Folia Phoniatrica et Logopaedica

Research Article

Folia Phoniatr Logop 2023;75:23–34
DOI: 10.1159/000525514

Received: September 8, 2021
Accepted: June 2, 2022
Published online: June 27, 2022

Tongue and Lip Acceleration as a Measure of Speech Decline in Amyotrophic Lateral Sclerosis

Kristin J. Teplansky[a] Alan Wisler[b] Jordan R. Green[c, d] Thomas Campbell[e] Daragh Heitzman[f] Sara G. Austin[g] Jun Wang[a, g]

[a]Speech, Language, and Hearing Sciences, University of Texas, Austin, TX, USA; [b]Mathematics and Statistics, Utah State University, Logan, UT, USA; [c]Communication Sciences and Disorders, MGH Institute of Health Professions, Boston, MA, USA; [d]Speech and Hearing Bioscience and Technology Program, Harvard University, Boston, MA, USA; [e]Callier Center for Communication Disorders, University of Texas at Dallas, Dallas, TX, USA; [f]MDA/ALS Clinic, Texas Neurology, Dallas, TX, USA; [g]Neurology, Dell Medical School, University of Texas, Austin, TX, USA

Keywords

Amyotrophic lateral sclerosis · Dysarthria · Speech kinematics · Speech motor control · Support vector machine

Abstract

Purpose: The goal of this study was to examine the efficacy of acceleration-based articulatory measures in characterizing the decline in speech motor control due to amyotrophic lateral sclerosis (ALS). ***Method:*** Electromagnetic articulography was used to record tongue and lip movements during the production of 20 phrases. Data were collected from 50 individuals diagnosed with ALS. Articulatory kinematic variability was measured using the spatiotemporal index of both instantaneous acceleration and speed signals. Linear regression models were used to analyze the relationship between variability measures and intelligible speaking rate (a clinical measure of disease progression). A machine learning algorithm (support vector regression, SVR) was used to assess whether acceleration or speed features (e.g., mean, median, maximum) showed better performance at predicting speech severity in patients with ALS. ***Results:*** As intelligible speaking rate declined, the variability of acceleration of tongue and lip movement patterns significantly increased ($p < 0.001$). The variability of speed and vertical displacement did not significantly predict speech performance measures. Additionally, based on R^2 and root mean square error (RMSE) values, the SVR model was able to predict speech severity more accurately from acceleration features ($R^2 = 0.601$, RMSE = 38.453) and displacement features ($R^2 = 0.218$, RMSE = 52.700) than from speed features ($R^2 = 0.554$, RMSE = 40.772). ***Conclusion:*** Results from these models highlight differences in speech motor control in participants with ALS. The variability in acceleration of tongue and lip movements increases as speech performance declines, potentially reflecting physiological deviations due to the progression of ALS. Our findings suggest that acceleration is a more sensitive indicator of speech deterioration due to ALS than displacement and speed and may contribute to improved algorithm designs for monitoring disease progression from speech signals.

Karger@karger.com
www.karger.com/fpl

Correspondence to:
Jun Wang, jun.wang@austin.utexas.edu

Karger

Introduction

Amyotrophic lateral sclerosis (ALS) is a rapidly progressive neurodegenerative disease that impairs voluntary motor movements of the limbs and the bulbar musculature [1]. The process of identifying bulbar ALS at an early disease stage and monitoring bulbar disease progression remain ongoing clinical challenges [2–6]. Diverse clinical presentations and varying rates of disease progression both contribute to diagnostic uncertainty and delay [7]. An official diagnosis of ALS can take up to 14 months [8], after which the patient has a median survival period of 2–3 years from the time of symptom onset [9]. ALS is typically classified based on the initial site of clinical presentation (e.g., spinal vs. bulbar onset [1]). Regardless of the site of onset, disease progression will lead to the impairment of speech motor function to the point of anarthria and eventually the loss of all voluntary muscle control. Clinical measurements, such as self-rated functional scores, are currently used to stage disease progression. The ALS Functional Rating Scale-Revised (ALSFRS-R) is a clinical assessment that consists of twelve self-reported questions to measure motor function, three of which pertain to bulbar function [10]. In clinical practice, speaking rate (number of words produced per minute, wpm) and speech intelligibility (the percentage of words that are understood by a listener) are two standards currently used to assess overall speech performance of patients with dysarthria [11]. While these measures are essential for monitoring disability and treatment planning, symptoms of bulbar dysfunction are subtle early in disease progression and often manifest before perceptual characteristics are detectable [6, 12]. Although current measures are clinically useful, it has been suggested that instrumental or performance-based measures are necessary for detecting early changes in bulbar motor function [13].

To address this need, researchers have been exploring the clinical utility of speech biomechanics. The characteristics of articulator movements such as speed (magnitude of velocity), duration, and amplitude have been investigated as they relate to the severity of dysarthria in ALS. Compared to neurotypical controls, researchers have found that individuals with ALS show longer jaw and lip movement duration [14], reduced tongue speed [15, 16], and have larger jaw movement displacements [14, 16, 17], particularly in the later and more severe stages of ALS. Researchers have also examined within-individual variability of tongue and lip movements using the spatiotemporal index (STI), a widely accepted measure of speech motor control [18–20]. Conventionally, STI is measured across multiple *displacement* signals of articulatory movement data [18, 19, 21–23]. STI captures both spatial and temporal aspects of movement patterns during multiple repetitions of a single phrase to determine the overall stability of articulatory movement patterns [22, 24]. A greater STI value represents less consistent (more variable) articulatory movement patterns and is typically associated with an impaired motor system; a low STI suggests stability. Greater variability has been observed in speakers with dysarthria due to brain injury and in stuttering [21, 25]. Interestingly, researchers have observed pathologically low variability in speech movement patterns of individuals with mild to moderate ALS while speaking at a habitual speaking rate; this observation was rate dependent [20]. When the same speakers were asked to reduce or increase their speaking rate, variability increased. Previous findings of pathologically low movement variability in speakers with ALS while speaking at a habitual speaking rate have been interpreted as potential compensatory adjustments to improve speech motor control (i.e., stabilizing the jaw in response to impaired tongue function) [20]. Associations between STI and speech performance suggest that the variability of the displacement signal may not be sensitive to degraded speech motor control until substantial muscle atrophy occurs and compensatory adjustments are no longer feasible [19]. Less is known about how variability changes in other properties of motion that may be more sensitive to degeneration of the oral musculature. Additional investigations of tongue and lip motion pattern variability may give additional insight into speech motor control challenges associated with ALS.

To our knowledge, no prior research on speakers with ALS has reported on the *variability* of tongue acceleration or velocity, yet impaired tongue function is one of the earliest indicators of bulbar impairment and is most often involved in ALS [2, 26]. As ALS progresses, the tongue shape becomes less curvilinear and significantly reduces in size due to glossal atrophy [27]. Research using needle EMG, an instrumental technique used to evaluate denervation in patients with ALS, has shown reduced motor unit action potentials in subclinical ALS [26]. Moreover, hypoglossal motor neurons have been reported to be more severely impaired than facial or trigeminal motor neurons, even in the absence of perceptual symptoms of dysarthria in patients with ALS [28]. Abnormalities such as reduced force and endurance of the tongue muscles have also been reported by references [15, 28–32]. Although adequate tongue strength is necessary for the production of intelligible speech, measuring articulatory strength

Table 1. Number of participants within each dysarthria severity group based on speech performance scores

Severity rating	Intelligible speaking rate	Participants, *n*
Normal	>136 IWPM	22
Mild	94–135 IWPM	15
Moderate	53–93 IWPM	6
Severe	28–41 IWPM	1
Profound	<28 IWPM	6

Intelligible speaking rate scores were categorized based on a classification scheme proposed by Stipancic et al. [46]. IWPM, intelligible words per minute.

alone may be limited since there is high variation in muscle strength among individuals, and maximum strength values are not used during the production of speech [28, 31]. Neuromuscular problems such as reduced speed [16] and the loss of functional strength due to the progression of ALS negatively impact speakers' ability to control articulatory movements efficiently. Thus, acceleration (the rate of change of velocity), a measure directly driven by force [33], may better capture neuromuscular changes than both speed and displacement as ALS-induced dysarthric speech requires more biomechanical effort to reach fixed targets due to motor neuron degeneration [17, 34, 35]. Acceleration has been suggested to be a good approximation of force in a prior study of mandibular musculature [34]. Prior work has shown that acceleration of the jaw and lips is a useful measure for classifying individuals with ALS into pre-symptomatic and symptomatic groups, suggesting that acceleration may be a useful kinematic feature for assessing the severity of bulbar impairment [36, 37]. Although the tongue is regarded as the primary articulator for speech, acceleration of the tongue has not been extensively studied, yet it is directly related to the forces acting on the articulators and may be more sensitive to changes in motor control [34].

A growing body of research is focused on the use of machine learning to analyze statistical features extracted from speech motion data [4, 38–41]. The use of machine learning has shown success in the automatic detection of neurological diseases such as depression [39, 40], traumatic brain injury [42], and Parkinson's disease [41, 43, 44]. Such advanced analytical methods have been used to classify individuals as healthy/ALS using speech data [38]. Support vector machines are widely used for classification problems in machine learning and have been used to estimate bulbar impairment in ALS from articulatory data and facial movements [6, 36, 37]. Support vector regression (SVR) follows the same principles as support vector machines and has been used to predict the severity of dysarthria from speech acoustic and articulatory samples [4, 38]. In this study, we implemented a machine learning model to test the diagnostic efficacy of displacement, speed, and acceleration-based features.

The purpose of this study was to determine if acceleration-based measures are more sensitive to speech performance decline than speed-based measures. To do this, we sought to answer the following questions:

1. Do movement acceleration features (interquartile range, maximum, average, median, and standard deviation) outperform vertical displacement and speed features in predicting speech severity (i.e., intelligible speaking rate) using machine learning?
2. Does the variability (measured by STI) of the acceleration of tongue and lip motion patterns correlate more strongly with speech severity than the variability of vertical displacement and speed?

Based on the current literature, we hypothesized that our machine learning model would show higher performance given acceleration-based measures than when given speed-based measures. We also hypothesized that the variability of acceleration would be a better indicator of speech performance than the variability of vertical displacement and speed.

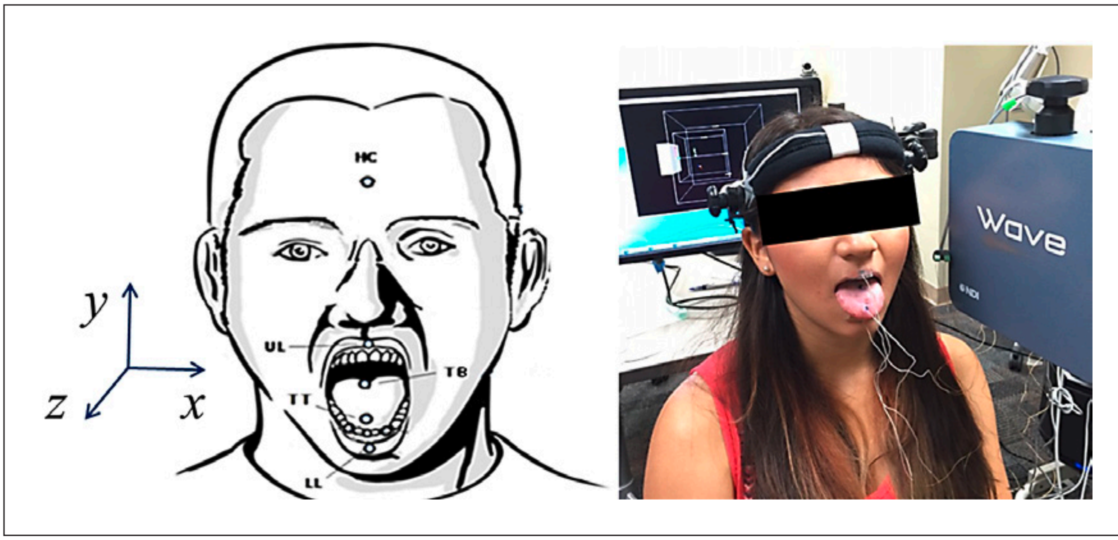
Materials and Methods

Participants

All data were collected under research protocols that have been approved by the University of Texas at Austin (IRB # 2019-12-0006). All participants provided written informed consent prior to the start of data collection. The sample for this study included 50 participants diagnosed with ALS (27 male, 23 female) with an average age of 58.31 years. The participants met the following conditions: (1) received a confirmed diagnosis of ALS by a board-certified neurologist; (2) were native speakers of English; (3) had no self-reported history of speech, language, or cognitive impairments; and (4) exhibited normal hearing capabilities. A hearing screening was performed prior to the start of formal data collection. To assess hearing, participants were screened at 25 dB HL at 1,000 Hz, 2,000 Hz, and 4,000 Hz.

Clinical Measures of Speech Severity

We measured speaking rate (wpm) and speech intelligibility (%) from sentence-level data using the Speech Intelligibility Test (SIT) Software [45]. The clinical assessment was administrated by a licensed speech-language pathologist. The SIT generates randomized sentences that increase from 5 to 15 words in length. The total scores were based on the 10 randomly generated sentences. The participants were judged to have a large range of severity with speech intelligibility scores ranging from 2.73% – 100% and speak-



Color version available online

Fig. 1. Illustration of sensor locations and NDI Wave System. HC, head center; TT, tongue tip; TB, tongue back; UL, upper lip; LL, lower lip.

ing rate ranging from 52.26 wpm–224.49 wpm. The average speaking rate was 135.03 wpm (SD = 44.16) and the average speech intelligibility was 86.81 (SD = 27.48). We selected intelligible speaking rate as the speech severity measure because it combines two areas of deficits that interact with one another: speech intelligibility and speaking rate. Moreover, intelligible speaking rate has been reported to be able to detect changes in speech function prior to speech intelligibility and speaking rate alone [46]. The intelligible speaking rate scores ranged from 2 to 218 intelligible words per minute (M = 122.93, SD = 57.95). This clinical measure has been used in a prior study [4]. Table 1 provides the number of participants that were judged to have mild, moderate, severe, or profound bulbar impairment based on intelligible speaking rate.

At each data collection session, participants produced 20 predefined short phrases such as "*I need to make an appointment*" and "*call me back when you can*" at their habitual speaking rate. The participants produced each of the 20 phrases four times. The sentences were selected because they are simple, commonly used in daily communication, and have previously been used to predict speech performance [5].

The NDI Wave System (Northern Digital Inc., Ontario, Canada) was used to collect articulatory data in real-time. The spatial accuracy of speech research system is 0.5 mm [47]. Audio recordings were collected at a 22-kHz sampling rate using a Shure Microflex microphone. The microphone was positioned approximately 15 cm from the mouth of the speaker. The NDI WaveFront Software allowed for time-synchronized acoustic and kinematic data recordings. We used the optimal four-sensor setup proposed by Wang and colleagues [48] to collect tongue and lip data. As illustrated in Figure 1, four sensor coils were secured to the following locations: tongue tip (TT) (5–10 mm from tongue apex), tongue back (TB) (20–30 mm from TT), and the vermillion borders of the upper and lower lips. PeriAcryl oral tissue adhesive (Glustich Inc., British Columbia, Canada) and medical tape were used to secure the sensors on the tongue and lips. To provide a point of reference and to isolate the articulatory movement data, subjects wore a plastic helmet with one sensor attached to the front.

Data Preprocessing

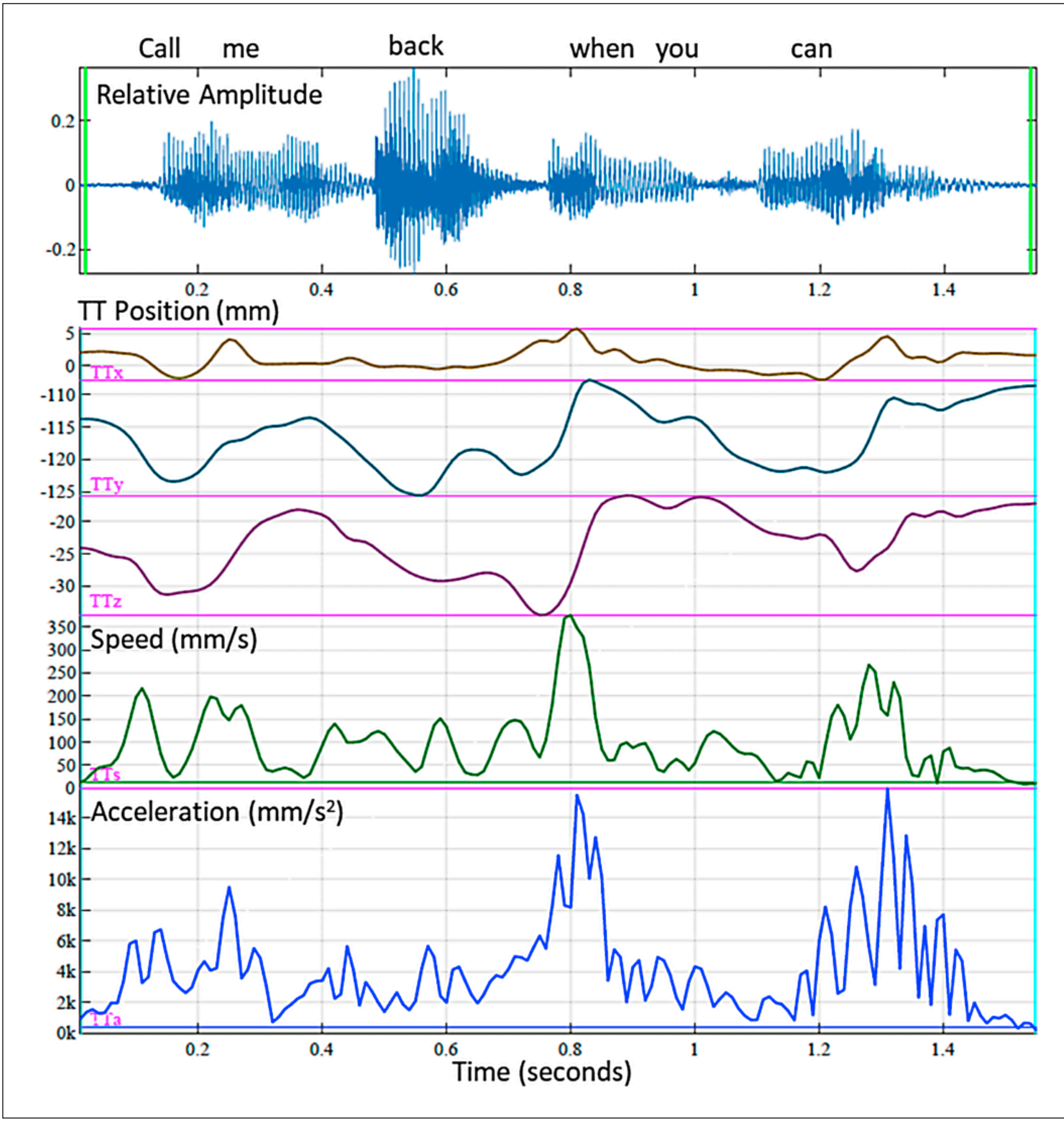
The data files collected from the NDI Wave System were imported into SMASH [49], a custom-written MATLAB (MathWorks, Natick, MA, USA)-based program used to analyze acoustic and kinematic data. Using this program, kinematic and time-matched acoustic data were parsed for further analysis. That is, each of the 20 phrases were identified and segmented from the original recording. The focal point of data segmentation was based on kinematic data. A tag was placed on the onset and offset of each phrase to represent the starting and stopping point of the associated articulatory motion. The motion data signal within the two tags was used for data analysis. Once the data were parsed, the x (lateral), y (vertical), and z (anterior-posterior) values of the four sensors (TT, TB, upper lip [UL], and lower lip [LL]) were used to calculate 3D speed and 3D acceleration measures. A 15-Hz Butterworth low-pass filter was applied to the positional data. An illustration of data parsing is provided in Figure 2.

Measurements: The STI of Displacement, Speed, and Acceleration

Tongue and lip movement data were analyzed using customized MATLAB programs (MathWorks, Natick, MA, USA). To calculate 3D acceleration, we can define Δs, Δv, and Δt as the change of displacement, change of velocity, and change of time in a 3D space. Note that for the purposes of this project, the direction of these signals was ignored as we were interested in speed (magnitude of velocity) and the magnitude of acceleration signals. First, instantaneous velocity v is calculated as follows:

$$v = \frac{||\Delta s||}{\Delta t} = \frac{\sqrt{(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2}}{\Delta t}. \tag{1}$$

Folia Phoniatr Logop 2023;75:23–34
DOI: 10.1159/000525514



Color version available online

Fig. 2. Segmented utterance (*"call me back when you can"*) produced by a speaker with ALS. Tongue tip (TT) positional signal is displayed with corresponding speed and acceleration signals. TTx, lateral tongue movements; TTy, superior-inferior movements; TTz, anterior-posterior tongue movements.

Second, where Δx is the change in position along the x-axis, Δy is the change in position along the y-axis, and Δz is the change in position along the z-axis. Since the measurements are based on the digital signals produced by the NDI Wave System, the differences are based on adjacent samples. Therefore, at time point n, $\Delta x = x[n] - x[n-1]$. Similarly, Δt is the difference in time between samples based on the sampling rate of the Wave System (100 Hz), thus $\Delta t = 0.01$. Instant acceleration is calculated as follows:

$$a = \frac{||\Delta^2 s||}{\Delta t^2} = \frac{\sqrt{\left(\Delta^2 x\right)^2 + \left(\Delta^2 y\right)^2 + \left(\Delta^2 z\right)^2}}{\Delta t^2}. \tag{2}$$

where $\Delta^2 x$ is calculated as $\Delta^2 x = \Delta x[n] - \Delta x[n-1] = x[n] - 2x[n-1] + x[n-2]$. For the y- and z-dimensions, $\Delta^2 y$ and $\Delta^2 z$ were calculated in the same manner.

We derived the STI of the displacement, instantaneous speed, and acceleration signals to measure trial-to-trial variability among multiple repetitions of a single phrase [22]. The following four steps were used to calculate STI:

1. Amplitude-normalize using z-scoring (subtract mean and divide by the standard deviation)
2. Time-normalize all speech samples to 1,000 data points by resampling
3. Calculate the standard deviation across productions for each 2% time interval of the signals
4. Sum the 50 standard deviation values from step three to calculate the final STI value

Data Analysis

The data were analyzed using two different approaches: (1) A machine learning algorithm (SVR) was performed to predict speech performance from features of acceleration and speed of the articulatory motion. We then investigated the degree to which each individual sensor contributed to the ability of each machine learning model to predict intelligible speaking rate. (2) Eight univariate linear regressions were carried out to investigate the relationship between intelligible speaking rate and the spatiotemporal variability of articulatory movement patterns.

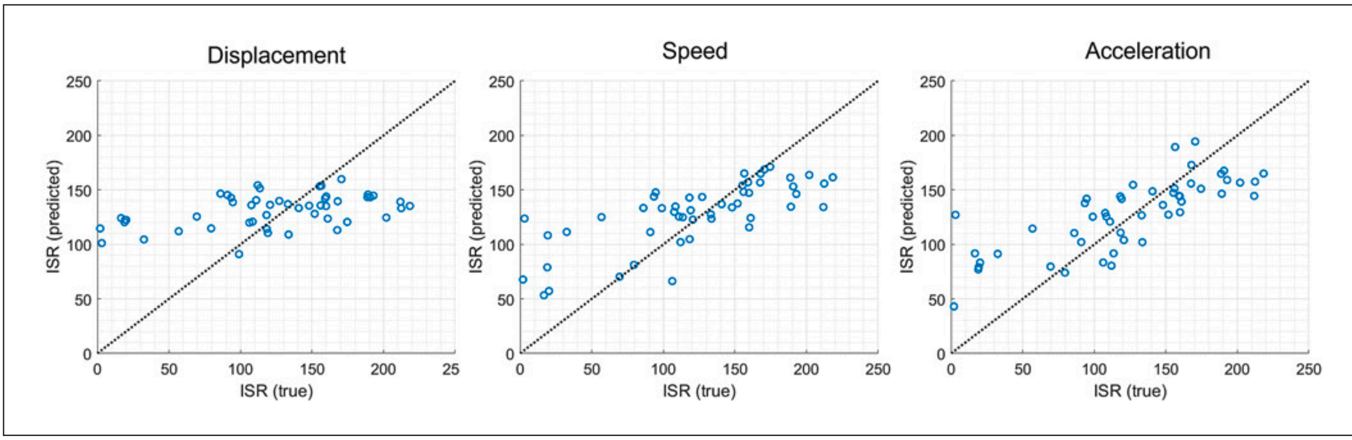
Machine Learning Analysis

One benefit to using SVR over a traditional linear regression is that SVR implements a supervised machine learning algorithm that is useful for recognizing patterns in complex data parameters without a substantial reliance on assumptions of data distributions [50]. Fundamentally, a machine learning-based approach is considered to be more robust, yields higher generalized performance than traditional least-squares regression models, and can handle more complex, nonlinear relationships among variables [51]. The goal of our SVR model was to predict intelligible speaking rate from statistical features of the speed and acceleration signals derived from articulatory motion of the tongue and lips. The model inputs were the velocity (i.e., $v_x = \Delta x/\Delta t$) and acceleration (i.e., $a_x = \Delta^2 x/\Delta t^2$) for each dimension and for each of the four sensors (see Measurements section). From each of the velocity and acceleration signals, we calculated five different statistics: interquartile range, maximum, average, median, and standard deviation. This process results in 80 velocity features and 80 acceleration features, that is, four velocity/acceleration signals per sensor for each dimension (x, y, and z) as well as instantaneous velocity/acceleration. We also assessed the y-dimension of the displacement signal for each sensor using the same procedures. All feature data were normalized by using z-scores. To test the performance of the SVR, a 5-fold cross-validation procedure was completed. This process was executed only once for simplicity and ease of interpretation. A random permutation was used to randomly assign participants into five groups (10 participants each). During each fold of the cross-validation procedure, one of these groups is removed for evaluation and the SVR is trained using data from the remaining forty participants. This process allows the model to be evaluated on the entire dataset while simultaneously avoiding any overlap in the training and test sets. Within each fold of the cross-validation procedure, the data are z-scored and principal component analysis was used to identify a 10-dimensional representation that captures maximum variances in the training set. Ten components were determined to sufficiently represent the variation in the data as 10 was the minimum number necessary to explain greater than 90% of the variance across the entire dataset, that is, averaged across both velocity and acceleration features. Although the entire dataset was used to select the appropriate number of components, which was held constant for this experiment, the actual mappings for these components were calculated solely using the training data, with unique mappings being learned for each stage of the five-fold cross-validation loop. Applying principle component analysis both helps reduce the dimensionality of the feature space and minimizes the risk of overfitting. These 10 principal components were passed to a SVR model, which utilizes a radial basis function kernel to map the input features to predictions of the corresponding participants' intelligible speaking rates.

In this study, we conducted both a sample-level and a participant-level classification analysis. At the sample level, each phrase produced by the participant was defined as a sample, whereas at the participant level, all of the speech samples from each subject were averaged prior to inputting the data into the machine learning model. To evaluate model performance, we derived the R^2 value and the root mean square error. Performance was averaged across the five folds to attain the out-of-sample performance across the entire dataset. To understand the degree to which each of the sensors and dimensions contributed more to the model in predicting intelligible speaking rate, we conducted an additional machine learning experiment. Instead of using all of the 80 displacement, velocity, and acceleration features (as we did in the previous experiment), we separated the feature data for each sensor prior to inputting the data into the machine learning model and compared the performances. A higher performance indicates a better contribution. To assess model performance, we followed the same procedures (i.e., cross-validation, dimensionality reduction, and SVR). To evaluate whether the differences in model performance are generalizable to the broader population of individuals with ALS, we separated the data using a data partitioning scheme based on random resampling. For this procedure, 100 different random 80–20 train/test splits were generated. The mean squared error results for each model across the 100 splits were compared using a corrected two-sample paired t test. This process helps prevent overlapping samples during the cross-validation procedure from increasing the false discovery rate of an experiment [52, 53].

Linear Regression Models (STI of Displacement, Speed, and Acceleration)

The MATLAB function *fitlm* was used for all statistical analyses. The *fitlm* function uses an ordinary least-squares method for



Color version available online

Fig. 3. Results from the SVR models at the participant level using movement displacement, speed, and acceleration features (interquartile range, maximum, average, median, and standard deviation) to predict intelligible speaking rate.

estimating model coefficients. A total of eight univariate linear regression models were performed. The spatiotemporal index of displacement ($STI_{displacement}$), the spatiotemporal index of instantaneous speed (STI_{speed}), and the spatiotemporal index of instantaneous acceleration ($STI_{acceleration}$) values were averaged across all 20 phrases for each of the four sensors. To limit the risk of false discoveries when testing multiple hypotheses, a Bonferroni-adjusted alpha criterion of 0.05/12 = 0.0041 was used to determine significance of the omnibus statistic. The efficacy of each model was determined based on (A) whether or not a statistical significant effect was observed, measured by the p value, and (B) the amount of variance in intelligible speaking rate explained by the model, measured by the coefficient of determination (R^2).

Table 2. Summary of SVR results for predicting intelligible speaking rate at the sample and participant level based on statistical features (interquartile range, maximum, average, median, and standard deviation) of displacement, speed, and acceleration

	Sample level		Participant level	
	R^2	RMSE	R^2	RMSE
Displacement	0.055	58.393	0.218	52.700
Speed	0.389	45.154	0.554	40.772
Acceleration	0.440	43.194	0.601	38.453

RMSE, root mean square error.

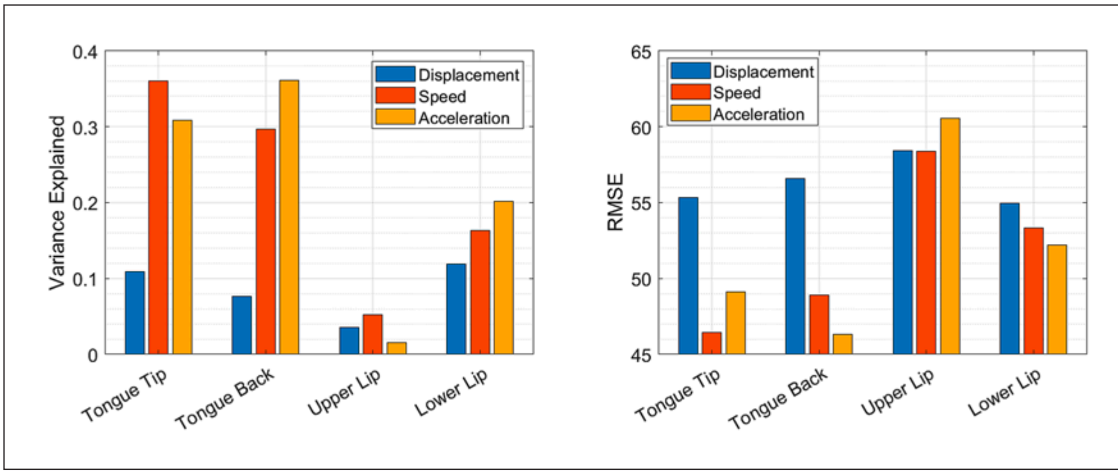
Results

Machine Learning Analyses: Movement Speed and Acceleration

The relationship between true and predicted intelligible speaking rate values generated by the acceleration and speed models are presented in Figure 3. The performance of each model is provided in Table 2. The contribution of each sensor to the model is presented in Figure 4. The SVR results indicated that the acceleration model achieved better overall performance than the displacement and speed models in terms of both R^2 and root mean square error; however, in the cross-validation procedure, the hypothesis testing within the repeated resampling procedure did not find these differences to be statistically significant at the 0.05 level. Model performance was higher at the participant level than the sample level for displacement, speed, and acceleration. Individual sensor analyses showed that, in isolation, the TT and TB models outperformed the upper and lower lip models. The UL contributed the least amount of useful information to our model for predicting intelligible speaking rate.

STI of Tongue and Lip Displacement, Speed, and Acceleration

Regression lines are presented in Figure 5. Descriptive statistics and linear regression model results are provided in Table 3. Please see the online supplementary material (see www.karger.com/doi/10.1159/000525514 for all online suppl. material) for regression models of speaking rate and speech intelligibility in isolation. The regression analysis found significant relationships between $STI_{acceleration}$ and intelligible speaking rate across three of the four sensors (TT, TB, and LL), with variability significantly increasing as speech integrity declined. This



Color version available online

Fig. 4. Results from the SVR models using displacement, speed, and acceleration movement features from individual sensors to predict intelligible speaking rate.

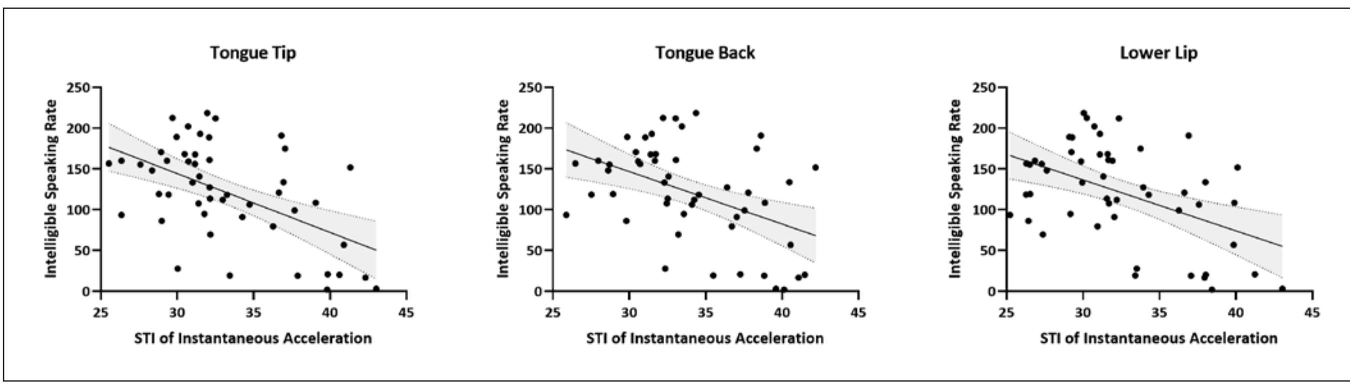


Fig. 5. Linear regression model results using the STI of instantaneous acceleration of the tongue tip and tongue back to predict intelligible speaking rate.

finding was consistent for the TT ($F(1,48) = 22.8$, $p < 0.001$); TB ($F(1,48) = 14.3$, $p < 0.001$); and LL ($F(1,48) = 14.1$, $p < 0.001$). There was no significant relationship between the STI_{speed} and intelligible speaking rate for any sensors. $STI_{acceleration}$ explained more variability in intelligible speaking rate than STI_{speed}. $STI_{acceleration}$ of the TT accounted for 32.2%, the TB accounted for 22.9%, and the LL accounted for 22.6%.

Although the primary focus of this study was on the variability of derivative measures of tongue and lip positional data rather than position ($STI_{displacement}$), this aspect must be considered because it has been previously used to measure articulatory pattern consistency in ALS and is a common stability measure in speech research [19, 20, 22, 54, 55]. We ran four univariate linear regression models for the $STI_{displacement}$ in the vertical dimension as previously described in prior work [19, 20, 22]. Our results showed no statistically significant relationships between the $STI_{displacement}$ and intelligible speaking rate for any sensors (TT [$F(1,48) = 0.47$, $p = 0.49$]; TB [$F(1,48) = 0.84$, $p = 0.37$]; UL [$F(1,48) = 2.16$, $p = 0.15$]; and LL [$F(1,48) = 0.52$, $p = 0.47$]).

Discussion

In this investigation, we examined the relationship between instantaneous acceleration and speed signals derived from the tongue and lips and a clinical measure of speech performance decline. Much of what is known

Table 3. Descriptive statistics for the variability of displacement, speed, and acceleration movement signals for each sensor and associated linear regression model

Model	Mean	SD	F	p value	R^2
Displacement					
TT	20.06	4.90	0.473	0.495	0.009
TB	19.48	5.46	0.836	0.365	0.017
UL	23.39	5.06	0.043	0.148	0.043
LL	21.36	5.08	0.010	0.473	0.011
Speed					
TT	30.34	4.13	4.17	0.046	0.079
TB	30.63	4.06	3.10	0.084	0.061
UL	35.05	4.09	5.14	0.027	0.097
LL	30.06	3.98	3.01	0.089	0.059
Acceleration					
TT	33.22	4.45	22.8	<0.001	0.322
TB	33.98	4.24	14.3	<0.001	0.229
UL	37.65	3.45	5.24	0.026	0.099
LL	32.47	4.55	14.1	<0.001	0.226

STI values were regressed on intelligible speaking rate.

about speech kinematics in ALS has been based on displacement or speed-based measures of articulatory activity. The present study extends knowledge about the use of kinematic measures for tracking speech performance decline by examining acceleration-based metrics as predictors of functional speech changes due to ALS. Our machine learning models showed higher performance accuracy when acceleration features were used to predict the severity of dysarthria. However, both speed and acceleration measures improved when predictions were aggregated across multiple phrases (as shown in Table 1), indicating that differences in tongue and lip motion patterns may be more reliably detected given larger speech samples. A previous study by Wang and colleagues (2018) aimed to predict intelligible speaking rate using statistical features derived from speech data and obtained a higher correlation (0.712) between the predicted and true ISR based than the current study. Their study used a large (thousands) of low-level (frame-to-frame), uninterpretable statistical features, which are useful for obtaining high performance [56]. The goal of the present study was to determine if acceleration was more sensitive than displacement and speed at predicting the decline in speech performance due to ALS. One advantage to using interpretable features is that we are able to compare which features contain more useful information for predicting variables of interest (e.g., speech performance decline). This approach allowed us to focus our investigation on specific changes in speech motor control due to ALS and can be applied to other neurogenic diseases. Based on the results of the current study, we suspect that acceleration may be more sensitive to small changes in speech motor control as an increased amount of effort is required to reach an articulatory goal [17, 57]. The change in instantaneous acceleration variability is likely multifactorial and potentially relates to both muscle weakness, control instability, and compensatory movement strategies. We must also note that the speed features that were included in our machine learning analysis (interquartile range, maximum, average, median, and standard deviation) were also sensitive indicators of speech performance decline. Similar findings have been previously reported in the literature by references [2, 14, 16, 17, 36, 37].

The single-sensor analyses provided information on how much predictive power individual sensors had in the SVR model. Not unexpectedly, our findings show that the tongue sensors provide the most relevant information for predicting ALS-induced dysarthria. This finding was evident for both acceleration and speed models. Current findings are in accordance with prior machine learning analyses demonstrating higher performance accuracy from tongue movement data [4, 6]. The tongue has also been consistently reported to be more impaired than the other articulators [2, 26, 28]. Because this study only included sensors on the tongue and lips, we are unable to determine the degree to which the jaw influences TB and LL movements in dysarthric speech. Additional studies are necessary to determine if the jaw is contributing to TB and LL movements to maintain intelligible speech.

In our secondary analysis, which analyzed the overall stability of tongue and lip movement patterns, we show that speakers with dysarthria due to ALS have less stable tongue and lip movements during the production of phrases when speech is more severely impaired. We must note that our study included fewer repetitions than the original formulation of STI, which was calculated based on 15 repetitions of a single phrase [22]. Although there is not an established minimum number of repetitions that should be used when calculating STI, having more repetitions improves reliability [58]. In our analysis, $STI_{acceleration}$ of the tongue and LL showed a statistically significant association with intelligible speaking rate. The amount of variance explained by $STI_{acceleration}$ reflects the potential importance of variability-based parameters of the tongue and LL in predicting speech performance decline in patients with ALS, particularly the TT. Increased variability of tongue and LL instantaneous *acceleration* patterns may reflect an increased physiologic effort to control the articulatory subsystem in response to bulbar deterioration.

Precise control of tongue and lip kinematics has been historically associated with a low STI value of movement amplitude relative to children and adults with speech motor impairments [22, 54, 59]. Our findings did not reveal a significant relationship between STI based on the amplitude signal and speech performance, which was not unexpected based on prior work. Kuruvilla-Dugdale and colleagues [19] reported increased displacement/amplitude variability in participants with severe ALS; however, this observation was only evident in two outliers. While muscle force was not directly investigated in this study, we suspect that the variability of instantaneous acceleration may be a more sensitive predictor of speech decline than the variability of displacement and instantaneous speed because of the direct and proportional relationship between acceleration and force as ALS impacts the ability to generate muscle force and control motor movements. Decreased muscle force and increased force variability of the lower extremities in individuals with ALS have been reported in the literature to negatively impact walking [60, 61]. This is not only evident during movements of the limbs; it has also been reported that patients with ALS have a lower rate of force recruitment in the tongue and lips compared to healthy controls [29]. Additional studies are needed to determine if increased variability in the instantaneous acceleration signal is related to a reduced ability to exert controlled forces during speech tasks, therefore, impacting the precision of speech patterns.

Our study provides support for the potential importance of assessing acceleration-based measures when investigating speech motor control in ALS. Additional studies are warranted to determine the relationship between tongue and lip force and the underlying neuromuscular mechanisms of dysarthric speech production due to ALS and whether the findings from this study are specific to ALS or extend to other neurogenic disorders (e.g., Parkinson's disease [62]). The goal of the present analysis was to predict intelligible speaking rate given acceleration-based kinematic measures. Additional work is required to explicate the clinical significance of these findings by evaluating longitudinal changes in speech function in ALS and comparing patients to healthy speakers. This research has the potential to improve algorithms for automatic assessments of ALS-induced dysarthria from speech signals [4, 63]. The current study provides additional support for the use of SVR models to assess speech motor control at a conceptual level. Earlier identification of physiological changes due to ALS may lead to the development of individualized therapeutic intervention strategies that can preserve functional speech for a longer duration and improve the quality of life for individuals with ALS.

Acknowledgments

We thank Dr. William Katz, Dr. Yana Yunusova, Jennifer McGlothlin, Brian D. Richburg, Satwik Dutta, Beiming Cao, Carolina Uzquiano, Alyssa Shrode, Brittany Shrode, and Bjorn Bleta for their support as well as the volunteering participants.

Statement of Ethics

This research has been approved by the Institutional Review Board (IRB) at the University of Texas at Austin (IRB # 2019-12-0006). Written informed consent was obtained from each subject after the details of the study were explained. This research was conducted ethically in accordance to the principles of the Declaration of Helsinki.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Funding Sources

This work was supported by the National Institutes of Health through grants R01DC013547 (PI: Green), R01DC016621 (PI: Wang), and R03DC013990 (PI: Wang).

Author Contributions

Kristin J. Teplansky conceived of the presented idea, processed, analyzed the data, and drafted the manuscript. Dr. Alan Wisler analyzed the data using machine learning and aided in interpreting the results. Drs. Jordan R. Green, Thomas Campbell, and Jun Wang were involved in the overall design and planning of the research and interpretation of these results. Drs. Daragh Heitzman and Sara G. Austin were involved in the implementation of the research. All the authors contributed to the revisions of the manuscript.

Data Availability Statement

The dataset is currently not publicly available but is under the plan to be ready for distribution in the future upon request. Further inquiries can be directed to the last author.

References

1 Kiernan MC, Vucic S, Cheah BC, Turner MR, Eisen A, Hardiman O, et al. Amyotrophic lateral sclerosis. Lancet. 2011 Mar;377(9769):942–55.
2 Green JR, Yunusova Y, Kuruvilla MS, Wang J, Pattee GL, Synhorst L, et al. Bulbar and speech motor assessment in ALS: challenges and future directions. Amyotroph Lateral Scler Frontotemporal Degener. 2013;14(7–8):494–500.
3 Meininger V. Getting the diagnosis right: beyond El Escorial. J Neurol. 1999;246 Suppl 3:III10–2.
4 Wang J, Kothalkar PV, Cao B, Heitzman D. Towards automatic detection of amyotrophic lateral sclerosis from speech acoustic and articulatory samples. Interspeech; 2016. p. 1195–9.
5 Wang J, Kothalkar PV, Kim M, Yunusova Y, Campbell TF, Heitzman D, et al. Predicting intelligible speaking rate of individuals with amyotrophic lateral sclerosis from a small number of speech acoustic and articulatory samples. Workshop Speech Lang Process Assist Technol. 2016:91–7.
6 Wisler A, Teplansky K, Green J, Yunusova Y, Campbell T, Heitzman D, et al. Speech-based estimation of bulbar progression in amyotrophic lateral sclerosis. Proceedings of the Eighth Workshop on Speech and Language Processing for Assistive Technologies. Minneapolis: Association for Computational Linguistics; 2019. p. 24–31.
7 van Es MA, Hardiman O, Chio A, Al-Chalabi A, Pasterkamp RJ, Veldink JH, et al. Amyotrophic lateral sclerosis. Lancet. 2017;390(10107):2084–98.
8 Matharan M, Mathis S, Bonabaud S, Carla L, Soulages A, Le Masson G. Minimizing the diagnostic delay in amyotrophic lateral sclerosis: the role of nonneurologist practitioners. Neurol Res Int. 2020 May;2020:14739811–8.
9 Turner MR, Scaber J, Goodfellow JA, Lord ME, Marsden R, Talbot K. The diagnostic pathway and prognosis in bulbar-onset amyotrophic lateral sclerosis. J Neurol Sci. 2010;294(1–2):81–5.
10 Cedarbaum JM, Stambler N, Malta E, Fuller C, Hilt D, Thurmond B, et al. The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. BDNF ALS Study Group (Phase III). J Neurol Sci. 1999 Oct;169(1–2):13–21.
11 Yorkston KM, Beukelman DR. Communication efficiency of dysarthric speakers as measured by sentence intelligibility and speaking rate. J Speech Hear Disord. 1981;46(3):296–301.
12 Rosenfield DB, Viswanath N, Herbrich KE, Nudelman HB. Evaluation of the speech motor control system in amyotrophic lateral sclerosis. J Voice. 1991 Jan;5(3):224–30.
13 Allison KM, Yunusova Y, Campbell TF, Wang J, Berry JD, Green JR. The diagnostic utility of patient-report and speech-language pathologists' ratings for detecting the early onset of bulbar symptoms due to ALS. Amyotroph Lateral Scler Frontotemporal Degener. 2017 Jul 3;18(5–6):358–66.
14 Yunusova Y, Weismer G, Westbury JR, Lindstrom MJ. Articulatory movements during vowels in speakers with dysarthria and healthy controls. J Speech Lang Hear Res. 2008 Jun;51(3):596–611.).
15 Langmore SE, Lehman ME. Physiologic deficits in the orofacial system underlying dysarthria in amyotrophic lateral sclerosis. J Speech Hear Res. 1994 Feb;37(1):28–37.
16 Shellikeri S, Green JR, Kulkarni M, Rong P, Martino R, Zinman L, et al. Speech movement measures as markers of bulbar disease in amyotrophic lateral sclerosis. J Speech Lang Hear Res. 2016 Oct;59(5):887–99.
17 Mefferd AS, Green JR, Pattee G. A novel fixed-target task to determine articulatory speed constraints in persons with amyotrophic lateral sclerosis. J Commun Disord. 2012;45(1):35–45.
18 Kuruvilla-Dugdale M, Chuquilin-Arista M. An investigation of clear speech effects on articulatory kinematics in talkers with ALS. Clin Linguist Phon. 2017 Oct 3;31(10):725–42.
19 Kuruvilla-Dugdale M, Mefferd A. Spatiotemporal movement variability in ALS: speaking rate effects on tongue, lower lip, and jaw motor control. J Commun Disord. 2017 May;67(May):22–34.
20 Mefferd AS, Pattee GL, Green JR. Speaking rate effects on articulatory pattern consistency in talkers with mild ALS. Clin Linguist Phon. 2014 Nov 11;28(11):799–811.
21 Smith A, Kleinow J. Kinematic correlates of speaking rate changes in stuttering and normally fluent adults. J Speech Lang Hear Res. 2000 Apr;43(2):521–36.
22 Smith A, Goffman L, Zelaznik HN, Ying G, McGillem C. Spatiotemporal stability and patterning of speech movement sequences. Exp Brain Res. 1995 Jun;104(3):493–501.
23 Mefferd AS. Associations between tongue movement pattern consistency and formant movement pattern consistency in response to speech behavioral modifications. J Acoust Soc Am. 2016;140(5):3728–37.
24 Smith A, Johnson M, McGillem C, Goffman L. On the assessment of stability and patterning of speech movements. J Speech Lang Hear Res. 2000 Feb;43(1):277–86.
25 McHenry MA. The effect of pacing strategies on the variability of speech movement sequences in dysarthria. J Speech Lang Hear Res. 2003;46(3):702–10.).
26 Finsterer J, Fuglsang-Frederiksen A, Mamoli B. Needle EMG of the tongue: motor unit action potential versus peak ratio analysis in limb and bulbar onset amyotrophic lateral sclerosis. J Neurol Neurosurg Psychiatry. 1997 Aug 1;63(2):175–80.
27 Cha CH, Patten BM. Amyotrophic lateral sclerosis: abnormalities of the tongue on magnetic resonance imaging. Ann Neurol. 1989 May;25(5):468–72.
28 DePaul R, Abbs JH, Caligiuri M, Gracco VL, Brooks BR. Hypoglossal, trigeminal, and facial motoneuron involvement in amyotrophic lateral sclerosis. Neurology. 1988 Feb 1;38(2):281–3.
29 DePaul R, Brooks BR. Multiple orofacial indices in amyotrophic lateral sclerosis. J Speech Hear Res. 1993 Dec;36(6):1158–67.
30 Weikamp JG, Schelhaas HJ, Hendriks JCM, de Swart BJM, Geurts ACH. Prognostic value of decreased tongue strength on survival time in patients with amyotrophic lateral sclerosis. J Neurol. 2012 Nov 24;259(11):2360–5.
31 Shellikeri S, Yunusova Y, Green JR, Pattee GL, Berry JD, Rutkove SB, et al. Electrical impedance myography in the evaluation of the tongue musculature in amyotrophic lateral sclerosis. Muscle Nerve. 2015 Oct;52(4):584–91.
32 Pizzorni N, Ginocchio D, Bianchi F, Feroldi S, Vedrodyova M, Mora G, et al. Association between maximum tongue pressure and swallowing safety and efficacy in amyotrophic lateral sclerosis. Neurogastroenterol Motil. 2020 Aug 26;32(8):e13859–11.
33 Newton I Isaac Newton's philosophiae naturalis principia mathematica. In: Koyré A, Cohen IB, editors. 3rd ed. Cambridge, MA: Harvard University Press; 1726.
34 Nelson WL, Perkell JS, Westbury JR. Mandible movements during increasingly rapid articulations of single syllables: preliminary observations. J Acoust Soc Am. 1984 Mar;75(3):945–51.
35 Nelson WL. Physical principles for economies of skilled movements. Biol Cybern. 1983 Feb;46(2):135–47.
36 Bandini A, Green JR, Wang J, Campbell TF, Zinman L, Yunusova Y. Kinematic features of jaw and lips distinguish symptomatic from presymptomatic stages of bulbar decline in amyotrophic lateral sclerosis. J Speech Lang Hear Res. 2018 May 17;61(5):1118–29.
37 Bandini A, Green JR, Zinman L, Yunusova Y. Classification of bulbar ALS from kinematic features of the jaw and lips: towards computer-mediated assessment. Interspeech. ISCAISCA; 2017. p. 1819–23.
38 An K, Kim M, Teplansky K, Green J, Campbell T, Yunusova Y, et al. Automatic early detection of amyotrophic lateral sclerosis from intelligible speech using convolutional neural networks. Interspeech. ISCAISCA; 2018. p. 1913–7.
39 Cummins N, Scherer S, Krajewski J, Schnieder S, Epps J, Quatieri TF. A review of depression and suicide risk assessment using speech analysis. Speech Commun. 2015 Jul;71:10–49.
40 Quatieri TF, Malyska N. Vocal-source biomarkers for depression: a link to psychomotor activity. Interspeech; 2012. p. 1058–61.

41 Camilo Vásquez Correa J, Rafael Orozdco Arroyave J, David Arias-Londoño J, Francisco Vargas Bonilla J, Nöth E. New computer aided device for real time analysis of speech of people with Parkinson's disease. Revista Facultad De Ingenieria-universidad De Antioquia. 2014;Vol. 72:87–103.
42 Falcone M, Yadav N, Poellabauer C, Flynn P. Using isolated vowel sounds for classification of mild traumatic brain injury. IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE; 2013. p. 7577–81.
43 Hahm S, Wang J. Parkinson's condition estimation using speech acoustic and inversely mapped articulatory data. Interspeech. 2015: 513–7.
44 Tsanas A, Little MA, McSharry PE, Spielman J, Ramig LO. Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease. IEEE Trans Biomed Eng. 2012 May;59(5):1264–71.
45 Yorkston KM, Beukelman DR, Hakel M, Dorsey M. Sentence intelligibility test for windows. Lincoln, NE; 2007.
46 Stipancic KL, Palmer KM, Rowe HP, Yunusova Y, Berry JD, Green JR. "You Say Severe, I Say Mild": toward an empirical classification of dysarthria severity. J Speech Lang Hear Res. 2021 Dec 13;64(12):4718–35.
47 Berry JJ. Accuracy of the NDI wave speech research system. J Speech Lang Hear Res. 2011; 54:1295–301.).
48 Wang J, Samal A, Rong P, Green JR. An optimal set of flesh points on tongue and lips for speech-movement classification. J Speech Lang Hear Res. 2016 Feb;59(1):15–26.
49 Green JR, Wang J, Wilson DL. SMASH: a tool for articulatory data processing and analysis. Lyon: Interspeech; 2013. p. 1331–5.
50 Cortes C, Vapnik V. Support-vector networks. Mach Learn. 1995;20(3):273–97.
51 Basak D, Pal S, Patranabis DC. Support vector regression. Neural Inform Process. 2007; 11(10):203–24.
52 Witten IH, Frank E, Hall MA. Data mining: practical machine learning tools and techniques. 3rd ed. Burlington, MA: Morgan Kaufmann; 2011. p. 629.
53 Gardner J, Brooks C. Statistical approaches to the model comparison task in learning analytics. 2017. p. 14.
54 Kleinow J, Smith A, Ramig LO. Speech motor stability in IPD: effects of rate and loudness manipulations. J Speech Lang Hear Res. 2001 Oct;44(5):1041–51.).
55 Chu SY, Barlow SM, Lee J, Wang J. Effects of utterance rate and length on the spatiotemporal index in Parkinson's disease. Int J Speech Lang Pathol. 2020;22(2):141–51.
56 Wang J, Kothalkar PV, Kim M, Bandini A, Cao B, Yunusova Y, et al. Automatic prediction of intelligible speaking rate for individuals with ALS from speech acoustic and articulatory samples. Int J Speech Lang Pathol. 2018 Oct 16;20(6):669–79.
57 Eshghi M, Stipancic KL, Mefferd A, Rong P, Berry JD, Yunusova Y, et al. Assessing oromotor capacity in ALS: the effect of a fixed-target task on lip biomechanics. Front Neurol. 2019 Dec 5;10:1288.
58 Wisler A, Goffman L, Zhang L, Wang J. Influences of methodological decisions on assessing the spatiotemporal stability of speech movement sequences. J Speech Lang Hear Res. 2022 Feb 9;65(2):538–54.
59 Green JR, Moore CA, Reilly KJ. The sequential development of jaw and lip control for speech. J Speech Lang Hear Res. 2002 Feb; 45(1):66–79.).
60 Hausdorff JM, Lertratanakul A, Cudkowicz ME, Peterson AL, Kaliton D, Goldberger AL, et al. Dynamic markers of altered gait rhythm in amyotrophic lateral sclerosis. J Appl Physiol. 2000;88(6):2045–53.
61 Jette DU, Slavin MD, Andres PL, Munsat TL. The relationship of lower-limb muscle force to walking ability in patients with amyotrophic lateral sclerosis. Phys Ther. 1999 Jul 1; 79(7):672–81.
62 Netsell R, Daniel B, Celesia GG. Acceleration and weakness in parkinsonian dysarthria. J Speech Hear Disord. 1975;40(2):170–8.
63 Wisler A, Teplansky K, Green JR, Austin SG, Wang J. Can machines learn continuous measures of speech severity from ordinal training labels? 33rd International Conference of the Florida Artifcial Intelligence Research Society. FLAIRS; 2020. p. 550–5.