionals regarding differential diagnosis and appropriate treatment (Borrie et al., 2012; Bunton et al., 2007; Kent, 1996). The accuracy of DAB characterizations could be further impacted when features are particularly subtle, such as in more mild cases of dysarthria (Allison et al., 2017; Rowe et al., 2020). Given the analytical validity demonstrated for the five acoustic features, our acoustic-based framework could provide clinicians with a more accurate and reliable means of characterizing and assessing speech motor function.

Limitations and Future Directions

A primary limitation of our study is its small sample size. Given that the kinematic analyses were performed on only half of the subjects, the associations found with the corresponding acoustic measures should be interpreted with caution, as they could change with additional data. Indeed, to assess analytical validity, we only used the acoustic data from the 11 participants with kinematic data, which greatly reduced our sample size and, consequently, leads to a higher likelihood of finding spurious correlations. Further research is needed with larger sample sizes to determine whether our correlations remain.

In addition, because we used data from an existing database, our study was limited in the demographic and descriptive information that was available for the subjects. The XRMB database did not include data on timing or site of disease onset, cognitive function, or bulbar status, which is often examined using the speech item from the ALS Functional Rating Scale – Revised (ALSFRS-R; Cedarbaum et al., 1999). In the absence of clinical measures of bulbar involvement, we used clinicians' VAS ratings of

severity. While prior work provides favorable evidence for the use of VAS ratings of intelligibility in research (Abur et al., 2019; Hustad, 2006; Stipancic et al., 2016), less literature is available on VAS ratings of *severity*. Furthermore, severity or intelligibility ratings are typically performed by the clinician, whereas the ALSFRS-R is a self-report measure. Additional research is, therefore, necessary to determine the relationship between VAS severity ratings and the current clinical standard of the ALSFRS-R. Moreover, because the XRMB dataset did not include the SIT, the continuous speech stimuli for assessing severity and intelligibility were different between databases; the same utterance from the Hunter passage was used for all participants from the XRMB dataset, whereas the participants from the MGH IHP dataset had distinct sets of sentences. Since perceptual judgement is currently the clinical standard for assessing severity, it is difficult to know whether the datasets truly were comparable in severity or whether familiarity with the utterances influenced clinicians' ratings of participants in the XRMB dataset.

Lastly, it is important to acknowledge the challenge of assessing criterion validity with an imperfect standard (i.e., perceptual judgement). Indeed, perceptual judgements can be unreliable and vulnerable to subjective biases (Borrie et al., 2012; Bunton et al., 2007; Kent, 1996). Thus, further evaluation of other standard accuracy characteristics, such as sensitivity and specificity, is needed to improve confidence in the clinical utility of new speech measures.

Conclusions

The results of this study demonstrate that our acoustically driven framework has potential as an objective, valid, and clinically useful tool for profiling articulatory deficits in individuals with speech motor disorders. Our findings also suggest that compared to clinician ratings, instrumental measures may be more sensitive to subtle differences in articulatory function. With further research, this framework could provide accurate and reliable characterizations of articulatory impairment, which may eventually increase the efficacy of diagnosis and treatment for patients with different articulatory phenotypes.

ACKNOWLEDGEMENTS

The authors acknowledge the grants awarded by the NIDCD, part of the NIH: R01DC009890 (Principal Investigators [PIs]: Jordan R. Green, Yana Yunusova), R01DC013547 (PI: Jordan R. Green), R01DC017291 (PIs: Yana Yunusova, Jordan R. Green), K24DC016312 (PI: Jordan R. Green), and F31DC019556 (PI: Hannah P. Rowe).

REFERENCES

Abur, D., Enos, N.M., & Stepp, C.A. (2019). Visual analog scale ratings and orthographic transcription measures of sentence intelligibility in Parkinson's disease with variable listener exposure. *American Journal of Speech Language Pathology*, *28*(3), 1222-1232.

Adams, S.G., Welsmer, G., & Kent, R.D. (1993). Speaking rate and speech movement velocity profiles. *Journal of Speech and Hearing Research*, *36*(1), 41-54.

Allison, K.M., Yunusova, Y., Campbell, T.F., Wang, J., Berry, J.D., & Green, J.R. (2017). The diagnostic utility of patient-report and speech-language pathologists' ratings for detecting the early onset of bulbar symptoms due to ALS. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, *18*(5-6), 358-366.

Allison, K.M., Cordella, C., Iuzzini-Seigel, J., & Green, J.R. (2020). Differential diagnosis of apraxia of speech in children and adults: A scoping review. *Journal of Speech, Language, and Hearing Research*, *63*(9), 2952-2994.

Audacity Team (2020). Audacity(R): Free Audio Editor and Recorder [Computer application]. Version 2.4.2 retrieved Sep 20th 2020 from https://audacityteam.org/.

Baehner, F.L. (2016). The analytical validation of the Oncotype DX Recurrence Score assay. *Ecancermedicalscience*, *10*(675), 1-11.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5*(9), 341-345.

Borrie, S.A., McAuliffe, M.J., & Liss, J.M. (2012). Perceptual learning of dysarthric speech: A review of experimental studies. *Journal of Speech, Language, and Hearing Research*, *55*(1), 290-305.

Bunton, K., Kent, R.D., Duffy, J.R., Rosenbek, J.C., & Kent, J.F. (2007). Listener agreement for

auditory-perceptual ratings of dysarthria. *Journal of Speech, Language, and Hearing Research*, *50*(6), 1481-1495.

Caruso, A.J., Abs, J.H., & Gracco, V.L. (1988). Kinematic analysis of multiple movement coordination during speech in stutterers. *Brain*, *111*, 439-455.

Cedarbaum, J.M., Stambler, N., Malta, E., Fuller, C., Hilt, D., Thurmond, B., Nakanishi, A., & BDNF ALS Study Group (Phase III) (1999). The ALSFRS-R: A revised ALS functional rating scale that incorporates assessments of respiratory function. *Journal of Neurological Sciences*, *169*(1-2), 13-21.

Chen, W.S. & Alwan, A. (2000). Place of articulation cues for voiced and voiceless plosives and fricatives in syllable-initial position. *Proceedings of Sixth International Conference on Spoken Language Processing*, 113-116.

Darley, F.L., Aronson, A.E., & Brown, J.R. (1969a). Differential diagnostic patterns of dysarthria. *Journal of Speech and Hearing Research*, *12*(2), 246-269.

Darley, F.L., Aronson, A.E., & Brown, J.R. (1969b). Cluster of deviant speech dimensions in the dysarthrias. *Journal of Speech, Language, and Hearing Research*, *12*(3), 462-496.

Delattre, P.C., Liberman, A.M., & Cooper, F.S. (1955). Acoustic loci and transitional cues for consonants. *Journal of the Acoustical Society of America*, *27*(4), 769-773.

Eklund, E., Ovist, J., Sandstrom, L., Viklund, F., Van Doorn, J., & Karlsson, F. (2015). Perceived articulatory precision in patients with Parkinson's disease after deep brain stimulation of subthalamic nucleus and caudal zona incerta. *Clinical Linguistics & Phonetics*, *29*(2), 150-166.

Enderby, P. & Palmer, R. (2008). Frenchay Dysarthria Assessment–Second Edition (FDA-2) [Measurement instrument]. Austin, TX: Pro-Ed, Inc.

Feenaughty, L., Tjaden, K., & Sussman, J. (2014). Relationship between acoustic measures and judgments of intelligibility in Parkinson's disease: A within-speaker approach. *Clinical Linguistics & Phonetics*, *28*(11), 857-878.

Fitts, P.M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, *47*(6), 381-391.

Fletcher, S.G. (1972). Time-by-count measurement of diadochokinetic syllable rate. *Journal of Speech,*

Language, and Hearing Research, *15*(4), 763-770.

Folker, J., Murdoch, B., Cahill, L., Delatycki, M., Corben, L., & Vogel, A. (2010). Dysarthria in Friedreich's ataxia: A perceptual analysis. *Folia Phoniatricia et Logopaedica*, *62*(3), 97-103.

Gracco, V.L. & Abbs, J.H. (1986). Variant and invariant characteristics of speech movements. *Experimental Brain Research*, *65*(1), 156-166.

Green, J.R., Moore, C.A., Higashikawa, M., & Steeve, R.W. (2002). The physiologic development of speech motor control. *Journal of Speech, Language, and Hearing Research*, *43*(1), 239-255.

Green, J.R., Wang, J., & Wilson, D.L. (2013). SMASH: A tool for articulatory data processing and analysis. *Proceedings of Interspeech,* 1331-1335.

Green, J.R., Yunusova, Y., Kuruvilla, M.S., Wang, J., Pattee, G.L., Synhorst, L., Zinman, L., & Berry, J.D. (2013). Bulbar and speech motor assessment in ALS: Challenges and future directions. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, *14*(7-8), 494-500.

Green, J.R. (2015). Mouth matters: Scientific and clinical applications of speech movement analysis. *Perspectives on Speech Science and Orofacial Disorders*, *25*(1), 6-16.

Grosjean, F. & Lane, H. (1981). Temporal variables in the perception and production of spoken and sign languages. In Eimas, P. and Miller, J. (Eds.), *Perspectives on the Study of Speech* (pp. 207-234). Hillsdale, New Jersey: Lawrence Erlbaum.

Gupta, R., Chaspari, T., Kim, J., Kumar, N., Bone, D., & Narayanan, S. (2016). Pathological speech processing: State-of-the-art, current challenges, and future directions. *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, 6470-6474.

Harris, P.A., Taylor, R., Minor, B.L., Elliott, V., Fernandez, M., O'Neal, L., McLeod, L., Delacqua, G., Delacqua, F., Kirby, J., & Duda, S.N. (2019). REDCap Consortium, The REDCap consortium: Building an international community of software partners. *Journal of Biomedical Informatics*, *95*, 1-24.

Heman-Ackah, Y.D., Michael, D.D., & Goding, G.S. (2002). The relationship between cepstral peak prominence and selected parameters of dysphonia. *Journal of Voice*, *16*(1), 20-27.

Hustad, K.C. (2006). Estimating the intelligibility of speakers with dysarthria. *Folia Phoniatricia et*

Logopaedica, *58*(3), 217-228.

Jiao, Y., Berisha, V., Liss, J., Hsu, S.C., Levy, E., & McAuliffe, M. (2016). Articulation entropy: An unsupervised measure of articulatory precision. *IEEE Signal Processing Letters*, *24*(4), 485-489.

Kelso, J.S., Tuller, B., Vatikiotis-Bateson, E., & Fowler, C. A. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(6), 812-832.

Kent, R.D. & Rosenbek, J.C. (1983). Acoustic patterns of apraxia of speech. *Journal of Speech and Hearing Research*, *26*(2), 231-249.

Kent, R.D., Kent, J.F., Weismer, G., Martin, R., Sufit, R.L., Brooks, B.R., & Rosenbek, J.C. (1989). Relationships between speech intelligibility and the slope of second formant transitions in dysarthric subjects. *Clinical Linguistics & Phonetics*, *3*(4), 347-358.

Kent, R.D. (1996). Hearing and believing: Some limits to the auditory-perceptual assessment of speech and voice disorders. *American Journal of Speech-Language Pathology*, *5*(3), 7-23.

Kent, R.D. & Kim, Y. (2003). Toward an acoustic typology of motor speech disorders. *Clinical Linguistics & Phonetics*, *16*(6), 427-445.

Ketcham, C.J. & Stelmach, G.E. (2003). Movement control in the older adult. In R.W. Pew & S.B. Van Hemel (Eds.), *Technology for adaptive aging* (pp. 64-92). Washington, DC: The National Academies Press.

Kim, H., Hasegawa-Johnson, M., & Perlman, A. (2011). Temporal and spectral characteristics of fricatives in dysarthria. *Journal of the Acoustical Society of America*, *130*(4), 2446-2446.

Kim, Y., 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.

Kleinow, J., Smith, A., & Ramig, L.O. (2001). Speech motor stability in IPD: Effects of rate and loudness manipulations. *Journal of Speech, Language, and Hearing Research*, *44*(5), 1041-1051.

Laboissiere, R., Ostry, R.J., & Feldman, A.G. (1996). The control of multi-muscle systems: Human jaw and hyoid movements. *Biological Cybernetics*, *74*, 373-384.

Lammert, A.C., Shadle, C.H., Narayanan, S.S., & Quatieri, T.F. (2018). Speed-accuracy tradeoffs in

782 human-speech production. *PLOS One*, *13*(9), 1-25.
783 Lansford, K.L., Liss, J.M., & Norton, R.E. (2013). Free-classification of perceptually similar speakers with
784 dysarthria. *Journal of Speech, Language, and Hearing Research*, *57*(6), 2051-2064.
785 Lansford, K.L. & Liss, J.M. (2014). Vowel acoustics in dysarthria: Mapping to perception. *Journal of*
786 *Speech, Language, and Hearing Research*, *57*(1), 68-80.
787 Lee, J., Hustad, K.C., & Weismer, G. (2014). Predicting speech intelligibility with a multiple speech
788 subsystems approach in children with cerebral palsy. *Journal of Speech, Language, and Hearing*
789 *Research*, *57*(5), 1666-1678.
790 MacKay, I. (2014). *Acoustics in Hearing, Speech and Language Sciences: An Introduction, Loose-Leaf*
791 *Version (Allyn & Bacon Communication Sciences and Disorders).* Boston, MA: Pearson.
792 MathWorks (2019). Matlab optimization toolbox (r2019a) [Computer software]. Natick, MA: The
793 MathWorks.
794 McNeil, M.R., Odell, K.H., Miller, S.B., & Hunter, L. (1995). Consistency, variability and target
795 approximation for successive speech repetitions among apraxic, conduction aphasic and ataxic
796 dysarthric speakers. *Clinical Aphasiology*, *23*, 39-55.
797 Mefferd, A.S., Green, J.R., & Pattee, G.L. (2014). Speaking rate effects on articulatory pattern variability in
798 talkers with mild ALS. *Clinical Linguistics and Phonetics*, *28*(11), 799-811.
799 Miller, N. (2013). Measuring up to speech intelligibility. *International Journal of Language and*
800 *Communication Disorders*, *48*(6), 601-612.
801 Nishio, M. & Nimi, S. (2006). Comparison of speaking rate, articulation rate and alternating motion rate in
802 dysarthric speakers. *Folia Phoniatrica et Logopaedica*, *58*(2), 114-131.
803 Owren, M.J. & Cardillo, G.C. (2006). The relative roles of vowels and consonants in discriminating talker
804 identity versus word meaning. *Journal of the Acoustical Society of America*, *119*(3), 1727-1739.
805 Quatieri, T.F., Talkar, T., & Palmer, J.S. (2020). A framework for biomarkers of covid-19 based on
806 coordination of speech-production subsystems. *IEEE Open Journal of Engineering in Medicine and*
807 *Biology*, *1*, 203-206.
808 R Core Team (2014). R: A language and environment for statistical computing [Computer software].

809 Vienna, Austria: R Foundation for Statistical Computing.
810 Rabinov, C.R., Kreiman, J., Gerratt, B.R., & Bielamowicz, S. (1995). Comparing reliability of perceptual
811 ratings of roughness and acoustic measures of jitter. *Journal of Speech and Hearing Research*,
812 *38*(1), 26-32.
813 Rong, P., Yunusova, Y., Wang, J., & Green, J.R. (2015). Predicting early bulbar decline in amyotrophic
814 lateral sclerosis: A speech subsystem approach. *Behavioural Neurology*, *215*, 1-11.
815 Rong, P., Yunusova, Y., Richburg, B., & Green, J.R. (2018). Automatic extraction of abnormal lip
816 movement features from the alternating motion rate task in amyotrophic lateral sclerosis.
817 *International Journal of Speech-Language Pathology*, *20*(6), 610-623.
818 Rong, P. (2020). Automated acoustic analysis of oral diadochokinesis to assess bulbar motor involvement in
819 amyotrophic lateral sclerosis. *Journal of Speech, Language, and Hearing Research*, *63*(1), 1-15.
820 Rowe, H. & Green, J.R. (2019). Profiling speech motor impairments in persons with amyotrophic lateral
821 sclerosis: An acoustic-based approach. *Proceedings of Interspeech*, 4509-4513.
822 Rowe, H.P., Gutz, S.E., Maffei, M., & Green, J.R. (2020). Acoustic-based articulatory phenotypes of
823 amyotrophic lateral sclerosis and Parkinson's disease: Towards an interpretable, hypothesis-driven
824 framework of motor control. *Proceedings of Interspeech*, 4816-4820.
825 Rusz, J., Benova, B., Ruzickova, H., Novotny, M., Tykalova, T., Hlavnicka, J., Uher, T., Vaneckova, M.,
826 Andelova, M., Novotna, K., Kadrnozkova, L., & Horakova, D. (2018). Characteristics of motor
827 speech phenotypes in multiple sclerosis. *Multiple Sclerosis and Related Disorders*, *19*, 62-69.
828 Sapir, S., Spielman, J.L., Ramig, L.O., Story, B.H., & Fox, C. (2007). Effects of intensive voice treatment
829 (the Lee Silverman Voice Treatment [LSVT]) on vowel articulation in dysarthric individuals with
830 idiopathic Parkinson disease: Acoustic and perceptual findings. *Journal of Speech, Language, and*
831 *Hearing Sciences*, *50*(4), 899-912.
832 Skodda, S., Visser, W., & Schlegel, U. (2011). Acoustical analysis of speech in progressive supranuclear
833 palsy. *Journal of Voice*, *25*(6), 725-731.
834 Sidtis, J.J., Ahn, J.S., Gomez, C., & Sidtis, D. (2011). Speech characteristics associated with three
835 phenotypes of ataxia. *Journal of Communication Disorders*, *44*(4), 478-492.

836 Simione, M., Wilson, E.M., Yunusova, Y., & Green, J.R. (2016). Validation of clinical observations of
837 mastication in persons with ALS. *Dysphagia*, *31*(3), 367-375.
838 Stevens, K.N. (1972). The quantal nature of speech: Evidence from articulatory-acoustic data. In P.B. Denes
839 & E.E. David Jr. (Eds.), *Human communication: A unified view* (pp. 51–66). New York:
840 McGraw-Hill.
841 Stipancic, K.L., Tjaden, K., & Wilding, G. (2016). Comparison of intelligibility measures for adults with
842 Parkinson's disease, adults with multiple sclerosis, and healthy controls. *Journal of Speech,*
843 *Language, and Hearing Research*, *59*(2), 230-238.
844 Stipancic, K. L., Yunusova, Y., Campbell, T. F., Wang, J., Berry, J. D., & Green, J. R. (2021). Two distinct
845 clinical phenotypes of bulbar motor impairment in amyotrophic lateral sclerosis. *Frontiers in*
846 *Neurology – NeuroRehabilitation*, *12*, 1-9.
847 Strand, E.A. (2013). Neurologic substrates of motor speech disorders. *Perspectives on Neurophysiology and*
848 *Neurogenic Speech and Language Disorders*, *23*(3), 98-104.
849 Tjaden K. (2000). A preliminary study of factors influencing perception of articulatory rate in Parkinson
850 disease. *Journal of Speech, Language, and Hearing Research*, *43*(4), 997-1010.
851 Tjaden, K.T. & Watling, E. (2003). Characteristics of diadochokinesis in multiple sclerosis and Parkinson's
852 disease. *Folia Phoniatrica et Logopaedica*, *55*(5), 241-259.
853 Tjaden, K., Sussman, J.E., & Wilding, G.E. (2014). Impact of clear, loud, and slow speech on scaled
854 intelligibility and speech severity in Parkinson's disease and multiple sclerosis. *Journal of Speech,*
855 *Language, and Hearing Research*, *57*(3), 779-792.
856 Turner, G.S. & Weismer, G. (1993). Characteristics of speaking rate in the dysarthria associated with
857 amyotrophic lateral sclerosis. *Journal of Speech & Hearing Research*, *36*(6), 1134-1144.
858 Hlavnicka, J., Cmejla, R., Tykalova, T., Sonka, K., Ruzicka, E., & Rusz, J. (2017). Automated analysis of
859 connected speech reveals early biomarkers of Parkinson's disease in patients. *Nature*, *7*(12), 1-11.
860 Waito, A.A., Wehbe, F., Marzouqah, R., Barnett, C., Shellikeri, S., Cui, C., Abrahao, A., Zinman, L., Green,
861 J.R., & Yunusova, Y. (2021). Validation of articulatory rate and imprecision judgments in speech of
862 individuals with amyotrophic lateral sclerosis. *American Journal of Speech-Language Pathology*,

30(1), 137-149.

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

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

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*(5), 1247–1261.

Westbury, J. (1994). *X-Ray Microbeam Speech Production Database User's Handbook (Version 1.0)*. Madison, WI: Waisman Center on Mental Retardation & Human Development, University of Wisconsin.

Whitfield, J.A. & Goberman, A.M. (2014). Articulatory-acoustic vowel space: Application to clear speech in individuals with Parkinson's disease. *Journal of Communication Disorders*, *51*, 19-28.

Williamson, J.R., Young, D., Nierenberg, A.A., Niemi, J., Helfer, B.S., & Quatieri, T.F. (2019). Tracking depression severity from audio and video based on speech articulatory coordination. *Computer Speech and Language*, *55*, 40-56.

Wooley, S.C., York, M.K., Moore, D.H., Strutt, A.M., Murphy, J., Schulz, P.E., & Katz, J.S. (2010). Detecting frontotemporal dysfunction in ALS: Utility of the ALS Cognitive Behavioral Screen (ALS-CBS). *Amyotrophic Lateral Sclerosis*, *11*(3), 303-311.

Yorkston, K.M., Beukelman, D.R., Hakel, M., & Dorsey, M. (2007). Speech Intelligibility Test for Windows [Measurement instrument]. Lincoln, NE: Institute for Rehabilitation Science and Engineering at Madonna Rehabilitation Hospital.

Yorkston, K.M., Hakel, M., Beukelman, D.R., & Fager, S. (2007). Evidence for effectiveness of treatment of loudness, rate, or prosody in dysarthria: A systematic review. *Journal of Medical Speech-Language Pathology*, *15*(2), 11-36.

Yumoto, E., Sasaki, Y., & Okamura, H. (1984). Harmonics-to-noise ratio and psychophysical measurement

of the degree of hoarseness. *Journal of Speech, Language, and Hearing Research*, *27*(1), 2-6.

Yunusova, Y., Green, J.R., Greenwood, L., Wang, J., Pattee, G.L., & Zinman, L. (2012). Tongue movements and their acoustic consequences in amyotrophic lateral sclerosis. *Folia Phoniatrica Et Logopaedica*, *64*(2), 94-102.

Ziegler, W. & Wessel, K. (1996). Speech timing in ataxic disorders: Sentence production and rapid repetitive articulation. *Neurology*, *47*(1), 208-214.