INTERSPEECH 2019
September 15–19, 2019, Graz, Austria

Profiling Speech Motor Impairments in Persons with Amyotrophic Lateral Sclerosis: An Acoustic-Based Approach

Hannah P. Rowe, Jordan R. Green

Massachusetts General Hospital Institute of Health Professions
hrowe@mghihp.edu, jgreen2@mghihp.edu

Abstract

The goal of this study was to profile the speech motor impairments that characterize dysarthria secondary to amyotrophic lateral sclerosis (ALS). This information is important for identifying optimal treatment strategies and guiding speech impairment subtype discovery, which may facilitate ongoing efforts to improve automatic speech recognition (ASR) of dysarthric speech. Speech motor impairments were profiled by introducing a novel framework that assesses four key components of motor control: *coordination*, *consistency*, *speed*, and *precision*. An individual acoustic feature was selected to represent each component. Specifically, *coordination* was indexed by the proportion of voice onset time (VOT) to syllable duration, *consistency* by the coefficient of variation of VOT between repetitions of /pataka/ within each distinct consonant, *speed* by the slope of the second formant (F2), and *precision* by the standard deviation of F2 slope between distinct consonants within each repetition of /pataka/. Acoustic measures were extracted from audio recordings of each participant (18 controls and 14 participants with ALS) during a sequential motion rate (SMR) task. Results revealed that the primary underlying speech motor impairments that characterize ALS are in *coordination*, *speed*, and *precision*. Further research is needed to validate the existence of speech-impairment-based subtypes across the continuum of speech motor disorders.

Index Terms: dysarthric speech, objective assessment, interpretable features, automatic speech recognition

1. Introduction

Dysarthria secondary to amyotrophic lateral sclerosis (ALS) has a profound impact on an individual's ability to communicate [1]. Despite the devastating effect of dysarthria on quality of life, only a few validated tools are available for assessing the motor impairments that give rise to abnormal speech patterns [2]. This paucity of impairment-specific information limits the accuracy of speech diagnostics and, arguably, the development of usable automatic speech recognition (ASR) systems for dysarthric speech [3]. The absence of robust ASR for dysarthric speech [4, 5, 6] is particularly unfortunate given that this technology has the potential to improve the quality of life for millions of people worldwide who suffer from speech impairments. Although speaker-dependent models may resolve some of these issues, this approach precludes use of commercial ASR systems and is unfeasible for many talkers with neurologic disorders due to physical fatigue involved in producing adequate training samples. One potential, less explored strategy for reducing the need for large training samples is to generate a limited but representative number of recognition models that accommodate groups of speakers with common speech errors (i.e., speech-impairment-based models).

Before these impairment-based models can be tested, however, conceptual frameworks are needed to guide speech impairment subtype discovery. Such frameworks can be data-driven, symptom-driven, or performance-driven. The proliferation of machine-based classifiers has given rise to widespread testing of data-driven approaches to detect and characterize abnormal speech features [7, 8]. Data-driven approaches have motivated the development of automatic acoustic feature extraction programs, such as OpenSMILE. Although these programs extract thousands of features, many of the features are uninterpretable and therefore of limited use when trying to understand the mechanisms of speech motor disorders. The importance of interpretable features has consequently been emphasized when assessing disordered speech [9].

Symptom-driven subtyping is a common component of the clinical speech exam and is based on patient or clinician perceptions [10, 11, 12, 13]. Clinician perceptions provide the basis for the most widely used speech-impairment classification system – the Darley, Aronson, and Brown (DAB) model – which is used for differential diagnosis to stratify dysarthrias into six different subtypes (i.e., spastic, flaccid, hyperkinetic, hypokinetic, mixed, ataxic) [10, 11]. However, stratifications made using this system are limited by poor reliability and disagreement among listeners [12, 14].

To overcome these limitations, researchers have been investigating the efficacy of performance-based measures of speech motor control [2]. This approach involves the use of objective speech measures derived from speech acoustics or sensors that record directly from the different speech subsystems. Among all of the performance-based measures, acoustic measures are by far the most frequently used because recording speech samples is noninvasive, efficient, and a procedure that is already embedded in most clinical environments. Within the speech acoustics literature, features from multiple speech subsystems (i.e., articulation, phonation, resonance, and respiration) have been tested for their ability to predict declines in speech motor function [15]. In comparison to measures from other subsystems, articulatory features were found to account for the majority of the loss of intelligibility in dysarthric speakers [15].

The current project represents the first step toward a speech-impairment-based subtyping paradigm: developing an acoustic feature set that, in the future, can be used to group persons with speech motor disorders based on their motor impairment profiles. Our proposed framework consists of four key components that have been shown to comprehensively characterize speech and limb motor control: *coordination* [16, 17, 18], *consistency* [16, 19, 20], *speed* [16, 21], and *precision* [16, 21]. While these components have been evaluated

http://dx.doi.org/10.21437/Interspeech.2019-2911

separately, no study to our knowledge has examined all four areas within a single framework for speech motor control. We used both novel and existing acoustic measures to represent each of the four components in our framework (see Figure 1).

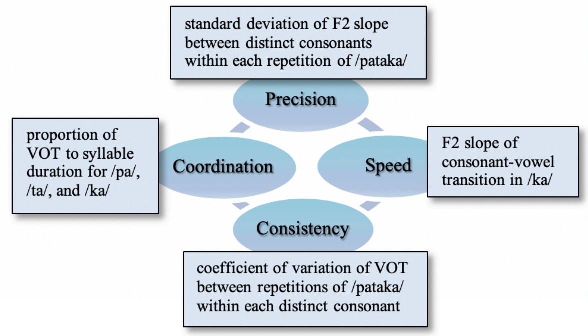


Figure 1: *Four key components of speech motor control and the corresponding acoustic features.*

Coordination in speech motor disorders has been regularly studied using voice onset time (VOT). Due to its reliability as a temporal parameter, VOT has served as an indirect measure of vocal tract motor control [22]. As such, abnormalities in VOT are interpreted as a loss of timing control [23]. While VOT has been examined in patients with dysarthria secondary to various conditions and neurodegenerative diseases [23], very few studies have examined VOT in patients with ALS. However, one study reported no significant difference in VOT between patients with ALS and healthy controls [24].

Consistency as it relates to speech motor disorders refers to the stability of speech sounds over time or across multiple repetitions of speech sequences [25]. Many researchers have analyzed consistency by investigating the variability in VOT across repetitions of the same utterance [23], but this acoustic measure has not been examined in patients with ALS.

An increasing amount of literature has determined the slope of the second formant (F2 slope) to be a viable acoustic correlate of articulatory movement *speed* [26, 27]. F2 slope offers a precise measure of articulator speed, as it represents the rate of change in the vocal tract configuration during speech sounds [27, 28]. In patients with ALS, several studies have found significant reductions in F2 slope compared to healthy controls [29, 30].

Lastly, few studies have investigated acoustic features associated with *precision*. However, studies examining healthy speakers found that F2 slope could be used as an acoustic cue for place of articulation in consonant-vowel pairs [31, 32]. Theoretically, large differences between F2 slopes of different consonants may reflect phoneme distinctiveness. Our study will thus use a novel acoustic correlate of precision: the standard deviation of F2 slope across distinct consonant-vowel pairs (i.e., /pa/, /ta/, and /ka/).

To assess the performance of patients with dysarthria secondary to ALS on these four components of speech motor control, we asked the following research questions:

1. Are there differences in performance on the four components between healthy controls and speakers with ALS?
2. How does performance on the four components change as a function of speech severity?

2. Method

2.1. Participants

Acoustic data from 7 individuals diagnosed with ALS and 18 controls were obtained from the x-ray microbeam (XRMB) dysarthria database collected at the University of Wisconsin—Madison [26]. Data from 7 additional individuals with ALS were obtained from an ongoing prospective study on speech deterioration due to ALS at the MGH Institute of Health Professions. Given that articulation rate typically declines prior to speech intelligibility and is considered to be the primary sign of onset of bulbar ALS [33], we calculated syllable rate from a sequential motion rate (SMR) task as a measure of speech severity level. In this task, participants are instructed to produce the syllable sequence /pataka/ as many times as possible on one breath. The ALS participants were representative of the severity spectrum, ranging from 1.72 syll/sec to 8.27 syll/sec. Participant demographics and mean SMRs for each group are reported in Table 1. Control participants had a negative history of speech, language, and hearing problems. All participants were native speakers of English.

Table 1: *Demographic and speech characteristics of the ALS group and the control group.*

	N	Age	Sex	SMR
ALS	14	52.6 (10.7)	9 M 5 F	4.47 (2.23)
Controls	18	59.3 (11.5)	9 M 9 F	6.75 (1.15)

2.2. Procedures

Participants were instructed to produce /pataka/ as many times as they could on one breath, as quickly and accurately as possible. A head-mounted, professional quality microphone was positioned at approximately 5 cm from the mouth during the recordings. Samples were recorded at a sampling rate of 22 kHz. Acoustic analyses of the first three repetitions of /pataka/ for each participant were conducted offline using Praat [27]. Formant settings were adjusted based on the gender of the participants. The maximum formant frequency was set to 5500 Hz for females and 5000 Hz for males. Analyses were conducted on the first three repetitions of /pataka/ for each participant.

2.3. Measurements

Figure 2 displays the waveforms, spectrograms, and textgrids of /pa/, /ta/, and /ka/ produced by a speaker in the ALS group. All acoustic measurements were completed in Praat [27].

2.3.1. Coordination

The proportion of VOT to syllable duration was used as our coordination measure. VOT was obtained manually based on a wideband spectrogram. First, the time interval between the acoustic energy of the stop consonant and the periodic wave energy of the subsequent vowel was determined. Vertical cursors were then placed at these two time points and the time between cursors was calculated as the VOT. The proportion of VOT to syllable duration was calculated for each unique consonant and then the mean proportion was calculated for each participant. A second trained researcher re-measured the VOTs

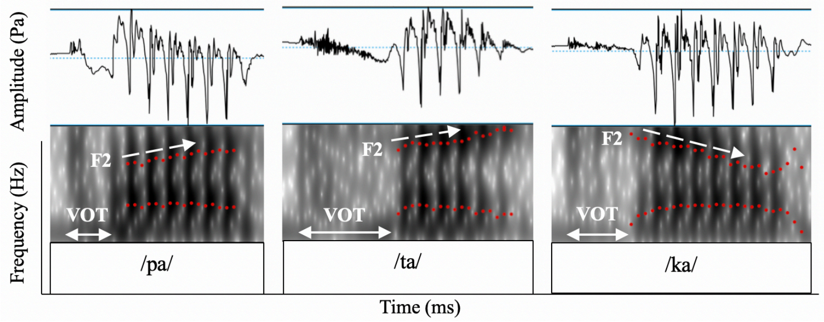


Figure 2: *Waveforms, spectrograms, and textgrids showing the segmentation of VOT (solid arrows) and F2 slope (dotted arrows) of /pa/, /ta/, and ka/ produced by a speaker in the ALS group.*

of 10% of the samples to ensure inter-rater reliability. The inter-rater reliability coefficient was at r = .95, with an average error of .003 milliseconds.

2.3.2. *Consistency*

The coefficient of variation of VOT between repetitions was used as our measure of consistency. First, we calculated the standard deviation of the VOT across each repetition of each distinct consonant-vowel pair (i.e., /pa/, /ta/, /ka/) for each participant. Thus, the standard deviation was computed on three repetitions of each syllable. To interpret the relative magnitude of the standard deviation, the coefficient of variation was calculated by dividing each standard deviation by the mean VOT for that participant and then multiplying the quotient by 100. The mean coefficient of variation was then calculated for each participant.

2.3.3. *Speed*

The F2 slope (i.e., F2 range/time) of the consonant-vowel transition in /ka/ was used as our measure of speed because it was the greatest of the three consonants (i.e., /p/, /t/, /k/). The typical F2 trajectory for /k/ is characterized by a decreasing transition between the consonant and the subsequent vowel.
This transition was identified by hand on a wideband spectrogram. First, the F2 of /k/ was found using automatic formant tracking in Praat [27]. Then, the entire vowel was segmented from the first glottal pulse to the last glottal pulse. The onset frequency of F2 slope was defined using the first glottal pulse, while the offset frequency was calculated using the midpoint of the vowel to control for coarticulatory affects from the subsequent consonant. The mean F2 slope of all repetitions was then calculated for each participant. A second trained researcher re-measured the F2 slopes of 10% of the samples to assess inter-rater reliability. The inter-rater reliability coefficient was at r = .92, with an average error of 34.60 Hz/ms.

2.3.4. *Precision*

The standard deviation of F2 slope between distinct consonants was used as our measure of precision. First, we calculated the standard deviation of the F2 slope across each distinct consonant-vowel pair of each repetition for each participant. The mean standard deviation was then calculated for each participant.

2.4. Statistical Analyses

Group comparisons were conducted using a linear mixed model (LMM). The analysis compared the ALS group and controls on the four acoustic measures with the diagnosis group as our fixed effect. To account for the interdependencies of multiple measures per participant, we treated individuals as our random effect. Correlation analyses were completed between SMR and each acoustic measure in the ALS group to examine how performance on the four components changed with speech severity. Correlation analyses were subsequently completed between all acoustic measures to ensure that the features represented independent constructs.

3. Results

3.1. Comparison Analyses

We found significant differences between the ALS group and controls in *coordination*, *speed*, and *precision* but not in *consistency* (see Table 2). Effect sizes and boxplots of the comparisons are displayed in Figure 3 and Figure 4, respectively.

Table 2: *Means and standard deviations for each group as well as p values and Cohen's ds for each comparison.*

Group	Mean (SD)	Comparison	*p*	Cohen's *d*
	Precision (Hz/ms)			
Con	8518.73 (3152.94)	Con-ALS	<.01*	1.08 [.30, 1.86]
ALS	4882.06 (3609.26)			
	Speed (Hz/ms)			
Con	-7636.10 (3991.20)	Con-ALS	=.04*	.79 [.03, 1.54]
ALS	-4341.85 (4433.82)			
	Consistency (%)			
Con	24.89 (6.51)	Con-ALS	=.11	.59 [-.15, 1.33]
ALS	29.48 (9.15)			
	Coordination			
Con	.53 (.23)	Con-ALS	<.01*	1.25 [.45, 2.04]
ALS	.32 (.15)			

Con = controls

3.2. Correlation Analyses

Pearson correlations were computed between SMR and each acoustic measure in the ALS group. Correlation coefficients and p values are displayed in Figure 4. Pearson correlations were also computed between all acoustic measures. *Speed* and *precision* were strongly correlated, $r(30) = .88$, $p < .01$; however, correlations were weak between all other variables: *coordination* and *consistency* ($r(30) = .36$, $p = .04$), *coordination* and *speed* ($r(30) = .29$, $p = .11$), *coordination* and *precision* ($r(30) = .31$, $p = .09$), *consistency* and *speed* ($r(30) = .20$, $p = .27$), and *consistency* and *precision* ($r(30) = .21$, $p = .26$).

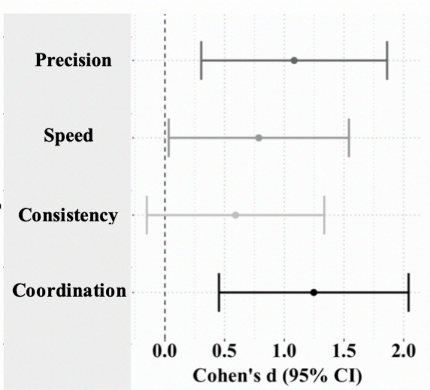
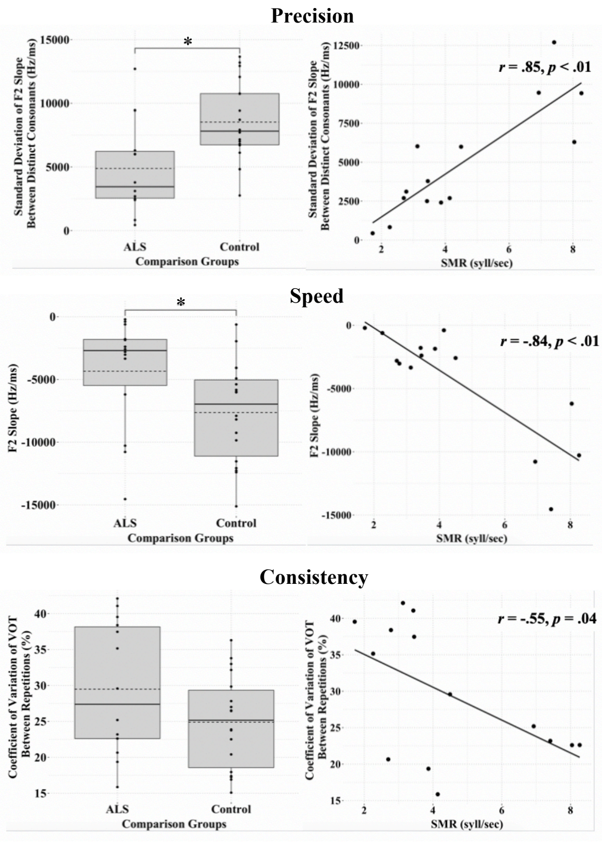


Figure 3: *Forest plot of effect sizes for the ALS group compared to the control group on each of the four components.*



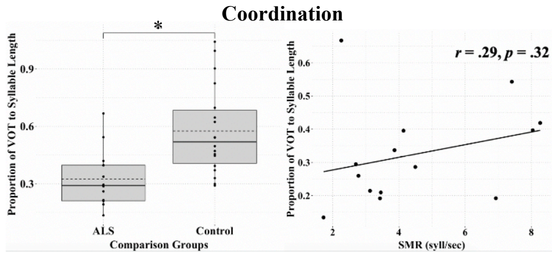


Figure 4: *Left panel: Boxplots of comparisons (dotted line = mean, solid line = median). Right panel: Correlations between SMR and each measure for the ALS group.*

4. Discussion

In this paper, we proposed a novel framework composed of four key components of speech motor control (i.e., *coordination*, *consistency*, *speed*, and *precision*) to characterize speech impairments in persons with ALS. Our findings revealed impairments across all components except *consistency*. We also found that *consistency*, *speed*, and *precision* were correlated with speech severity in the ALS group as indexed by SMR.

Dysarthria secondary to ALS is characterized by progressive muscle paresis, resulting in weakness of the speech articulators [1]. Our findings are consistent with previous studies on patients with ALS, which collectively show that this loss of strength has a broad impact on multiple articulatory components [28, 29, 34, 35, 36, 37]. Our findings are also in agreement with speech kinematic literature that has found consistency to be preserved in patients with mild dysarthria secondary to ALS [36]. Although our results demonstrate a more global profile of impairment, the four components appear to be distinct constructs as indicated by the lack of association in our correlation analyses, except between speed and precision, which should be further investigated.

Collectively, these data suggest that the paresis associated with ALS may be characterized by underlying impairments in *coordination*, *speed*, and *precision*. Coordination deficits were evidenced by our VOT findings, which reflect dyscoordination between supra and subglottal gestures [35]. The decreased capacity to generate speed was evidenced by reductions in F2 slope, which has been shown to be a common manifestation of ALS [28, 29]. Lastly, deficits in articulatory precision may account for prior reports of reduced phoneme distinctiveness and its detrimental impact on speech intelligibility [37].

Our correlation analyses indicate that *consistency* may be preserved in the early stages of ALS but later affected as the disease progresses. Thus, the lack of a significant between-group difference may be due to the variability in the ALS group on this measure.

Additional research is required to determine whether these four components of speech motor control correlate with functional measures of speech such as intelligibility. Further research is also needed to test this framework with other types of dysarthria that are likely to be characterized by different patterns of impairments. The results of such studies will inform impairment-specific ASR models and determine the use of this framework as an effective assessment tool.

5. Acknowledgements

This work was supported by NIH-NIDCD under Grants R01DC009890, R01DC0135470, and K24DC016312.

6. References

[1] Duffy, J.R. (2013). Motor speech disorders: Substrates, differential diagnosis, and management (3rd edition). St. Louis, MI: Elsevier.

[2] 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.

[3] Green, P., Carmichael, J., Enderby, P., & Hawley, M.S. (2003). Automatic speech recognition with sparse training data for dysarthric speakers. In Proceedings of *8TH European Conference on Speech Communication and Technology, EUROSPEECH 2003-INTERSPEECH 2003*. Switzerland, 2003.

[4] Blaney, B. & Wilson, J. (2000). Acoustic variability in dysarthria and computer speech recognition. *Clinical Linguistics and Phonetics*, *14*(4), 307-327.

[5] Deller, J., Hsu, D., & Ferrier L. (1991). On the use of hidden Markov modelling for recognition of dysarthric speech. *Computer Methods and Programmes in Biomedicine*, *35*, 125-139.

[6] Doyle, P., Leeper, H., Kotler, A-L, Thomas-Stonell, N., O'Neill, C., Dylke, M-C., Rolls, K. (1997). Dysarthric speech: A comparison of computerised speech recognition and listener intelligibility. *Journal of Rehabilitation Research and Development*, *34*(3), 309-316.

[7] Bayestehtashk, A., Asgari, M., Shafran, I., & McNames, J. (2015). Fully automated assessment of the severity of Parkinson's disease from speech. *Computer Speech and Language*, *29*(1), 172-185.

[8] Wang, Y., Kent, R.D., Duffy, J.R., & Thomas, J.E. (2009). Analysis of diadochokinesis in ataxic dysarthria using the Motor Speech Profile Program. *Folia Phoniatrica Et Logopaedica: Official Organ of the International Association of Logopedics and Phoniatrics (IALP)*, *61*, 1-11.

[9] Tu, M., Berisha, V., & Liss, J. (2017). Interpretable objective assessment of dysarthric speech based on deep neural networks. In Proceedings of *Interspeech 2017.* Sweden, August 2017.

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

[11] Darley, F.L., Aronson, A.E., & Brown, J.R. (1969). Differential diagnostic patterns of dysarthria. *Journal of Speech, Language, and Hearing Research*, *12*, 246–269.

[12] Zyski, J.B. & Weisiger, B.E. (1987). Identification of dysarthria types based on perceptual analysis. *Journal of Communication Disorders*, *20*, 367–378.

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

[14] 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.

[15] 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.

[16] 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.

[17] Green, J.R., Moore, C.A., Higashikawa, M., & Steeve, R.W. (2000). The physiologic development of speech motor control. *Journal of Speech, Language, and Hearing Research*, *43*(1), 239-255.

[18] Gracco, V.L. & Lofqvist, A. (1994). Speech motor coordination and control: Evidence from lip, jaw, and laryngeal movements. *Journal of Neuroscience*, *14*(11), 6585-6597.

[19] Sharkey, S.G. & Folkins, J.W. (1985). Variability of lip and jaw movements in children and adults. *Journal of Speech, Language, and Hearing Research*, *28*(1), 8-15.

[20] Abs, J.H. & Gracco, V.L. (1984). Control of complex motor gestures: Orofacial muscle responses to load perturbations of lip during speech. *Journal of Neurophysiology*, *51*(4), 705-723.

[21] Gracco, V.L. (1990). Chapter 1 characteristics of speech as a motor control system. *Advances in Psychology*, *70*, 3-28.

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

[23] Auzou, P., Ozsancak, O., Morris, R.J., Jan, M., Eustache, F. & Hannequin, D. (2000). Voice onset time in aphasia, apraxia of speech, and dysarthria: A review. *Clinical Linguistics & Phonetics*, *14*(2), 131-150.

[24] Caruso, A.J. & Burton, E.K. (1987). Temporal acoustic measures of dysarthria associated with amyotrophic lateral sclerosis. *Journal of Speech and Hearing Research*, *30*, 80-87.

[25] McNeil, M.R., Odell, K.H., Miller, S.B., & Hunter, L. (1995). Consistency, variability and target approximation for successive speech repetitions among apraxic, conduction aphasic and ataxic dysarthric speakers. *Clinical Aphasiology*, *23*, 39–55.

[26] Westbury, J.R. (1994). *X-ray microbeam speech production database user's handbook.* Madison, WI: Waisman Center on Mental Retardation & Human Development.

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

[28] 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: Official Organ of the International Association of Logopedics and Phoniatrics (IALP)*, *64*, 94-102.

[29] Kim, Y., Weismer, G., Kent, R.D., Duffy, & Green, J.R. (2009). Statistical models of F2 slope in relation to severity of dysarthria. *Folia Phoniatrica Et Logopaedica: Official Organ of the International Association of Logopedics and Phoniatrics (IALP)*, *61*, 329–335.

[30] Kent, R.D., Kent, J.F., Weismer, G., Martin, R.E., 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.

[31] Chen, W.S. & Alwan, A. (2000). Place of articulation cues for voiced and voiceless plosives and fricatives in syllable-initial position. In Proceedings of *Interspeech 2000.* China, October 2000.

[32] Delattre, P.C., Liberman, A.M., & Cooper, F.S. (1955). Acoustic loci and transitional cues for consonants. *JASA*, *27*, 769-773.

[33] Yorkston, K.M., Strand, E., Miller, R., Hillel, A., & Smith, K. (1993). Speech deterioration in amyotrophic lateral sclerosis: implications for the timing of intervention. *Journal of Medical Speech- Language Pathology*, *1*, 35–46.

[34] Samlan, R.A. & Weismer, G. (1995). The relationship of selected perceptual measures of diadochokinesis to speech intelligibility in dysarthric speakers with amyotrophic lateral sclerosis. *American Journal of Speech-Language Pathology*, *4*(2), 9-13.

[35] Tyler, A.A. & Watterson, T.L. (1991). Vot as an indirect measure of laryngeal function. *Seminars in Speech and Language*, *12*(2), 131-141.

[36] Mefferd, A.S., Pattee, G.L., & Green, J.R. (2014). Speaking rate effects on articulatory pattern consistency in talks with mild ALS. *Clinical Linguistics & Phonetics*, *28*(11), 799-811.

[37] Weismer, G. & Green, J.R. (2015). Speech production in motor speech disorders: Lesions, models, and a research agenda. In M.A. Redord (Ed.), The handbook of speech production (pp. 298-330). Hoboken, N.J.: John Wiley & Sons, Inc.