Research Note

A Phonetic Complexity-Based Approach for Intelligibility and Articulatory Precision Testing: A Preliminary Study on Talkers With Amyotrophic Lateral Sclerosis

Mili Kuruvilla-Dugdale,[a] Claire Custer,[a] Lindsey Heidrick,[b] Richard Barohn,[c] and Raghav Govindarajan[d]

Purpose: This study describes a phonetic complexity-based approach for speech intelligibility and articulatory precision testing using preliminary data from talkers with amyotrophic lateral sclerosis.
Method: Eight talkers with amyotrophic lateral sclerosis and 8 healthy controls produced a list of 16 low and high complexity words. Sixty-four listeners judged the samples for intelligibility, and 2 trained listeners completed phoneme-level analysis to determine articulatory precision. To estimate percent intelligibility, listeners orthographically transcribed each word, and the transcriptions were scored as being either accurate or inaccurate. Percent articulatory precision was calculated based on the experienced listeners' judgments of phoneme distortions, deletions, additions, and/or substitutions for each word.

Articulation errors were weighted based on the perceived impact on intelligibility to determine word-level precision.
Results: Between-groups differences in word intelligibility and articulatory precision were significant at lower levels of phonetic complexity as dysarthria severity increased. Specifically, more severely impaired talkers showed significant reductions in word intelligibility and precision at both complexity levels, whereas those with milder speech impairments displayed intelligibility reductions only for more complex words. Articulatory precision was less sensitive to mild dysarthria compared to speech intelligibility for the proposed complexity-based approach.
Conclusions: Considering phonetic complexity for dysarthria tests could result in more sensitive assessments for detecting and monitoring dysarthria progression.

T he current clinical gold standard to detect and track dysarthria progression in talkers with amyotrophic lateral sclerosis (ALS) involves a brief perceptual evaluation based on sentence or word reading to determine speech intelligibility and rate. However, these measures only provide an overall index of severity and are relatively insensitive to dysarthria onset, as suggested by prior ALS literature (Ball, Beukelman, & Pattee, 2002;

Ball, Willis, Beukelman, & Pattee, 2001). Moreover, because these clinical speech measures are coarse, it is difficult to identify and track early changes in articulatory impairments, which contribute significantly to dysarthric speech in the form of consonant imprecision, vowel distortion, and reduced rate. Despite these limitations, intelligibility and rate are widely used clinically to time speech interventions and introduce assistive communication options for those with progressive speech loss due to ALS. It is, therefore, crucial to identify sensitive stimuli that can be used to detect subtle, incremental changes in intelligibility and articulatory precision, so clinicians can accurately monitor dysarthria progression to improve the timing of speech interventions.

Phonetic complexity, characterized by the articulatory motor adjustments required for consonant and vowel production, may be an important attribute to consider when identifying sensitive test stimuli, because speech motor problems may be best observed during complex movements (Duffy, 2013; Forrest, Weismer, & Turner, 1989). Indeed,

[a]Department of Communication Science and Disorders, University of Missouri, Columbia
[b]Department of Hearing and Speech, University of Kansas Medical Center, Kansas City
[c]Department of Neurology, University of Kansas Medical Center, Kansas City
[d]Department of Neurology, University of Missouri, Columbia
Correspondence to Mili Kuruvilla-Dugdale: kuruvillam@missouri.edu

the earliest report on phonetic complexity effects showed that differences between typical speakers and those with Parkinson's disease, for kinematic and acoustic events, were greater for speech segments that were higher in complexity (Forrest et al., 1989). Similarly, in talkers with multiple sclerosis, phonetic contexts that require rapid or radical constrictions of the vocal tract were found to be more sensitive to dysarthria than other contexts (Rosen, Goozée, & Murdoch, 2008). More recent research on talkers with nonprogressive dysarthria due to cerebral palsy (CP) demonstrated that later acquired sounds, which are motorically complex targets, had a greater negative impact on speech intelligibility than earlier acquired, less complex targets (Allison & Hustad, 2014). Notably, these researchers reported a 6%–17% drop in intelligibility with increased phonetic complexity and sentence length in children with dysarthria secondary to CP (Allison & Hustad, 2014). Another recent study on how phonetic complexity affects the frequency of articulation errors showed that young adults with CP misarticulated complex consonants more often than less complex consonants (Kim, Martin, Hasegawa-Johnson, & Perlman, 2010). In addition, speakers with low intelligibility were found to reduce the articulatory complexity of target consonants more frequently than those with high intelligibility (Kim et al., 2010). Growing evidence for the detrimental effects of phonetic complexity on speech production suggests that it is an important factor to consider in the design of assessment and treatment stimuli because complex stimuli are challenging for those with motor speech disorders to produce.

Phonetic complexity of individual speech sounds has been estimated in different ways. In the framework proposed by R. D. Kent (1992), phonetic complexity is characterized based on the articulatory motor adjustments required to produce various consonants and vowels. According to this framework, sounds acquired at an early age, such as bilabials (e.g., /b/), require simple motor adjustments like velopharyngeal valving or regulation of articulatory movement speed. Sounds acquired later require more refined articulatory skills and a complex pattern of interarticulator coordination, for example, tongue configuration and fine force regulation for fricative production. Using these principles, researchers have assigned consonants and vowels into different levels of phonetic complexity based on their speech motor demands (Allison & Hustad, 2014; Kim et al., 2010). Specifically, vowels have four levels and consonants have five levels of complexity, where higher numbers correspond with greater speech motor complexity. For any given word, the complexity levels of its constituent consonants and vowels are summed up to determine its overall complexity. For example, for the word *unerring* (/ənerɪŋ/), the complexity levels of its consonants are 3, 5, and 5, respectively; vowels are 1, 4, and 2, respectively, giving the word an overall score of 20. Importantly, when this complexity framework was applied to dysarthria, researchers observed that later acquired sounds, which are motorically complex targets, were more commonly misarticulated by even the most intelligible adult talkers with dysarthria and had a

greater negative impact on speech intelligibility than earlier acquired, less complex targets (Allison & Hustad, 2014; Kim et al., 2010). Another classification system for phonetic complexity is based on the frame–content theory advanced by MacNeilage and Davis (1990). According to their theory of speech evolution, cyclical closing and opening of the jaw is the basic movement or motor frame upon which content (i.e., more complex consonant vowel patterns involving the other articulators) is superimposed, to fulfill complex communication needs. In other words, as children slowly acquire coordination and control over articulators, they produce different consonants or vowels (i.e., content) that are differentiated from the basic jaw-based frame. The complexity metric based on this theory was applied to speech from adults who stutter and revealed an increase in stuttering rates for phonetically complex content words (Dworzynski & Howell, 2004). In the current study, we used the classification system recapitulated from R. D. Kent (1992) that has been successfully applied to demonstrate the effects of phonetic complexity at the auditory–perceptual level in adults and children with nonprogressive dysarthria.

Articulatory imprecision is a hallmark characteristic of progressive dysarthria and contributes significantly to the intelligibility deficit. However, articulatory errors are not routinely assessed and monitored in clinic as a way to determine dysarthria onset and progression. If included as part of assessment, clinicians and researchers typically use interval rating scales to indicate the severity of articulatory problems such as imprecise consonants, distorted vowels, and prolonged phonemes. Interrater agreement and reliability for articulatory dimensions such as irregular articulatory breakdowns and distorted vowels are susceptible to moderate levels of listener disagreement (Sheard, Adams, & Davis, 1991). Relatedly, prior research shows a poor correspondence for consonant imprecision ratings between tasks, such as syllable repetition and passage reading in talkers with spastic dysarthria (Zeplin & Kent, 1996). As an alternate approach, researchers have used direct magnitude estimation to scale intelligibility solely based on articulatory precision (Weismer, Jeng, Laures, Kent, & Kent, 2001; Weismer, Laures, Jeng, Kent, & Kent, 2000). However, a scaled estimate of speech intelligibility as an index of articulatory precision is a separate entity than segmental articulation errors derived from phonetic transcription. The latter allows specification of the articulatory underpinnings of reduced intelligibility, in terms of phoneme additions, deletions, distortions, and/or substitutions (Bent, Bradlow, & Smith, 2007; Weismer, 2008). In contrast, the former does not tap into all the potential levels of articulatory problems and may only be marginally affected by the deterioration of individual sound segments (Weismer et al., 2000). One way to reduce the dissociation between articulatory error analyses and speech intelligibility ratings reported in prior studies (Platt, Andrews, & Howie, 1980; Platt, Andrews, Young, & Quinn, 1980; Whitehill, 2002) is to weight an individual articulation error based on its contribution to speech intelligibility. Such an approach will

provide not only more detailed information about the nature of the misarticulations, but also which errors interfere most with intelligibility. More importantly, the segmental analysis could be used by clinicians to target and remediate errors that have the most impact on speech intelligibility. Overall, phoneme-level analyses will allow us to increase the information obtained from auditory–perceptual judgments and, if tracked as a function of phonetic complexity, may help capture subtle articulatory problems as well as small, incremental articulatory changes that will help improve detection and monitoring of dysarthria progression.

Although phonetic complexity effects on speech performance have been studied using auditory–perceptual measures in nonprogressive dysarthria, no studies to date have examined these effects in talkers with progressive dysarthria. Therefore, the primary purpose of this study was to determine the diagnostic value of a complexity-based intelligibility approach using preliminary data from talkers with ALS who were stratified into two groups: those with mild speech impairments and those with more severe impairments. Specifically, this study aimed to investigate group differences in intelligibility as a function of phonetic complexity between healthy controls and the two ALS groups. In contrast to the current clinical approach for intelligibility testing where phonetically balanced test stimuli are routinely used, our approach uses words based on their phonetic complexity. This may allow us to capture slight changes in intelligibility and improve the timing of speech treatments, without missing the critical window during which these options must be introduced to be most successful and yield the greatest impact on quality of life.

Because articulatory impairments are known to contribute significantly to reduced intelligibility, assessing articulatory precision across phonetic complexity categories can supplement more traditional methods of tracking dysarthria progression. Therefore, the second aim of our study was to evaluate a complexity-based approach for identifying articulatory impairment, where articulation errors are weighted based on their contribution to speech intelligibility. Specifically, this study aimed to investigate group differences in articulatory precision as a function of phonetic complexity between healthy controls and those with mild dysarthria as well as more severe dysarthria. Based on previous research (Allison & Hustad, 2014; Forrest et al., 1989; Kim et al., 2010; Rosen et al., 2008), we hypothesized that between-groups differences in word intelligibility and articulatory precision would increase as phonetic complexity increased. Furthermore, phonetic complexity effects on word intelligibility and precision were expected to depend on dysarthria severity. More specifically, significant between-groups differences were predicted only for more complex words in less impaired talkers, whereas more severely impaired talkers were expected to have significantly lower intelligibility and articulatory precision scores compared to controls, even for low complexity words. Lastly, we predicted that sentence intelligibility scores based on the Sentence Intelligibility Test (SIT; K. Yorkston, Beukelman, Hakel, & Dorsey, 2007) would be significantly different

only between controls and more severely impaired talkers, which is in contrast to the predictions for our proposed approach.

Method

Participants

Speakers

Eight adults with a diagnosis of definite ALS (five men, three women) and eight healthy adults (five men, three women) were recruited for the study. The specific inclusionary criteria for all speakers in the study were as follows: (a) be a native speaker of American English; (b) have no prior history of speech, language, and hearing problems; and (c) have no existing neurological diagnosis such as dementia, history of vascular disease, and/or trauma to the head. For the ALS group, participants were included if they had (a) a diagnosis of definite ALS with either bulbar or spinal onset and (b) dysarthria types typically seen in the ALS population, that is, flaccid, spastic, or mixed flaccid–spastic dysarthria.

The Montréal Cognitive Assessment (Nasreddine et al., 2005) and SIT (K. Yorkston et al., 2007) were used to assess the cognitive and speech abilities of each participant (see Table 1 for details). The Montréal Cognitive Assessment was not administered on two participants with ALS due to time constraints, but both participants and their family members denied a cognitive impairment. No formal hearing screening was conducted, but participants were able to follow task instructions and conversation at normal loudness levels without any difficulty. Participants with ALS were broadly classified as being mildly impaired or more severely impaired by taking into account both their SIT-based intelligibility and speaking rate scores (see Table 1).

Listeners

Sixty-four undergraduate speech-language pathology students who were blinded to the group membership of the speakers (i.e., dysarthria vs. healthy controls) and with minimal experience with motor speech disorders performed the word intelligibility ratings. Sentence intelligibility was calculated from the orthographic transcriptions of SIT sentences completed by four inexperienced listeners who were not involved in the word transcription task. For articulatory precision judgments, two highly trained listeners with moderate experience in motor speech disorders completed the ratings. Listeners were included in the study if they (a) had no prior history of a speech, language, or hearing impairment; (b) were native speakers of American English; (c) were between 18 and 40 years of age; and (d) had no history of neurological, vascular, and/or traumatic insults. All listeners passed a hearing screening at 500 Hz, 1 kHz, 2 kHz, 4 kHz, and 8 kHz. None of the listeners self-identified speech, language, and/or cognitive impairments. Listeners were all women (mean age = 21.76 years, SD = 2.33).

Table 1. Biographical details and assessment outcomes of participants.

Group	Gender	Age (years) M (SD)	MoCA M (SD)	Sentence intelligibility (%) M (SD)	Speaking rate (words per minute) M (SD)	Dysarthria severity
Controls	3F, 5M	60.81 (9.81)	24.67 (1.22)	99.20 (1.25)	188.49 (23.15)	Normal speech
ALS	3F, 5M	63.20 (10.18)	23.85 (3.84)	80.19 (25.31)	146.33 (57.42)	Mild–severe
ALS 1	F	48.06	24	96.36	210.91	Mild
ALS 2	M	54.06	28	98.48	230.27	Mild
ALS 3	M	65.10	19	96.33	164.39	Mild–moderate
ALS 4	F	55.08	—	99.10	161.40	Mild–moderate
ALS 5	F	73.09	18	86.97	116.63	Moderate
ALS 6	M	68.04	—	79.71	137.98	Moderate
ALS 7	M	78.08	26	55.55	79.57	Severe
ALS 8	M	64.08	26	29.10	69.5	Severe

Note. MoCA = Montréal Cognitive Assessment; ALS = amyotrophic lateral sclerosis; F = female; M = male.

Word Stimuli by Phonetic Complexity

For this study, we used 16 words from the Hoosier Mental Lexicon (Nusbaum, Pisoni, & Davis, 1984) that were originally selected as part of a larger study to determine the effects of lexical variables on language acquisition and speech motor control. The words were given a phonetic complexity score based on the consonant and vowel classification system recapitulated from R. D. Kent (1992) and used in recent phonetic complexity studies involving dysarthric speakers (Allison & Hustad, 2014; Kim et al., 2010). Each word was transcribed using the International Phonetic Alphabet, after which vowels, consonants, and consonant clusters that made up each word were given a score based on their speech motor complexity (see Table 2). Segmental complexity scores were summed for each word, and the words were then split into low and high complexity categories based on their overall score (see Table 3). Specifically, the low complexity words had an average score of 19 (range = 17–20), and the high complexity words had an average

score of 29.13 (range = 28–30). The low complexity words had an average length of 2.4 syllables (SD = 0.52), and the high complexity words had an average length of 3.6 syllables (SD = 0.52).

Because the words were part of a larger study about the influence of lexical properties, objective variables such as word frequency and phonotactic probability were well controlled for in this study. Word frequency obtained using SUBTLEX$_{US}$ (Brysbaert & New, 2009; available from http://subtlexus.lexique.org/) was 3.16 (SD = .75) for the low complexity words and 3.25 (SD = 1.02) for the high complexity words. Phonotactic probability determined using the KU phonotactic probability calculator (Vitevitch & Luce, 2004; available from https://phonotactic.drupal.ku.edu/) was .005 (SD = .003) for both the low and high complexity words. Overall, both the low and high complexity words were very similar in terms of word frequency and phonotactic probability.

Regarding syllable strength, all the low complexity words had stressed initial syllables, whereas three of the

Table 2. Classification of vowels and consonants based on articulatory motor complexity as proposed by R. D. Kent (1992).

Complexity level[a]	Phonemes	Underlying articulatory motor adjustments
Vowels		
1	/ɑ, ə/	Slight elevation of tongue from low carriage; mastery of back and forth movements of the tongue
2	/u, i, o/	Maximally dissimilar vowels develop creating a triangle for the boundaries of vowel production
3	/ɛ, ɔ, ai, au, ɔi/	Vowel triangle shifts to quadrilateral with appearance of vowel /ɛ/; differentiation of similar sounds /ɑ/ and /ɔ/; tongue movement precision for gliding of diphthongs
4	/I, e, æ, ʊ/	Coordination of tongue and mandible for front vowels
5	/ɝ, ɚ/	Retroflexion of the tongue
Consonants		
3	/p, m, n, w, h/	Certain rapid, high velocity movements (/p, m, n/) & steady, low velocity movements (/w, h/) Velopharyngeal valving (stops and nasals); bilabial, alveolar, and glottal places of articulation
4	/b, d, k, g, j, f/	Additional rapid, high velocity movements (/b, d, k, g/) & steady, low velocity movements (/j/) Control of fine force for frication (/f/); additional place of articulation (velar)
5	/t, ŋ, r, l/	Additional rapid, high velocity movements (/t, ŋ/); tongue positioning for /r, l/
6	/s, z, ʃ, v, θ, ð, tʃ, dʒ/	Tongue positioning and control of fine force for dental, alveolar, and palatal fricatives
7	2 consonant clusters	Fine motor control to transition quickly and efficiently between articulatory placements
8	3 consonant clusters	Fine motor control to transition to three different articulatory placements successively

[a]Complexity starts at Level 1 for vowels and Level 3 for consonants based on the age of acquisition for the earliest acquired vowels and consonants.

Table 3. Words characterized by phonetic complexity.

Low complexity	High complexity
Entourage	Absolutely
Faction	Community
Meanest	Complicate
Ocular	Dishevel
Phantom	Hospitable
Twinkle	Periwinkle
Unerring	Recompense
Union	Unanimous

eight high complexity words had unstressed initial syllables with primary stress on the second syllable. Syllable stress for each word was determined by the second author and confirmed using the MRC psycholinguistic database (Wilson, 1988). Subjective lexicosemantic variables such as word familiarity and predictability were not carefully controlled for in this study; however, studies show good overall correlation between objective word frequency, which we controlled for, and these subjective measures (Gordon, 1985; Nusbaum et al., 1984).

Data Collection

Acquisition of Speech Samples

Participants were asked to repeat the word stimuli, which were presented through loudspeakers, so that the presentation mode was consistent across speakers, irrespective of their reading skills and/or dysarthria severity. Because the data were collected as part of a larger study on speech kinematics, each speaker produced 11 lists of the target words, randomized in a different order for each list. Productions were recorded on a solid state digital recorder (Marantz PMD670) via a condenser microphone (Shure, PG42) placed 20 cm from each participant. The sampling rate for the audio signals was 22,000 Hz. Of the 11 lists, one list from the middle was selected from each speaker for the intelligibility and precision judgments, to minimize production variability that is more likely for the initial lists and to minimize fatigue effects likely during latter productions.

Listener Judgments of Word and Sentence Intelligibility

Each speaker's speech sample was judged by four different listeners, because intelligibility scores can vary widely between listeners (K. M. Yorkston & Beukelman, 1980). To determine word intelligibility, listeners were seated in a sound booth (IAC Acoustics, North Aurora, IL) and were instructed to orthographically transcribe each word, which they heard via loudspeakers at 50 dB HL. Listeners were told that the spoken words were all real English words and were instructed to produce orthographic transcriptions of real words. The order of words presented to the listeners was randomized.

Similarly, for sentence intelligibility, four listeners blinded to the study completed orthographic transcriptions

of the 11 SIT sentences recorded from each participant. The 5–15 word sentences for each listener were randomly generated by the test software to minimize rater familiarity of the test stimuli. Listeners were asked to listen to each sentence via noise-canceling headphones and write down what they heard as accurately as possible. For both the word and sentence intelligibility tasks, listeners were allowed to listen to each speech sample no more than two times.

Listener Judgments of Articulatory Precision

For the articulatory precision judgments, the two trained listeners listened to the stimuli from all 16 speakers via noise-canceling headphones and broadly transcribed each word using the International Phonetic Alphabet. Listeners were asked to indicate which phonemes were misarticulated and the type of articulation error (i.e., deletion, substitution, distortion, or addition) detected for each misarticulated sound. In addition to errors stemming from tongue, lip, and/or jaw dysfunction, articulation errors resulting from other speech subsystem impairments were also considered for the precision ratings. For example, articulatory imprecision resulting from velopharyngeal dysfunction (i.e., nasalized oral consonants) or from respiratory problems (i.e., weak pressure consonants) were also included in the precision ratings.

Data Analysis

Word and Sentence Intelligibility

Listeners' orthographic transcriptions of each word were scored by a trained research assistant as being either correct or incorrect, which is consistent with the binomial scoring procedure used routinely for sentence intelligibility estimation. Words were scored as incorrect only after the second author ruled out spelling errors by listening to the audio samples. Sentence intelligibility was calculated by each listener by counting the number of correctly identified words and dividing by the total number of words possible, multiplied by 100. Intelligibility scores from the four listeners were used to compute the average word and sentence intelligibility score for each speaker.

Interrater reliability was determined using Spearman's rank order correlation, and there was a strong agreement between raters for intelligibility (r_s = .86, p = .01). Average agreement between raters was calculated as the mean of the correlation values from each possible pair of raters.

Articulatory Precision

The trained listeners' judgments of phoneme substitutions, omissions, distortions, and/or additions were compared, and when the two listeners disagreed about consonant and/or vowel precision for a word, they listened to the word again to reach consensus. The consensus ratings were used to calculate an articulatory precision score for each word, where each word started with an accuracy score of 1 and all phonemes within a word received the same

accuracy score based on the total number of phonemes. For example, each phoneme in a six-phoneme word was given a score of 0.166 (i.e., 1/6). If all the sounds in a word were produced accurately, then the accuracy score for that word was 1. If sounds were misarticulated, then the articulation errors were weighted based on their contribution to speech intelligibility, which was determined using the extant dysarthria literature (Lee, Sung, Sim, Han, & Song, 2012; Rudzicz, Namasivayam, & Wolff, 2012; K. M. Yorkston & Beukelman, 1980). The impact of this weighting on accuracy scores were as follows: additions reduced the phoneme accuracy score by 25%, distortions by 50%, substitutions by 75%, and omissions by 100%. Omissions were weighted most heavily because these errors are not only significant predictors of intelligibility decline in dysarthria (Lee et al., 2012) but also represent changes in syllable and word structure (Preston, Ramsdell, Oller, Edwards, & Tobin, 2011). Substitutions were given the next highest weighting because of the significant negative impact these errors have on word intelligibility, which are shown to differentiate dysarthric and nondysarthric speech (Rudzicz et al., 2012). Although most dysarthric articulation errors are distortions (K. M. Yorkston & Beukelman, 1980), these errors were weighted less heavily than substitutions because they were predicted to have a lesser impact on word intelligibility. Lastly, addition errors received the lowest weighting because these errors were predicted to have the least impact on intelligibility. The weighted phonetic accuracy or articulatory precision score for each word was calculated as the sum of the accuracy scores of each individual phoneme comprising the word. Our weighting system was broadly based on the Weighted Speech Sound Accuracy measure developed by Preston et al. (2011) for differentially weighting developmental speech sound errors.

Interrater reliability was determined using Spearman's rank order correlation, and there was a strong agreement between raters for articulatory precision (r_s = .86, p = .01). The average agreement between raters was calculated as the mean of the correlation values from each possible pair of raters.

Statistical Analysis

Because of the relatively small number of participants within each group (n = 8), nonparametric statistics were completed. Specifically, Kruskal–Wallis H tests were used to determine between-groups differences for each complexity level. Post hoc comparisons were conducted using Mann–Whitney U tests. An alpha level of .05 was used for significance testing. Effect sizes were calculated using the formula $r = Z/\sqrt{N}$, and Cohen's effect size estimates were used to interpret the meaning of the r score (Cohen, 1988).

Results
Phonetic Complexity Effects on Word Intelligibility

Kruskal–Wallis H tests were used to compare word intelligibility scores obtained from the three groups, that is,

talkers with mild speech impairments (n = 4), those with moderate to severe impairments (n = 4), and healthy controls (n = 8), for each complexity category. Findings revealed a significant group effect for low complexity words, $H(2)$ = 8.845, p < .05, and high complexity words, $H(2)$ = 9.572, p < .01. Findings of the post hoc analysis are shown in Figure 1. For both complexity categories, the ALS groups that were significantly different from the healthy controls had mean intelligibility scores that were 2 SDs below the control group mean. A large effect size (r = .79) was observed between the control and moderate–severe group for the low complexity words. Similarly, for the high complexity words, large effects sizes were observed between the control and mild group (r = .79) as well as the moderate–severe group (r = .78).

Phonetic Complexity Effects on Articulatory Precision

Articulatory precision scores for each complexity category obtained from the three groups (i.e., talkers with mild speech impairments, n = 4; those with more severe impairments, n = 4; and healthy controls, n = 8) were compared using Kruskal–Wallis H tests. Findings revealed a significant group effect for low complexity words, $H(2)$ = 9.581, p < .01, and high complexity words, $H(2)$ = 10.185, p < .01. Findings of the post hoc analysis are shown in Figure 2. The more severely impaired ALS group had mean precision scores that were 2 SDs below the control group mean for both complexity categories. For the low complexity words, a large effect size (r = .82) was observed between the control and moderate–severe group. Similarly, for the high complexity words, a large effect size was observed between the control and the moderate–severe group (r = .88).

Figure 1. Word intelligibility (±SE) across phonetic complexity categories for healthy controls, mildly impaired, and more severely impaired talkers with ALS. Brackets are used to denote statistically significant differences between groups.
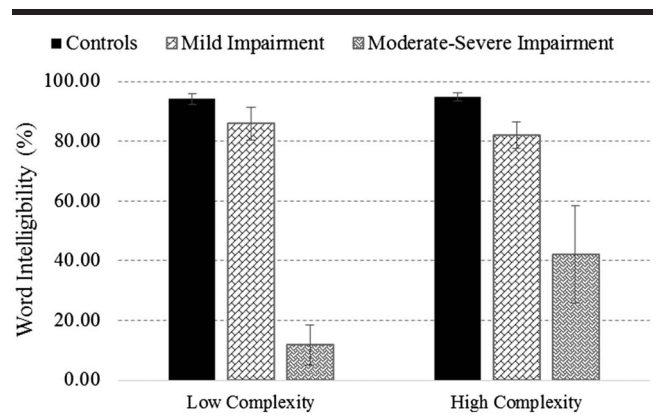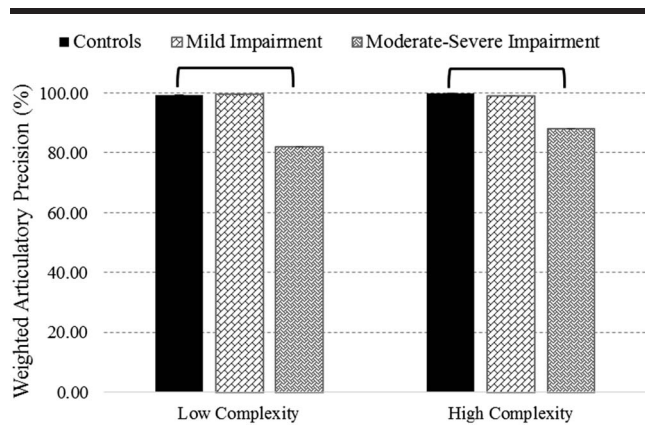
Figure 2. Weighted articulatory precision (±*SE*) across phonetic complexity categories for healthy controls, mildly impaired, and more severely impaired talkers with ALS. Brackets are used to denote statistically significant differences between groups.
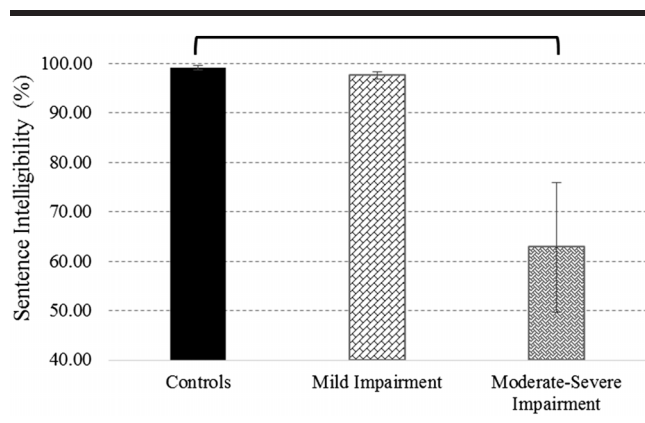

Sentence Intelligibility

SIT-based sentence intelligibility scores from the two ALS groups (i.e., mild vs. moderate–severe impairment) as well as healthy controls were compared using the Kruskal–Wallis *H* test. Findings revealed a significant effect of group on sentence intelligibility scores, $H(2) = 10.276$, $p < .01$. Findings of the post hoc analysis are shown in Figure 3.

Discussion

The purpose of this study was to determine the diagnostic value of a complexity-based approach for intelligibility and articulatory precision testing using preliminary data from talkers with ALS. Overall, the results support our hypothesis that word intelligibility is significantly different between less impaired talkers and healthy controls

Figure 3. Sentence intelligibility scores (±*SE*) for healthy controls and talkers with amyotrophic lateral sclerosis. Brackets denote significant between-groups differences.

only for more complex words. On the other hand, severely impaired talkers display significantly lower word intelligibility and precision scores compared to healthy controls even for words with low complexity. Furthermore, our results are in agreement with prior research that suggests that sentence intelligibility estimates are relatively insensitive to mild speech decline in talkers with ALS (Ball et al., 2001, 2002; Green et al., 2013), because we observed significant differences in sentence intelligibility only between controls and moderate to severely impaired talkers.

The observed severity-dependent effects of phonetic complexity on intelligibility were also reported by Allison and Hustad (2014), who demonstrated a significant negative effect of increased complexity on intelligibility for all groups of children included in their study, but with the greatest impact on children with dysarthria, for sentences above the first complexity level. However, the authors observed variable relationships between phonetic complexity and sentence intelligibility, which did not follow a consistent pattern across individual children with dysarthria. For example, in their study, the two children with the lowest and highest overall intelligibility scores did not demonstrate a significant effect of phonetic complexity on sentence intelligibility. Determining individual variability in phonetic complexity effects on intelligibility will be essential for the further refinement of speech assessments to improve identification and tracking of dysarthria progression.

Compared to intelligibility, the predicted severity-dependent effect of phonetic complexity on articulatory precision was not observed in those with mild speech impairments. Our results differ from findings reported by Kim and colleagues (2010), because these authors found a significant increase in error frequency with increased consonant complexity for both the low and high intelligibility groups of talkers with CP. However, speakers with low intelligibility displayed more manner of articulation errors and reduced the complexity of their utterances compared to high intelligibility speakers, suggesting that phonetic complexity has a greater impact on the speech of those with more severe impairments. Intact articulation observed in the mildly impaired group of the current study is also in contrast to the findings reported in other ALS studies, where articulatory dysfunction was observed before detectable declines in intelligibility (Green et al., 2013; Rong, Yunusova, Wang, & Green, 2015). Direct comparison between the current study and this previous ALS study is impossible because of methodological differences, that is, the use of auditory–perceptual versus kinematic analysis, respectively. That being said, one study focused on determining the phonetic basis of the intelligibility deficit in ALS reported that, even in talkers with comparatively high intelligibility, articulatory impairments stemming from jaw–tongue dysfunction contributed to the intelligibility deficit (J. F. Kent et al., 1992; R. D. Kent et al., 1990).

With further refinement, the weighted articulatory precision measure proposed in this study may show promise to determine the articulatory underpinnings of

the intelligibility deficit in progressive dysarthrias. A future direction of this research is to weight errors based on the specific phonetic features that are affected. For example, errors that involve tongue height dysfunction (high vs. low vowels) may need to be weighted higher than errors involving tongue advancement (front vs. back vowels), a feature that was found to be preserved even in the least intelligible talkers with ALS (R. D. Kent et al., 1990). Admittedly, as the weighting system becomes more complex, its diagnostic utility for clinical purposes will diminish, particularly in large multidisciplinary clinics where the time demands on clinicians are high. However, such an approach has value in identifying specific sound segments or contrasts that can be targeted in therapy and will allow for the planning of more effective therapy. Future work will also need to determine if words with phonetic complexity levels higher than those used in the current study will elicit significant articulatory precision differences both between and within groups, particularly for those showing early bulbar decline.

Interpreting intelligibility scores based on stimuli beyond the word level is difficult, because it is known to be impacted by several factors such as syntactic and semantic context, speaker familiarity, and predictability of the test stimuli (Hustad, Beukelman, & Yorkston, 1998; R. Kent, 1993; Weismer et al., 2001). On the other hand, evaluation of intelligibility based on single words might have less ecological validity, as single word utterances are not representative of an everyday communication context (Hustad et al., 1998; R. Kent, 1993; Weismer et al., 2001). Although some research suggests that word and sentence intelligibility are highly correlated (K. M. Yorkston & Beukelman, 1978), Weismer and colleagues argue that the high correlation between the two may be interpreted as a third variable artifact. That is, severity, which is the third variable, is highly correlated with each of these intelligibility scores, which, in turn, produces high correlations between the two intelligibility scores (Weismer et al., 2001). From a speech production perspective, it is assumed that talkers with ALS can effectively utilize compensatory inter- and intrasubsystem adjustments for connected speech compared to single word contexts, which may result in an inflation of sentence intelligibility, particularly during early stages of the disease (DePaul & Brooks, 1993; Green et al., 2013; Yunusova et al., 2010). Thus, for the purposes of early diagnosis and improved tracking of dysarthria progression, a more conservative, ecologically less valid estimate of intelligibility using single words may have an advantage over a less conservative estimate like sentence intelligibility.

Limitations and Future Directions

One limitation of the current study is the small sample size, which was further reduced because of stratification by dysarthria severity. Therefore, these results should be considered preliminary and interpreted with considerable caution. Second, syllable stress, which was not well controlled for in this study, may have influenced the word

intelligibility findings for the high complexity category where three of the eight words had an unstressed initial syllable and primary stress on the second syllable. Visual inspection of word intelligibility scores with those three words excluded revealed that intelligibility was higher for all three groups, that is, controls (+6%), mildly impaired (+3%), and moderate–severely impaired (+2%), but this difference became smaller as dysarthria severity increased. This finding supports the notion that, as syllabic contrasts are reduced with increasing dysarthria severity, listeners' ability to use syllabic strength to recognize word onsets is diminished (Liss, Spitzer, Caviness, Adler, & Edwards, 1998). Importantly, when those three words were excluded from the one-way analysis and post hoc comparisons, between-groups differences in word intelligibility similar to those depicted in Figure 1 were obtained. In terms of word length, the complexity categories were also different (i.e., two to three syllables for the low complexity category and three to four syllables for the high complexity category). However, stimulus length and phonetic complexity are inextricably linked, and just like phonetic complexity, increased word length also increases the demands on the speech motor system, which can lead to reduced intelligibility. One goal of our future work involving a larger sample size is to resolve the potential confounding effect of word length with phonetic complexity using statistical analysis. For example, these two variables could be entered in a regression analysis as potential predictors of reduced intelligibility, which will allow us to determine the unique contribution of each variable to the intelligibility deficit. Lastly, ceiling effects were observed for both healthy controls and mildly impaired speakers. For these two groups, between-groups differences may have emerged if the listening task was made more difficult by using multitalker babble, for example.

Clinical Implications and Conclusions

Although preliminary, our findings suggest that a phonetic complexity-based approach holds promise to capture disease-related declines in intelligibility, particularly in talkers with ALS who have mild speech impairments. In contrast, sentence intelligibility showed limited sensitivity to early speech impairments. Similarly, we were unable to detect differences between the controls and mildly impaired group using our complexity-based approach for estimating articulatory precision. With further refinements to our precision measure and by using more challenging stimuli in future studies, we may be able to detect subtle differences in articulatory performance in early ALS. Overall, our study shows that considering phonetic complexity while selecting speech test stimuli may help with the early detection and accurate monitoring of dysarthria progression in the ALS population. In addition, our approach for estimating articulatory precision by differentially weighting errors is novel and may be used to supplement more traditional methods of tracking dysarthria progression.

Acknowledgments

References

Allison, K. M., & Hustad, K. C. (2014). Impact of sentence length and phonetic complexity on intelligibility of 5-year-old children with cerebral palsy. *International Journal of Speech-Language Pathology, 16*(4), 396–407.

Ball, L. J., Beukelman, D. R., & Pattee, G. L. (2002). Timing of speech deterioration in people with amyotrophic lateral sclerosis. *Journal of Medical Speech-Language Pathology, 10*(4), 231–235.

Ball, L. J., Willis, A., Beukelman, D. R., & Pattee, G. L. (2001). A protocol for identification of early bulbar signs in amyotrophic lateral sclerosis. *Journal of the Neurological Sciences, 191*(1), 43–53.

Bent, T., Bradlow, A. R., & Smith, B. L. (2007). Segmental errors in different word positions and their effects on intelligibility of non-native speech. *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege,* 331–347.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*(4), 977–990.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

DePaul, R., & Brooks, B. R. (1993). Multiple orofacial indices in amyotrophic lateral sclerosis. *Journal of Speech and Hearing Research, 36*(6), 1158–1167.

Duffy, J. R. (2013). *Motor speech disorders: Substrates, differential diagnosis, and management.* St. Louis, MO: Elsevier Mosby.

Dworzynski, K., & Howell, P. (2004). Predicting stuttering from phonetic complexity in German. *Journal of Fluency Disorders, 29*(2), 149–173.

Forrest, K., Weismer, G., & Turner, G. S. (1989). Kinematic, acoustic, and perceptual analyses of connected speech produced by Parkinsonian and normal geriatric adults. *The Journal of the Acoustical Society of America, 85*(6), 2608–2622.

Gordon, B. (1985). Subjective frequency and the lexical decision latency function: Implications for mechanisms of lexical access. *Journal of Memory and Language, 24*(6), 631–645.

Green, J. R., Yunusova, Y., Kuruvilla, M. S., Wang, J., Pattee, G. L., Synhorst, 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.

Hustad, K. C., Beukelman, D. R., & Yorkston, K. M. (1998). Functional outcome assessment in dysarthria. *Seminars in Speech and Language, 19*(3), 291–302. Stuttgart, Germany: Thieme.

Kent, J. F., Kent, R. D., Rosenbek, J. C., Weismer, G., Martin, R., Sufit, R., & Brooks, B. R. (1992). Quantitative description of the dysarthria in women with amyotrophic lateral sclerosis. *Journal of Speech and Hearing Research, 35*(4), 723–733.

Kent, R. (1993). Speech intelligibility and communicative competence in children. In A. P. Kaiser & A. D. Gray (Eds.), *Enhancing children's communication* (pp. 233–239). Baltimore, MD: Brookes.

Kent, R. D. (1992). The biology of phonological development. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 65–90). Timonium, MD: York Press.

Kent, R. D., Kent, J. F., Weismer, G., Sufit, R. L., Rosenbek, J. C., Martin, R. E., & Brooks, B. R. (1990). Impairment of speech intelligibility in men with amyotrophic lateral sclerosis. *Journal of Speech and Hearing Disorders, 55*(4), 721–728.

Kim, H., Martin, K., Hasegawa-Johnson, M., & Perlman, A. (2010). Frequency of consonant articulation errors in dysarthric speech. *Clinical Linguistics & Phonetics, 24*(10), 759–770.

Lee, Y. M., Sung, J. E., Sim, H. S., Han, J. H., & Song, H. N. (2012). Analysis of articulation error patterns depending on the level of speech intelligibility in adults with dysarthria. *Communication Sciences & Disorders, 17*(1), 130–142.

Liss, J. M., Spitzer, S., Caviness, J. N., Adler, C., & Edwards, B. (1998). Syllabic strength and lexical boundary decisions in the perception of hypokinetic dysarthric speech. *The Journal of the Acoustical Society of America, 104*(4), 2457–2466.

MacNeilage, P. F., & Davis, B. (1990). Acquisition of speech production: Frames, then content. In M. Jeannerod (Ed.), *Attention and performance 13: Motor representation and control* (pp. 453–476). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., . . . Chertkow, H. (2005). The Montréal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society, 53*(4), 695–699.

Nusbaum, H. C., Pisoni, D. B., & Davis, C. K. (1984). Sizing up the Hoosier mental lexicon. *Research on Spoken Language Processing Report, No. 10,* 357–376.

Platt, L. J., Andrews, G., & Howie, P. M. (1980). Dysarthria of adult cerebral palsy: II. Phonemic analysis of articulation errors. *Journal of Speech and Hearing Research, 23*(1), 41–55.

Platt, L. J., Andrews, G., Young, M., & Quinn, P. T. (1980). Dysarthria of adult cerebral palsy: I. Intelligibility and articulatory impairment. *Journal of Speech and Hearing Research, 23*(1), 28–40.

Preston, J. L., Ramsdell, H. L., Oller, D. K., Edwards, M. L., & Tobin, S. J. (2011). Developing a weighted measure of speech sound accuracy. *Journal of Speech, Language, and Hearing Research, 54*(1), 1–18.

Rong, P., Yunusova, Y., Wang, J., & Green, J. R. (2015). Predicting early bulbar decline in amyotrophic lateral sclerosis: A speech subsystem approach. *Behavioural Neurology, 2015,* 1–11. https://doi.org/10.1155/2015/183027

Rosen, K. M., Goozée, J. V., & Murdoch, B. E. (2008). Examining the effects of multiple sclerosis on speech production: Does phonetic structure matter? *Journal of Communication Disorders, 41*(1), 49–69.

Rudzicz, F., Namasivayam, A. K., & Wolff, T. (2012). The TORGO database of acoustic and articulatory speech from speakers with dysarthria. *Language Resources and Evaluation, 46*(4), 523–541.

Sheard, C., Adams, R. D., & Davis, P. J. (1991). Reliability and agreement of ratings of ataxic dysarthric speech samples with varying intelligibility. *Journal of Speech and Hearing Research, 34*(2), 285–293.

Vitevitch, M. S., & Luce, P. A. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, & Computers, 36*(3), 481–487.

Weismer, G. (2008). Speech intelligibility. In M. J. Ball, M. Perkins, N. Müller, & S. Howard (Eds.), *The handbook of clinical linguistics* (pp. 568–582). Oxford, UK: Blackwell.

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., Laures, J. S., Jeng, J. Y., Kent, R. D., & Kent, J. F. (2000). Effect of speaking rate manipulations on acoustic and perceptual aspects of the dysarthria in amyotrophic lateral sclerosis. *Folia Phoniatrica et Logopaedica, 52*(5), 201–219.

Whitehill, T. L. (2002). Assessing intelligibility in speakers with cleft palate: A critical review of the literature. *The Cleft Palate-Craniofacial Journal, 39*(1), 50–58.

Wilson, M. (1988). MRC psycholinguistic database: Machine-usable dictionary, Version 2.00. *Behavior Research Methods, Instruments, & Computers, 20*(1), 6–10.

Yorkston, K., Beukelman, D., Hakel, M., & Dorsey, M. (2007). *Speech intelligibility test* [Computer software]. Lincoln, NE: Institute for Rehabilitation Science and Engineering at the Madonna Rehabilitation Hospital.

Yorkston, K. M., & Beukelman, D. R. (1978). A comparison of techniques for measuring intelligibility of dysarthric speech. *Journal of Communication Disorders, 11*(6), 499–512.

Yorkston, K. M., & Beukelman, D. R. (1980). A clinician-judged technique for quantifying dysarthric speech based on single-word intelligibility. *Journal of Communication Disorders, 13*(1), 15–31.

Yunusova, Y., Green, J. R., Lindstrom, M. J., Ball, L. J., Pattee, G. L., & Zinman, L. (2010). Kinematics of disease progression in bulbar ALS. *Journal of Communication Disorders, 43*(1), 6–20.

Zeplin, J., & Kent, R. D. (1996). Reliability of auditory-perceptual scaling of dysarthria. In D. A. Robin, K. M. Yorkston, & D. R. Beulkman (Eds.), *Disorders of motor speech: Recent advances in assessment, treatment, and clinical characterization* (pp. 145–154). Baltimore, MD: Paul H. Brookes.