

# Speech-based Estimation of Bulbar Regression in Amyotrophic Lateral Sclerosis

Alan Wisler<sup>1</sup>, Kristin Teplansky<sup>1,2</sup>, Jordan R. Green<sup>3</sup>, Yana Yunusova<sup>4</sup>,  
Thomas F. Campbell<sup>2</sup>, Daragh Heitzman<sup>5</sup>, Jun Wang<sup>1,2</sup>

<sup>1</sup>Speech Disorders & Technology Lab, Department of Bioengineering

<sup>2</sup>Callier Center for Communication Disorders

University of Texas at Dallas, Richardson, TX, USA

<sup>3</sup> Department of Communication Sciences and Disorders

MGH Institute of Health Professions, Boston, MA, USA

<sup>4</sup> Department of Speech-Language Pathology, University of Toronto, Toronto, ON, Canada

<sup>5</sup> MDA/ALS Center, Texas Neurology, Dallas, TX, USA

## Abstract

Amyotrophic Lateral Sclerosis (ALS) is a progressive neurological disease that leads to degeneration of motor neurons and, as a result, inhibits the ability of the brain to control muscle movements. Monitoring the progression of ALS is of fundamental importance due to the wide variability in disease outlook that exists across patients. This progression is typically tracked using the ALS functional rating scale - revised (ALSFRS-R), which is the current clinical assessment of a patient's level of functional impairment including speech and other motor tasks. In this paper, we investigated automatic estimation of the ALSFRS-R bulbar subscore from acoustic and articulatory movement samples. Experimental results demonstrated the ALSFRS-R bulbar subscore can be predicted from speech samples, which has clinical implication for automatic monitoring of the disease progression of ALS using speech information.

## 1 Introduction

Amyotrophic Lateral Sclerosis (ALS, also known as Lou Gehrig's disease) is a progressive neurological disease that destroys nerve cells and inhibits the normal voluntary motor function of the affected individual. The progression of this disease rapidly limits the patient's ability to perform normal daily tasks such as walking, speaking, and eventually even breathing. Although there is currently no cure for ALS, early detection and accurate tracking of disease progression is crucial to the planning of treatment strategies and therapeutic intervention (Kiernan et al., 2011). The currently used clinical measure for the disease progression is the patient self-reported ALSFRS-R score, which estimates the degree of functional

impairment across motor tasks such as speaking and walking, as well as common daily tasks such as getting dressed and climbing the stairs (Cedarbaum et al., 1999).

ALSFRS-R has a collection of 12 questions, with a total score ranging from 0 to 48, which is composed of three factors: bulbar functions, fine and gross motor functions, and respiratory function (Franchignoni et al., 2013). Bulbar functions include speaking, salivating, and swallowing. The efficacy of the ALSFRS-R for measuring motor-function and levels of self-sufficiency of individuals with ALS has been thoroughly demonstrated. The ALSFRS-R has shown high inter-rater reliability, test-retest reliability, and internal consistency (Cedarbaum and Stambler, 1997; Brinkmann et al., 1997). Additionally, the ALSFRS-R is highly correlated with the clinical stage of ALS (Balendra et al., 2014) and has been shown to be a useful predictor of patient survival (Magnus et al., 2002). Despite the utility and reliability of the ALSFRS-R, it is only able to quantify specific degradations in motor function along a five point scale. As such, it lacks the resolution to capture more subtle changes in motor function that can be observed through instrumentation-based measures (Allison et al., 2017).

Recently, there has been a surge of research using speech analytics to detect and track a range of neurological diseases such as Parkinson's (Orozco-Arroyave et al., 2016a,b; Hsu et al., 2017; Benba et al., 2015) and ALS (An et al., 2018; ill; Norel et al., 2018; Wang et al., 2016a,b, 2018). Efforts towards tracking disease progression in this area have typically focused on the estimation of speech specific measures such as speech intelligibility (Berisha et al., 2013; Kim et al., 2015),

speaking rate (Jiao et al., 2016; Martens et al., 2015), or severity (Tu et al., 2017; Asgari and Shafran, 2010). While these efforts have shown success in the ability to objectively measure functional changes directly related to speech, whether speech can be used to measure functional impairment along other tasks in ALS remains largely unexplored.

In this paper we sought to address this question by examining how well speech and articulation data can predict the ALSFRS-R bulbar subscore (ranges from 0 to 12). The long-term goal of this research is to develop objective measures for broad level motor function. At this early stage, we focused on the bulbar score first. To our knowledge, this paper is the first to predict ALSFRS-R (bulbar) score directly from speech information. Two regression models, a simple linear ridge regression model and a machine learning algorithm (support vector machine), were used in the regression analysis.

## 2 Data Collection

### 2.1 Participants

Sixty-six speakers diagnosed with ALS at early-onset participated in this study at up to four data collection sessions with an interval of four to six months. At each session, participants or caregivers completed the ALSFRS-R, which included the bulbar subscore. Speech intelligibility (percentage of understandable words, judged by listeners) and speaking rate (words produced per minute) were assessed by a speech-language pathologist using the Sentence Intelligibility Test (SIT) software (Dorsey et al., 2007). Intelligible speaking rate, called communication efficiency, was also calculated, which is the percentage of understandable words per minute (speech intelligibility  $\times$  speaking rate) (Yorkston and Beukelman, 1981). The whole data set was used for the basic correlation analysis between ALSFRS-R and speech performance measures, while the data from 28 participants were used for regression analysis. This subset includes 15 male and 13 female participants, whose age averaged 57.3 years with a standard deviation of 10.7 years.

### 2.2 Stimuli and Procedure

The participants were asked to produce 20 sentences in a fixed order, such as *I need some assistance* and *call me back when you can*. A com-

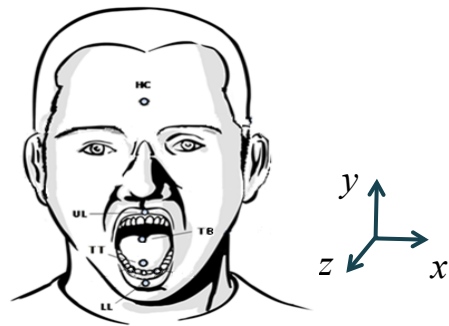


Figure 1: Sensor locations for the Wave system

plete list of the stimuli used for data collection is included in the Appendix. The sentences were selected because they are commonly used in augmentative and alternative communication (AAC) devices (Beukelman et al., 1984). All speech stimuli were presented on a TV screen in front of the participants. The stimuli were repeated for a total of four recordings at the participants habitual speaking rate among other speech tasks.

The NDI Wave System (Northern Digital Inc., Waterloo, Canada) was used to collect articulatory movement data with an accuracy of 0.5 mm (Berry, 2011). An optimal four sensor set-up (Wang et al., 2016c) was used to collect articulatory data from the tongue tip (TT, 5 mm from apex), tongue back (TB, 10 mm from TT), upper lip (UL, vermillion border) and lower lip (LL, vermillion border). The sensors were attached using nontoxic dental glue (PeriAcryl 90, GluStitch) or medical tape. A lightweight helmet with a 6 degree-of-freedom sensor served as a point of head reference. Prior to the start of each data collection session, the speakers had 3-5 minutes to adapt to the wired sensors prior to formal data collection. In this paper, we used  $x$ ,  $y$ , and  $z$  to represent lateral, vertical, and anterior-posterior movements, respectively. A visual depiction of the sensor locations and coordinate system is displayed in Figure 1. To capture acoustic signals simultaneously, a Shure Microflex microphone with a sampling rate of 22kHz was positioned approximately 15 cm from each speaker’s mouth.

### 2.3 Data Processing

Head rotation and translation movements were removed from articulatory data prior to analysis. A low pass filter of 15hz was applied to remove noise (Wang et al., 2016c). SMASH (Green et al., 2013a), a Matlab based software, was used seg-

ment the time-matched articulatory and acoustic data into individual phrase samples.

## 2.4 Relationship between ALSFRS-R scores and speech performance measures

In this section, we evaluated the relationship between traditional speech metrics, such as speaking rate and speech intelligibility, and ALSFRS-R scores. Because speech represents only a small component of the broad motor function assessed by the ALSFRS-R, we not only compared the relationship between speech and the ALSFRS-R as a whole, but also at the Bulbar subscore, which reflects the portion of the ALSFRS-R related to speaking, salivating and swallowing.

There are several important factors to consider when evaluating the relationship between these measures and ALSFRS-R score. First, neither the speech metrics nor the ALSFRS-R scores being compared are perfect measures of the underlying decline in motor function that they attempt to quantify. Speaking rate is highly sensitive to natural deviation between speakers and compensatory strategies that can mask changes in motor function (Green et al., 2013b). Speech intelligibility suffers from ceiling and floor effects that prevent it from tracking disease progression outside of a fixed severity range (Yorkston and Beukelman, 1981). Second, because the ALSFRS-R measures each motor component along a 5-point scale it cannot capture subtle changes to motor control that occur between points on this scale. Despite this limitation, the ALSFRS-R has been proven reliable in test-retest analysis (Cedarbaum and Stambler, 1997) and correlates highly with the clinical stage of individuals with ALS (Balendra et al., 2014).

Figure 2 displays the relationship between speech intelligibility, speaking rate, and intelligible speaking rate (ISR) and the ALSFRS-R bulbar subscore for participants in our data set. Although all three scatter plots show a correlation between the measures of speech and the ALSFRS-R bulbar subscore, there exists significant variability in the ALSFRS-R that cannot be explained by the measures of speech. This is particularly true of intelligibility, where participants could score as low as 4/12 of the ALSFRS-R bulbar subscore, while maintaining near-perfect intelligibility. Among the three speech measures, ISR had the highest correlation with ALSFRS Bulbar subscore.

To better understand the relationship between

these speech measures and the different components of the ALSFRS-R, we performed a correlation analysis between three measures of speech intelligibility, speaking rate, and ISR, and three ALSFRS-R component scores (Table 1). The three component scores were (a) the total score, which provides a broad assessment of motor function, (b) the bulbar subscore, which includes functions of motor control most closely related to speech including assessment of speaking, swallowing and salivating, and (c) the non-bulbar component, which is the difference between the total ALSFRS-R score and the bulbar subscore. This analysis found a strong correlation between all three measures of speech and the Bulbar subscore with all correlations between 0.5 and 0.7 and all p-values less than  $10^{-6}$ . Although there exists a statistically significant relationship between each of the speech measures and the total ALSFRS-R score, this significance disappears if the bulbar component is removed. Therefore this relationship is simply more evidence of the speech measures ability to track the bulbar component.

## 3 Methods

As mentioned earlier, this analysis was based on a subset of twenty eight participants from the previously described data set whose speech data has been manually parsed for automatic processing. Fifteen of the patients only made a single visit, six made two visits, five made three visits, and only two made the full four visits. Though the number of samples collected for each patient is usually 80 per session, some of the participants were not able to complete all of the recording tasks. In these cases predictions were made based on the reduced set of samples that were available.

### 3.1 Acoustic Features

The acoustic features used in this paper were based on the frame-level Mel-frequency cepstral coefficients (MFCCs). Although MFCCs were originally popularized due to their effectiveness in automatic speech recognition systems, they have recently seen increasing usage in a range of other speech assessment tasks, including the detection of motor speech disorders like Parkinson’s disease (Benba et al., 2015). As the mel cepstrum encodes spectral magnitude information related to the shape of the vocal tract, MFCC’s can capture articulatory changes resulting from conditions like

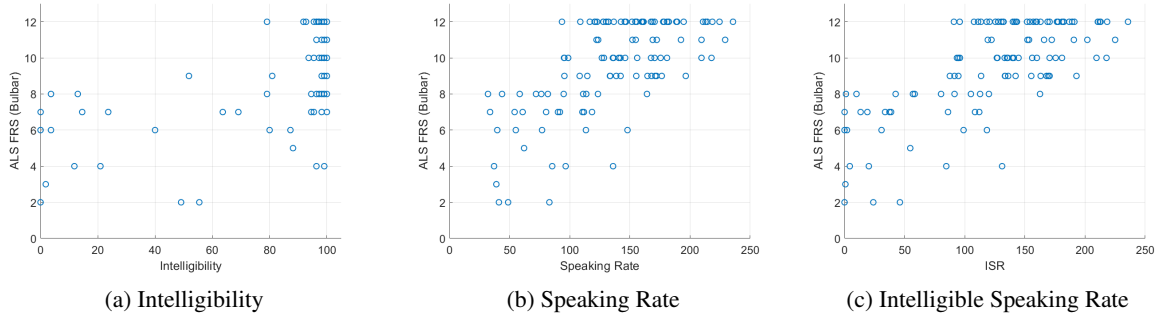


Figure 2: Scatter plots depicting the relationship between the three speech metrics and ALSFRS-R Bulbar subscore for each participant & recording session in the data set.

	ALSFRS-R		ALSFRS-R (Bulbar)		ALSFRS-R (non-Bulbar)	
	Correlation	p-value	Correlation	p-value	Correlation	p-value
Intelligibility	0.1960	0.0366	0.5840	$< 10^{-6}$	0.0005	0.9954
Speaking Rate	0.2419	0.0095	0.6422	$< 10^{-6}$	0.0286	0.7625
ISR	0.2331	0.0126	0.6957	$< 10^{-6}$	0.0002	0.9981

Table 1: Correlations between the speech measures and ALSFRS-R overall, bulbar, and non-bulbar scores.

depression (Williamson et al., 2014) or dysarthria (Fraile et al., 2008). For each frame we extracted 14 MFCCs, along with their first and second temporal derivatives  $\Delta$ MFCC and  $\Delta\Delta$ MFCC. From these 42 variables across time, we calculate 3 different summary statistics, mean, standard deviation and pairwise variability yielding a total of 126 features.

### 3.2 Articulatory Features

For a specific sensor, we have three positional arrays  $\mathbf{x} = [x_1, \dots, x_N]$ ,  $\mathbf{y} = [y_1, \dots, y_N]$ , and  $\mathbf{z} = [z_1, \dots, z_N]$  corresponding to dimensions  $x$ ,  $y$ , and  $z$ . For any index  $i \in [1, \dots, N - 1]$ , we can calculate the corresponding distance traveled as

$$d_i = \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2 + (z_{i+1} - z_i)^2} \quad (1)$$

and form the corresponding distance matrix

$$\mathbf{D} = [d_1, \dots, d_{N-1}]. \quad (2)$$

This distance matrix forms the basis for the articulation features used in this paper. From  $\mathbf{D}$ , we extracted eight summary statistics: mean, standard deviation, skewness, kurtosis, maximum, minimum, range, and pairwise variability. In addition to these baseline features, we considered three procedures for normalizing the feature based on measurements of the overall distance trajectory. The first normalization procedure was dividing the features by the overall distance traveled  $\sum \mathbf{D}$ , in

order to control for the distance of the overall articulation motion which was heavily dependent on the specific phrase being produced. The second and third normalization procedures attempted to control for the size of each individual’s articulation motion space, the first by normalizing between the maximum overall distance between any two points in  $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ . The second was forming a convex hull around the articulation path and normalizing based on the volume of the resulting hull. Combining the 8 different statistics and four normalization methods (including no normalization) with the four different sensors (Tongue-tip, tongue-back, lower lip and upper lip) yielded a total of 128 features.

To illustrate how these articulatory features might help assess motor function for different individuals, we plotted the articulation data from two different patients on opposite ends of the severity spectrum in our dataset (Figure 3). The first sample was from participant DA001 on his/her first visit. This participant experienced minimal decline in their speaking abilities and scored a perfect 12/12 on the ALSFRS-R bulbar subscore. The second sample was drawn from patient DA016 on her second visit, where she had a severe speech decline already scoring only 3/12 points on the Bulbar subsection. Figure 3 displays the tongue-tip articulation tracks for the two participants, along with a box plot comparing the distribution of the distance values between points

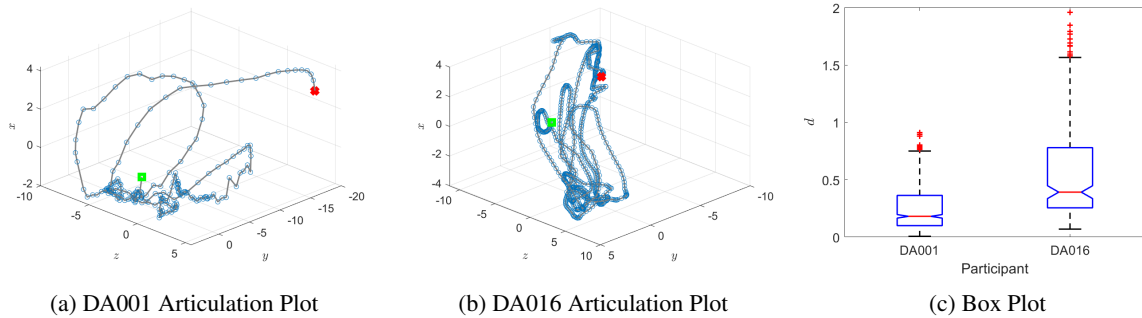


Figure 3: Articulation plots for two participants (a) DA001 (normal speech) and (b) DA016 (severe speech deficit) while speaking the phrase “I need to see a doctor”, along with (c) a box plot comparing the distributions of distances between adjacent points in the articulation track for each participant. In the articulation plots, the starting point of the articulatory motion is marked with a red cross and the end is marked with a green square.

for each participant. As shown in Figures 3a and 3b, participant DA016 has a significantly reduced articulation space relative to DA001, where DA001’s space had a volume that almost doubled DA016’s (335.96 vs. 181.55). Additionally, the box plot in Figure 3c shows a stark difference between the distribution of pairwise distances across the two participants. The distances for participant DA016 are significantly lower on average than DA016, indicating a pronounced reduction in the speed of articulation.

### 3.3 Regression Analysis

The regression analysis conducted in this experiment began with the  $3959 \times 254$  dimension feature matrix extracted via the procedure outlined in the previous section. To ensure the regression model’s ability to generalize to new speakers, it is evaluated by leave-one-speaker-out cross-validation. Thus at every stage of cross-validation, the model is trained using 27 participants and evaluated based on the single left-out participant. When a participant with multiple recording sessions is moved to the validation set, all sessions are moved to the validation set as a group and unique predictions are made and evaluated for each session.

All the data samples were z-scored (subtracted the mean and divided by the standard deviation) to obtain the normalized feature data. This procedure helped prevent the scale of different attributes affecting how much they contribute to the model.

Two regression models were used in this analysis, a simple ridge regression model and a support vector machine (SVM). Ridge regression is similar to ordinary least-squares regression, but utilizes an  $L_2$  regularization term in order to bet-

ter model data that is subject to multicollinearities (Hoerl and Kennard, 1970). Unlike traditional regression models that minimize observed training error, support vector regression (SVR) minimizes a generalization bound in order to ensure the model performs well on out-of-sample data (Basak et al., 2007). This factor, combined with the ability of SVMs to use non-linear kernels to model complex non-linear patterns in data, has made them widely used for both classification and regression problems. The SVMs used in this paper employed a linear kernel and were trained using the sequential minimal optimization (SMO) algorithm.

In addition to the baseline model (using all previously described acoustic and articulation features), we also tested the performance of other five feature groups, acoustic only, acoustic + lips, tongue, lips, and tongue + lips. The initial predictions were made on individual samples (phrases), and were then averaged to form a final prediction for each patient-session pair.

Two measures were used for the regression performance, root mean squared error (RMSE) and the correlation of the resulting set of predictions with the true ALSFRS-R (bulbar) scores. Low RMSE indicates that the small difference between the predicted and true ALSFRS-R values. High correlation indicates that changes in the predicted ALSFRS-R values are likely corresponding to a proportional changes in the true values.

## 4 Results

The results for each of the six feature group and two regression models are displayed in terms of both RMSE and correlation in Figure 4. The high-

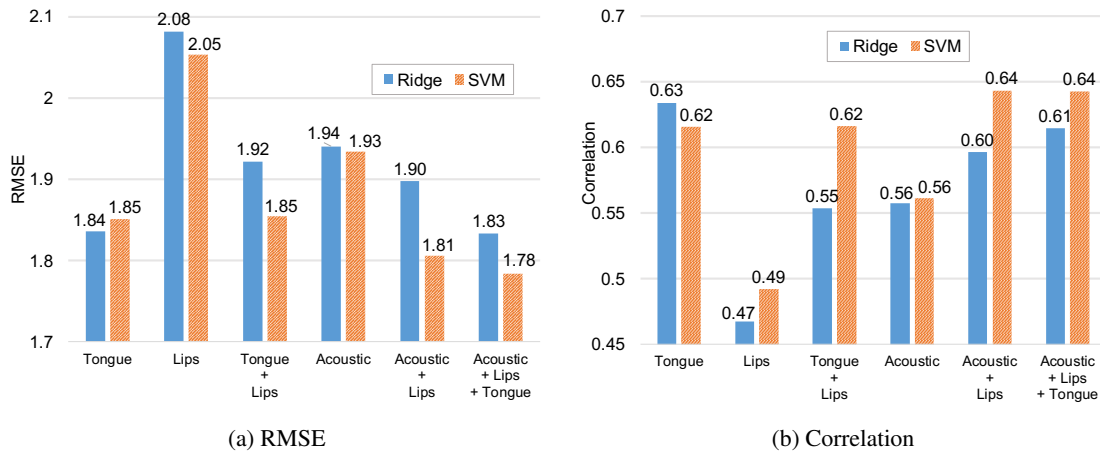


Figure 4: Bar graphs describing the performance of Ridge (blue) and SVM (orange) models across the six feature groupings described along the x-axis in terms of both root mean-squared error and correlation.

est performance was achieved by the SVR model using acoustic data along with all articulatory motion data (RMSE = 1.78,  $r = 0.64$ ).

Figure 4 indicated a few interesting findings. First, we found that the models trained on both the articulation motion data and the acoustic data tended to outperform either grouping by itself. This is consistent with the literature on both ISR prediction (Wang et al., 2016b, 2018) and ALS early detection (Wang et al., 2016a), which have shown the performance benefits of adding articulatory data to acoustic models.

In addition, the performance on data from tongue or lips separately shows that the tongue sensors were significantly more powerful than lips for predicting ALSFRS-R scores when viewed in isolation, which is not surprising, as the tongue is the primary articulator. Wang and colleagues also found tongue information outperformed lip information in predicting intelligible speaking rate for ALS (Wang et al., 2018).

Interestingly, when comparing the performances between the “Acoustic+Lips” group and the “All Features” group, we found that the addition of the tongue data (on top of acoustic and lip motion data) did not significantly improve the performance. Further studies, however, are required to verify this finding with a deeper analysis on the performance of additional tongue information on top of acoustic and lip information. Future research should investigate the degree to which non-invasive video-based measures of lip motion can be substituted for the more traditional motion-sensors. This finding supports the idea that a mobile app for recording speech and lip motion (via

a webcam) would be beneficial for future home-based data collection from patients.

Finally, when comparing the performance of the two regression models that were tested, SVM tended to perform slightly better than the ridge regression models. The lone exception to this was in the case of tongue-only articulation data, where the ridge regression model slightly outperformed the SVM. Future work will involve more complicated models such as convolutional neural networks (CNNs), which have recently shown potential in ALS early detection (An et al., 2018).

## 5 Conclusion

This paper explored automatic estimation of the ALSFRS-R bulbar score from speech information, where both acoustic and articulatory motion data collected during speech production were used. Two regression models, support vector regression and ridge regression, were applied on six different feature groups/sets. The highest performance was achieved by the SVR model using acoustic data along with all articulatory motion data. To our knowledge, for the first time, we demonstrated the feasibility of automatic prediction of ALSFRS-R bulbar score from speech samples. Future research on this topic will focus on the degree to which non-speech information can be included to predict ALS motor function decline more broadly.

## Acknowledgments

This work was supported by the National Institutes of Health (NIH) under award numbers R01DC013547 and R03DC013990.

## References

- Kristen M Allison, Yana Yunusova, Thomas F Campbell, Jun Wang, James D Berry, and Jordan R Green. 2017. The diagnostic utility of patient-report and speech-language pathologists' ratings for detecting the early onset of bulbar symptoms due to als. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, 18(5-6):358–366.
- KwangHoon An, Myungjong Kim, Kristin Teplansky, Jordan R Green, Thomas F Campbell, Yana Yunusova, Daragh Heitzman, and Jun Wang. 2018. Automatic early detection of amyotrophic lateral sclerosis from intelligible speech using convolutional neural networks. *Proc. Interspeech 2018*, pages 1913–1917.
- Meysam Asgari and Izhak Shafran. 2010. Predicting severity of parkinson's disease from speech. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pages 5201–5204. IEEE.
- Rubika Balendra, Ashley Jones, Naheed Jivraj, Catherine Knights, Catherine M Ellis, Rachel Burman, Martin R Turner, P Nigel Leigh, Christopher E Shaw, and Ammar Al-Chalabi. 2014. Estimating clinical stage of amyotrophic lateral sclerosis from the als functional rating scale. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, 15(3-4):279–284.
- Debasish Basak, Srimanta Pal, and Dipak Chandra Patranabis. 2007. Support vector regression. *Neural Information Processing-Letters and Reviews*, 11(10):203–224.
- Achraf Benba, Abdelilah Jilbab, and Ahmed Hammouch. 2015. Detecting patients with parkinson's disease using mel frequency cepstral coefficients and support vector machines. *International Journal on Electrical Engineering and Informatics*, 7(2):297.
- Visar Berisha, Rene Utianski, and Julie Liss. 2013. Towards a clinical tool for automatic intelligibility assessment. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2825–2828. IEEE.
- Jeffrey J Berry. 2011. Accuracy of the NDI wave speech research system. *Journal of Speech, Language, and Hearing Research*, 54:1295–1301.
- David R. Beukelman, Kathryn M. Yorkston, Miguel Poblete, and Carlos Naranjo. 1984. Frequency of word occurrence in communication samples produced by adult communication aid users. *Journal of Speech and Hearing Disorders*, 49:360–367.
- James R Brinkmann, Patricia Andres, Michelle Mendoza, and Mohammed Sanjak. 1997. Guidelines for the use and performance of quantitative outcome measures in ALS clinical trials. *Journal of the Neurological Sciences*, 147(1):97–111.
- Jesse M Cedarbaum and Nancy Stambler. 1997. Performance of the amyotrophic lateral sclerosis functional rating scale (alsfrs) in multicenter clinical trials. *Journal of the Neurological Sciences*, 152:s1–9.
- Jesse M Cedarbaum, Nancy Stambler, Errol Malta, Cynthia Fuller, Dana Hilt, Barbara Thurmond, Arline Nakanishi, BDNF ALS Study Group, 1A complete listing of the BDNF Study Group, et al. 1999. The alsfrs-r: a revised als functional rating scale that incorporates assessments of respiratory function. *Journal of the Neurological Sciences*, 169(1-2):13–21.
- Matt Dorsey, Kathryn Yorkston, David Beukelman, and Mark Hakel. 2007. Speech intelligibility test for windows.
- Rubén Fraile, Juan Ignacio Godino-Llorente, Nicolás Sáenz-Lechón, Víctor Osmar-Ruiz, and Pedro Gómez Vilda. 2008. Use of cepstrum-based parameters for automatic pathology detection on speech-analysis of performance and theoretical justification. volume 1, pages 85–91.
- Franco Franchignoni, Gabriele Mora, Andrea Giordano, Paolo Volanti, and Adriano Chiò. 2013. Evidence of multidimensionality in the alsfrs-r scale: a critical appraisal on its measurement properties using rasch analysis. *J Neurol Neurosurg Psychiatry*, 84(12):1340–1345.
- Jordan R Green, Jun Wang, and David L Wilson. 2013a. SMASH: a tool for articulatory data processing and analysis. In *Interspeech*, pages 1331–1335. IEEE.
- Jordan R Green, Yana Yunusova, Mili S Kuruvilla, Jun Wang, Gary L Pattee, Lori Synhorst, Lorne Zinman, and James D Berry. 2013b. Bulbar and speech motor assessment in als: Challenges and future directions. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, 14(7-8):494–500.
- Arthur E Hoerl and Robert W Kennard. 1970. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67.
- Sih-Chiao Hsu, Yishan Jiao, Megan J McAuliffe, Visar Berisha, Ruey-Meei Wu, and Erika S Levy. 2017. Acoustic and perceptual speech characteristics of native mandarin speakers with parkinson's disease. *The Journal of the Acoustical Society of America*, 141(3):EL293–EL299.
- Yishan Jiao, Ming Tu, Visar Berisha, and Julie Liss. 2016. Online speaking rate estimation using recurrent neural networks. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5245–5249. IEEE.

- Matthew C Kiernan, Steve Vucic, Benjamin C Cheah, Martin R Turner, Andrew Eisen, Orla Hardiman, James R Burrell, and Margaret C Zoing. 2011. Amyotrophic lateral sclerosis. *The lancet*, 377(9769):942–955.
- Myung Jong Kim, Younggwon Kim, and Hoirin Kim. 2015. Automatic intelligibility assessment of dysarthric speech using phonologically-structured sparse linear model. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 23(4):694–704.
- T Magnus, M Beck, R Giess, I Puls, M Naumann, and KV Toyka. 2002. Disease progression in amyotrophic lateral sclerosis: predictors of survival. *Muscle & Nerve: Official Journal of the American Association of Electrodiagnostic Medicine*, 25(5):709–714.
- Heidi Martens, Tomas Dekens, Gwen Van Nuffelen, Lukas Latacz, Werner Verhelst, and Marc De Bodt. 2015. Automated speech rate measurement in dysarthria. *Journal of Speech, Language, and Hearing Research*, 58(3):698–712.
- Raquel Norel, Mary Pietrowicz, Carla Agurto, Shay Rishoni, and Guillermo Cecchi. 2018. Detection of amyotrophic lateral sclerosis (ALS) via acoustic analysis. In *Proc. Interspeech*, pages 377–381.
- JR Orozco-Arroyave, F Hönig, JD Arias-Londoño, JF Vargas-Bonilla, K Daqrouq, S Skodda, J Rusz, and E Nöth. 2016a. Automatic detection of parkinson’s disease in running speech spoken in three different languages. *The Journal of the Acoustical Society of America*, 139(1):481–500.
- Juan Rafael Orozco-Arroyave, JC Vdsquez-Correa, Florian Hönig, Julián D Arias-Londoño, JF Vargas-Bonilla, Sabine Skodda, Jan Rusz, and E Noth. 2016b. Towards an automatic monitoring of the neurological state of parkinson’s patients from speech. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6490–6494. IEEE.
- Ming Tu, Visar Berisha, and Julie Liss. 2017. Objective assessment of pathological speech using distribution regression. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5050–5054. IEEE.
- Jun Wang, Prasanna V Kothalkar, Beiming Cao, and Daragh Heitzman. 2016a. Towards automatic detection of amyotrophic lateral sclerosis from speech acoustic and articulatory samples. In *Interspeech*, pages 1195–1199.
- Jun Wang, Prasanna V Kothalkar, Myungjong Kim, Andrea Bandini, Beiming Cao, Yana Yunusova, Thomas F Campbell, Daragh Heitzman, and Jordan R Green. 2018. Automatic prediction of intelligible speaking rate for individuals with ALS from speech acoustic and articulatory samples. *International journal of speech-language pathology*, 20(6):669–679.
- Jun Wang, Prasanna V Kothalkar, Myungjong Kim, Yana Yunusova, Thomas F Campbell, Daragh Heitzman, and Jordan R Green. 2016b. Predicting intelligible speaking rate in individuals with amyotrophic lateral sclerosis from a small number of speech acoustic and articulatory samples. In *Workshop on Speech and Language Processing for Assistive Technologies*, volume 2016, page 91. NIH Public Access.
- Jun Wang, Ashok Samal, Panying Rong, and Jordan R Green. 2016c. An optimal set of flesh points on tongue and lips for speech-movement classification. *Journal of Speech, Language, and Hearing Research*, 59(1):15–26.
- James R Williamson, Thomas F Quatieri, Brian S Helfer, Gregory Ciccarelli, and Daryush D Mehta. 2014. Vocal and facial biomarkers of depression based on motor incoordination and timing. In *Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge*, pages 65–72. ACM.
- Kathryn M Yorkston and David R Beukelman. 1981. Communication efficiency of dysarthric speakers as measured by sentence intelligibility and speaking rate. *Journal of Speech and Hearing Disorders*, 46(3):296–301.

## Appendix

Index	Phrase
1	<i>I love you</i>
2	<i>Good afternoon</i>
3	<i>I changed my mind</i>
4	<i>I need some assistance</i>
5	<i>Come back again</i>
6	<i>Nice to see you</i>
7	<i>Not very good today</i>
8	<i>I need to see a doctor</i>
9	<i>I am fine</i>
10	<i>Thanks for stopping by</i>
11	<i>How are you</i>
12	<i>Call me back when you can</i>
13	<i>Great to see you again</i>
14	<i>Can I have that</i>
15	<i>This is an emergency</i>
16	<i>What are you doing</i>
17	<i>Good-bye</i>
18	<i>When will you be back</i>
19	<i>Give me one minute</i>
20	<i>I need to make an appointment</i>

Table 2: List of stimuli used for data collection.