



HHS Public Access

Author manuscript

Amyotroph Lateral Scler Frontotemporal Degener. Author manuscript; available in PMC
2021 February 01.

Published in final edited form as:

Amyotroph Lateral Scler Frontotemporal Degener. 2020 February ; 21(1-2): 42–50.

doi:10.1080/21678421.2019.1697888.

Reliability and Validity of Speech & Pause Measures during Passage Reading in ALS

CAROLINA BARNETT^{1,2}, JORDAN R. GREEN^{3,4}, REEMAN MARZOUQAH⁵, KAILA L. STIPANCIC³, JAMES D. BERRY⁶, LAWRENCE KORNGUT⁷, ANGELA GENGE⁸, CHRISTEN SHOESMITH⁹, HANNAH BRIEMBERG¹⁰, AGESSANDRO ABRAHAO^{11,12}, SANJAY KALRA^{13,14}, LORNE ZINMAN^{11,12,15}, YANA YUNUSOVA^{5,12,15,16}

¹Division of Neurology, Department of Medicine, University of Toronto and University Health Network, Toronto, ON, Canada

²Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

³Department of Communication Sciences and Disorders, MGH Institute of Health Professions, Boston, MA, United States

⁴Speech and Hearing Biosciences and Technology Program, Harvard University, Cambridge, MA, United States

⁵Department of Speech-Language Pathology, University of Toronto, Toronto, Ontario, Canada

⁶Massachusetts General Hospital, Boston, Massachusetts, United States of America

⁷University of Calgary, Calgary, Canada,

⁸Montreal Neurological Institute, Montreal, QC, Canada

⁹Department of Clinical Neurological Sciences, University of Western Ontario, London, Ontario, Canada

¹⁰Division of Neurology, University of British Columbia, Vancouver, BC, Canada

¹¹Department of Medicine, Division of Neurology, Sunnybrook Health Sciences Centre, Toronto, Canada

¹²Hurvitz Brain Sciences Program, Sunnybrook Research Institute, Ontario, Canada

¹³Neuroscience and Mental Health Institute, University of Alberta, Edmonton, Canada

¹⁴Division of Neurology, University of Alberta, Edmonton, Canada

¹⁵L.C. Campbell Cognitive Neurology Research Unit, Sunnybrook Research Institute, University of Toronto, Toronto, Canada

¹⁶University Health Network—Toronto Rehabilitation Institute, Toronto, Ontario, Canada

Correspondence: Yana Yunusova, PhD., Department of Speech-Language Pathology, University of Toronto, 160 - 500 University Avenue, Toronto, ON, Canada, M5G 1V7, Tel: 416 978-6890, Fax: 416 978-1596, yana.yunusova@utoronto.ca.

Disclosure of interest: The authors report no conflict of interest.

Abstract

Objective: The use of speech measures is becoming a common practice in the assessment of bulbar disease progression in amyotrophic lateral sclerosis (ALS). This study aimed to establish psychometric properties (e.g., reliability, validity, sensitivity, specificity) of speech and pause timing measures during a standardized passage.

Methods: A large number of passage recordings (ALS N=775; Neurotypical controls N=323) was analyzed using a semi-automatic method (Speech and Pause Analysis, SPA).

Results: The results revealed acceptable reliability of the speech and pause measures across repeated recording by the control participants. Strong construct validity was established via significant group differences between patients and controls and correlation statistics with clinical measures of overall ALS and bulbar disease severity. Speaking rate, pause events, and mean pause duration were able to detect ALS participants at the presymptomatic stage of bulbar disease with a good discrimination ability (AUC 0.81).

Conclusions: Based on the current psychometric evaluation, performing passage recording and speech and pause timing analysis was deemed useful for detecting early and progressive changes associated with bulbar ALS.

Keywords

Bulbar ALS; speaking rate; pauses; reliability; validity; passage reading

Introduction

The emergence of bulbar signs is an important milestone in the progression of amyotrophic lateral sclerosis (ALS), leading to significant functional effects on speech communication, swallowing, and overall quality of life (1). Monitoring the onset and progression of bulbar disease is of great importance for predicting the course of ALS and patient survival, as well as for planning management practices (e.g., feeding tube insertion, assistive communication devices, etc.). Sensitive measures of bulbar disease are also critically needed to provide objective outcomes for clinical trials (2,3). The current clinical practices of bulbar disease monitoring are limited (4), with the ALS Functional Rating Scale – Revised (ALSFRRS-R) (5) being used as the only means of bulbar assessment in nearly 90% of clinics in the USA (6). However, one of the major limitations of the bulbar subscore of the ALSFRS-R is its insensitivity to the early stages of bulbar disease (7), when bulbar monitoring is arguably of the utmost importance. Speaking rate has been recommended as a preferred objective measure of bulbar decline by speech-language pathologists (SLP). SLPs employ speaking rate monitoring in order to predict the loss of speech intelligibility and to time introduction of assistive speech technologies (8,9).

To date, speaking rate has often been measured using the Speech Intelligibility Test (SIT) (10), a computerized assessment during which a set of 11 sentences spoken aloud is recorded, timed by locating the sentence onsets and offsets, and speaking rate calculated as the number of words per minute (WPM) of speech (11,12). Clinicians hesitate to use the SIT in busy ALS clinics, however, for the amount of work needed to obtain the measures. A

passage reading task can be used as an excellent alternative to the SIT, as it can provide a measure of the overall speaking rate as well as the more specific measures of speaking time (i.e., articulatory rate) and pause duration (13,14). Speaking and pause measures obtained during this task have been shown to distinguish bulbar disease at different stages, with or without respiratory deficits (14). In a recent preliminary analysis, a percent pause measure demonstrated sensitivity to presymptomatic stages of bulbar disease (7). Moreover, the same measure detected changes after treatment in a Phase II clinical trial (15,16). However, psychometric properties (e.g., reliability, validity) of the speaking and pausing measures, as well as their sensitivity to the earliest stages of the disease, have not been established in a large patient cohort. Additionally, their properties have not been directly compared to measures derived from the SIT (17). Delineation of these properties is an essential step in ensuring the acceptance of these measures into clinical practice and trial design.

In this study, we assessed the psychometric properties of the speaking and pause measures obtained in a passage reading task in a large cohort of speakers with ALS and neurologically healthy controls. Specifically, we assessed (1) test-retest reliability of the measures as well as the minimally detectable change (MDC), which is the change in scores outside of the measurement error; (2) construct validity of the measures in speakers with ALS; as well as (3) sensitivity, specificity, and predictive ability of the measures to early (presymptomatic) phase of the bulbar disease, defined by the ALSFRS-R bulbar subscore. Based on earlier works, we hypothesized that speaking and pause measures in a passage would reveal adequate psychometric properties and ability to detect bulbar signs prior to the onset of obvious bulbar symptoms.

Methods

Participants

A data set of 1098 passage recordings was compiled from 3 observational longitudinal studies of ALS. The total number of participants with ALS and neurotypical controls was 526, but we removed 3 controls who had slurred speech on the recordings (total N=253). Summary demographic and clinical information is presented in Table 1. 172 patients and 46 controls were recorded more than once. The time interval between recordings varied from 3 to 6 months. Participants with ALS were diagnosed with possible, probable, and definite ALS, as defined by the El Escorial Criteria (18). The ALSFRS-R was administered to document overall and bulbar functional disease effects. The Speech Intelligibility Test (SIT) (10) was collected to estimate bulbar disease severity via measures of speech intelligibility (i.e., % words transcribed correctly by an unfamiliar listener) and speaking rate (WPM). As part of their clinical assessment, patients with ALS completed a pulmonary function test which supplied % Forced Vital Capacity (% FVC).

Protocol/Procedures

All participants read a standardized Bamboo passage designed with the purpose of aiding automatic pause boundary detection (see Appendix). The passage contains voiced consonants (like 'b', 'd') at word and phrase boundaries in order to enhance automatic pause identification (13). Participants were familiarized with the passage by reading it silently a

few times. During recording, they were instructed to read the passage aloud in their natural reading manner, with normal rate and loudness. ‘Theatrical’ and rushed readings were discouraged and rerecorded when occurred.

Digital acoustic recordings were obtained at 44 kHz/ 16-bit resolution. They were analyzed with Speech Pause Analysis (SPA) software, a semi-automated Matlab pause identification procedure (13). The minimum speech duration threshold value was set at 25 msec and the minimum pause threshold was chosen to be 200 msec for controls and 300 msec for patients (19). As a result, the boundaries associated with each pause, below the signal amplitude threshold, were identified on the waveform (see Figure 1). A listening test was conducted for each recording to verify the accuracy of the reading and pause locations by the software. Misread words were included in the analyses without editing. Word repetitions or insertion as well as the pause that immediately followed any of these events were deleted from the waveforms; however, these occurrences were rare (see (14)).

SPA Measures

The SPA software output comprised the time stamps associated with each speech and pause events and their summary statistics. The primary variables of interest were selected based on their ability to distinguish patients with ALS from controls in our previous study (14). They included:

1. Speaking rate, Passage (words per minute, WPM): a measure of overall rate of speaking during the passage (total number of words =97), includes all pauses within the passage.
2. Total duration (sec): the duration of the recording from the onset of the first sentence to the offset of the last sentence.
3. Speech duration (sec): the sum of durations of each phrase (pauses excluded).
4. % Pause: the percentage of total reading time spent pausing.
5. Pause events (count): the number of times the participant paused while reading the passage.
6. Mean pause (sec): the average duration of all pauses.
7. Mean phrase (sec): represents the average duration of a phrase. Phrases were defined as sections of continuous speech between pauses.
8. Coefficient of variation of phrase durations (CV phrase duration): a normalized measure of variability of phrase durations.
9. Coefficient of variation of pause duration (CV pause duration): a normalized measure of variability in the duration of the pauses.

We expected that the SPA speech-based measures – which evaluated the integrity of speech articulation, including mean phrase and CV phrase duration – would be sensitive to bulbar motor impairment. The pausing measures, in contrast, might also reflect respiratory dysfunction, as well as bulbar impairment. Speaking rate, as a combined measure, would be affected by both deficits.

Statistical analysis

We used means and standard deviations, or counts and proportions, to describe variable distributions, as appropriate. To address question one, we assessed the test-retest reliability of the SPA measures in individuals from the control group who had at least two visits. For continuous measures, we used intra-class correlations coefficients (ICC), using a random effects model (ICC 2,1)(20). For the measure of speech intelligibility, the ICC was not meaningful because the range of scores was limited. In this case, we estimated total agreement and examined the Bland-Altman plot to assess limits of agreement (21). We also used the ICCs to calculate the standard error of measurement (SEM), which is the error around a single measurement, as follows: $SD \cdot \sqrt{1-ICC}$ (22). To estimate the error around repeated measures, we calculated the minimally detectable change (MDC); this is the smallest change in score that is outside of error. We calculated the MDC at the 95% confidence level as follows: $MDC_{95} = 1.96 \times \sqrt{2} \times SEM$ (22).

To address question two, we assessed the construct validity of the SPA measures by assessing their correlations with the total and bulbar ALSFRS-R scores in the patient group and we expected moderate correlations between SPA measures and ALSFRS-R scores. We also correlated these measures with %FVC as they may depend on the integrity of the respiratory musculature (14). Because many patients had repeated assessments, we used the rmcrr package in R (23) to account for multiple observations per individual. We also tested known-groups validity by comparing the mean scores in all SPA measures among controls, bulbar pre-symptomatic patients and bulbar symptomatic patients. To do so, we classified ALS patients as having bulbar symptoms based on the ALSFRS-R bulbar subscore, which has a maximum of 12 points. We considered patients as bulbar presymptomatic when they had scores of 11 or 12, and anyone below a score of 11 as symptomatic; this was done to account for error in measurement, as a difference between 11 and 12-point score can be due to error. We used ANOVA to compare the mean scores across the three groups (controls, pre and symptomatic patients), and Tukey test to compare specific pairs (i.e. controls vs. pre-symptomatic). We expected significant differences across groups.

To answer question three, we used the data from the control sample to estimate cut points for normality for each SPA measures. For each test, we calculated the mean +2SD as a cut point of normality. Then we then calculated the proportion of SPA measures outside of the cut points for normality in bulbar symptomatic and presymptomatic patients. The proportion of patients with an abnormal SPA test and symptomatic bulbar disease conceptually corresponds to the sensitivity of a given test to detect bulbar impairments, when considering the ALSFRS-R bulbar score as the gold standard. The bulbar presymptomatic patients with abnormal SPA tests represents patients with sub-clinical impairments of bulbar function. We compared the proportion of normal and abnormal tests in bulbar symptomatic and presymptomatic groups using chi-squared statistics. We also built receiver operating characteristic curves (ROC) to assess the performance of each SPA measure to differentiate symptomatic and presymptomatic patients. The area under the curve (AUC) reflects the performance of each measure, whereby AUC values close to 1 indicate good classification performance and values closer to 0.5 indicate poor performance (24).

Finally, we built a logistic regression model to predict symptomatic bulbar disease. For model building, we first included all SPA measures to assess multicollinearity using variation inflation factors (VIF), aiming for VIF values below 5 (24). We reduced variables first considering known correlations (i.e., removing a variable highly correlated with another), then removing SPA tests with poor performance (i.e., low reliability or poor differentiation between symptomatic and presymptomatic patients). We also tested potential non-linear effects with partial residual plots, and we built models using cubic splines for those variables flagged. We chose the final model using likelihood ratio (LR) tests and comparing C statistics and optimism after bootstrapping with 100 repetitions. The goal was to find the best performing model, while minimizing overfitting. We calibrated the final model using calibration plots.

All analyses were done with R software version 3.5.1 (25). We considered p values <0.05 as markers of statistical significance. When appropriate, we used Bonferroni correction for multiple testing.

Results

The final dataset had 1095 assessments, 775 for patients and 320 for controls. 620 patient assessments had the ALSFRS-R bulbar sub score; 36 assessments had missing SIT intelligibility data; and 41 had missing SIT speaking rate.

1. Test-retest reliability, SEM, and MDC₉₅

Reliability statistics as shown in Table 2 were calculated for the measures obtained from controls with at least 2 assessments (n=46). The data revealed moderate reliability (ICCs>0.5) for both measures of speaking rate, total and speech durations, and mean phrase durations. CV phrase and pause as well as mean pause duration had poor reliability (ICCs: 0.19-0.36).

2. Construct validity: Correlations with ALS measures

Table 3 reports correlations between SIT and SPA measures and clinical measures of disease, including ALSFRS-R and %FVC. The majority of bulbar measures were correlated with clinical scores. Moderate correlations ($r>0.4$) with ALSFRS-R total and bulbar subscores were observed for speaking rates obtained in the passage, total and speech duration, % pause and pause events. Neither SIT speaking rate or speech duration in a passage were correlated with %FVC.

3. Construct validity: Known-groups validity

ANOVA showed significant differences across the three groups (controls, symptomatic, and pre-symptomatic patients) in all the measures. Tukey's test showed significant differences between all pairs of comparisons (i.e. control vs. pre-symptomatic; pre-symptomatic vs. symptomatic), with the following exceptions: % intelligibility (non-significant between controls and pre-symptomatic), CV phrase duration (non-significant between symptomatic and pre-symptomatic patients), and CV pause duration (non-significant between pre-symptomatic and controls).

4. Performance of SPA tests to differentiate between bulbar symptomatic and presymptomatic patients

Of 620 assessment of ALS patients with complete ALSFRS-R data, 271 had an ALSFRS-R bulbar subscore <11 (symptomatic). Table 5 reports the proportion of abnormal SPA measures – using the cut-points derived from the control group – in the subgroups of symptomatic and presymptomatic bulbar patients, and the AUC obtained from ROC curves.

Total duration and speech duration were highly correlated to each other, and also to speaking rate, resulting in multicollinearity. The final logistic regression model had the following variables: passage speaking rate, pause events, and mean pause (VIF <2.5). Non-linear effects were not retained. The final model had AUC: 0.81, LR chi-square 196.97, $p < 0.0001$, Brier's = 0.17. Bootstrapping showed low optimism values (indicating minimal overfitting): 0.02 for Dxy, 0.03 for R2, and 0.07 for slope. Model coefficients are shown in Table 6.

Discussion

The study reported psychometric properties of speech and pause measures (SPA) (13) obtained in a passage reading task, as compared to SIT (26) measures, in neurotypical controls and patients with ALS with varying degrees of bulbar disease. Overall, the results demonstrated acceptable reliability, strong validity, and significant predicative abilities of SPA measures based on classification and regression analyses.

Reliability

The repeated measures collected during speaking tasks from the control participants demonstrated moderate reliability. This is not surprising as the study was not initially designed to study reliability, and the recordings were collected with a relatively long interval between the sessions (average of 138.42 days, $SD = 40.48$). The recordings were collected both in laboratory and in multiple clinics, and reliability of measurements may have been affected by the variations in the research assistants' styles and instructions. The extent of background noise in different recordings occasionally interfered with automatic extraction of SPA measures. Furthermore, the measures across control speakers had a relatively narrow range (e.g., speech durations ranged between 24-30 seconds); this directly affects the reliability coefficients that depend on the variance of the sample. The pause measures were somewhat less reliable than speech (e.g., speech duration, mean phrase) or speech + pause measures (e.g., speaking rate, total duration). This is not surprising as speakers tend to naturally vary the duration of their pauses more than their locations, which, in the neurologically healthy speakers, are determined by linguistic factors (e.g., syntactic boundaries) (27). A potential way to improve reliability coefficients in the future would be to record participants twice per session and use the average scores for all SPA measures, (20) and to have a shorter and standardized time-frame between assessments.

SIT reliability statistics have been reported in the past for participants with ALS and normal controls, showing much stronger results than the reliability of the SPA measures reported here for the first time (17). These differences might be in the way reliability statistic was calculated. Stipanovic et al. calculated reliability using a split-half method with only one

assessment per individual. Here, we used a more conservative approach, relying on data from two separate (test-retest) recording sessions. Arguably, for a test intended to measure change over time, test-retest reliability is more important, as we need to know the noise introduced by normal variations over time (22).

In addition to ICC, we also reported the SEM and MDC_{95} for SIT and SPA measures. If SEMs reflect the variation in one measurement (i.e., cross-sectional study) around the true scores, the MDC_{95} values are important as they help interpret change over time. For example, an individual should change by at least 7.45 seconds on the measure of total passage duration to consider the change to be outside of error, with 95% confidence. Our SIT intelligibility and speaking rate results compared well to those previously reported (17). Stipancic et al. reported the MDC_{95} SIT intelligibility of 3.5%; here we estimated it at 3.13%. The earlier paper reported MDC_{95} SIT speaking rate to be around 40 WPM for a group of speakers with high speaking rate (>160 WPM), similar to that reported in the current study. Stipancic et al. also showed a clear difference in MDC_{95} with disease progression, when the scores decreased with decline in speaking rate (17). These statistics for SPA measures had not been reported prior to the current study.

Construct Validity

All speech and pause measures distinguished the patient group from the control group, except coefficient of variation of pause duration (13,14). Yet, the majority of SPA measures had at best moderate correlations with ALSFRS-R total and bulbar scores. This is not surprising as ALSFRS-R scores – including the bulbar subscore – reflect a combination of symptoms resulting from dysfunction in different systems (e.g., speech, swallowing, salivation) and are not sensitive to early stages of bulbar decline, where we have the great majority of our data (7). Therefore, the variance explained by a single speech or pause measure is relatively low. However, the correlation statistics may be suitable for determining the most useful SPA measures. Specifically, mean pause, mean phrase, CV phrase and CV pause durations had particularly low or no correlation with clinical measures.

Sensitivity, specificity, and ROC AUC for the speech and pause measures

We were interested in determining if speech and pause measures would be able to detect bulbar changes in otherwise bulbar presymptomatic patients with ALS, using ALSFRS-R as the gold standard measure (6). The measures that combine both speech and pause estimates (e.g., speaking rate in the passage, total duration of the passage) performed best with AUCs of >0.8 – relatively good detection performance for such broad, easy to obtain measures. These passage-based measures performed comparably to the SIT speaking rates, indicating their interchangeability. Total speech duration, which in meaning is close to a measure of articulatory rate, was the next contender (AUC = 0.79). In our previous report (7), % pause stood out with an AUC >0.8 and measure of speaking rate had an AUC ~0.7, which was not confirmed in this study. The difference may be the result of a larger sample size and performance variation in the patient group in the current report.

We expanded our previous work by creating a multivariate predictive model of symptomatic bulbar disease. This model included speaking rate in the passage, pause events, and mean

pause duration. The model was well-calibrated and had a reasonably good discrimination ability (AUC = 0.81). Interestingly, it did not perform much better than the bivariate model of either passage speaking rate or total duration, and in fact, speaking rate was the main predictor of bulbar symptomatic status in the model. This is, in part, explained by the correlation between the different SPA measures.

Limitations and future directions

The key limitation of this work is that we were not able to assess reliability statistics as well as MDC of the speech and pause measures in the patient group. We only assessed these estimates in the controls, because the time frame between assessments in patients was long and we expected deterioration based on the natural history of the disease. From prior work on the SIT test we know that the estimates should vary with disease severity (17). Therefore, future work is needed to establish the reliability and MDC in patients with ALS at different points on the disease progression continuum, so that the SPA measures can be interpreted in this population. Another limitation of this work is in treating all of the data cross-sectionally. This choice was made due to the expansive nature of the analysis; the longitudinal responsiveness of the SPA measures will be reported separately. The differences in the time intervals between sessions across participants will be considered in this subsequent analysis. Additionally, it is important to note that cognitive impairments would affect reading and particularly pause measures (14). An evaluation of SPA across a sample of patients with ALS and varied cognitive abilities is required to delineate the contribution of cognitive abnormalities on a reading task.

Conclusions and recommendations

Based on the current psychometric evaluation, performing passage recording and speech and pause timing analysis was deemed useful for detecting changes associated with bulbar ALS. Among the SPA measures, speaking rate (i.e., the number of words produced per minute) as well as total duration and speech duration performed comparably and any of these can be used for bulbar disease tracking purposes. In comparison to the SIT measures, they are easy to use and require simply a timer to mark the onset and offset of the recording. It is, however, recommended that the patient is familiarized with the text of the passage prior to the recording to avoid false stops and starts, reading related hesitations, or error corrections.

Acknowledgements:

The funding for this work was provided by National Institutes of Health (NIH-NIDCD) grants R01DC009890, R01DC013547 and R01DC017291 as well as the Canadian Institutes for Health Research (CIHR), ALS Canada and Brain Canada for the Canadian ALS Neuroimaging Consortium (CALSNIC). Clinical data management and quality control was facilitated in part by the Canadian Neuromuscular Disease Registry.

Appendix

Bamboo Passage

Bamboo walls are getting to be very popular. They are strong, easy to use, and good-looking. They provide a good background and can create a look of a Japanese garden.

Bamboo is a grass, and is one of the most rapidly growing grasses in the world. Many varieties of bamboo are grown in Asia, although it is also grown in America. Last year we bought a new home and have been working on the flower garden. In a few more days, we will be done with the bamboo wall in our garden. We have really enjoyed the project.

References

1. Hecht M, Hillemacher T, Grasel E, Tigges S, Winterholler M, Heuss D, et al. Subjective experience and coping in ALS. *ALS Other Mot Neuron Disord.* 2002;3:225–32.
2. Benatar M, Boylan K, Jeromin A, Rutkove SB, Berry J, Atassi N, et al. ALS biomarkers for therapy development: State of the field and future directions. *Muscle and Nerve.* 2016;53(2):169–82. [PubMed: 26574709]
3. Yunusova Y, Plowman EK, Green JR, Barnett C, Bede P. Clinical measures of bulbar dysfunction in ALS. *Front Neurol.* 2019;10:1–11. [PubMed: 30761061]
4. Yunusova Y, Ansari J, Ramirez J, Shellikeri S, Stanisz GJ, Black SE, et al. Frontal anatomical correlates of cognitive and speech motor deficits in amyotrophic lateral sclerosis. *Behav Neurol.* 2019;2019:1–11.
5. 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. *J Neurol Sci.* 1999;169(1–2):13–21. [PubMed: 10540002]
6. Plowman EK, Tabor LC, Wymer J, Pattee G. The evaluation of bulbar dysfunction in amyotrophic lateral sclerosis: Survey of clinical practice patterns in the United States. *Amyotroph Lateral Scler Front Degener.* 2017;18(5–6):351–7.
7. Allison K, Yunusova Y, Campbell T, Wang J, Berry J, 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;18(5–6):358–66. [PubMed: 28355886]
8. Yorkston KM, Strand EA, Miller R, Hillel A, Smith K. Speech deterioration in amyotrophic lateral sclerosis: Implications for the timing of intervention. *J Med Speech Lang Pathol.* 1993;1:35–46.
9. Ball, Laura J and Willis, Amy and Beukelman, David R and Pattee GL. A protocol for identification of early bulbar signs in amyotrophic lateral sclerosis. *J Neurol Sci.* 2001;191:43–53. [PubMed: 11676991]
10. Yorkston KM, Beukelman D, Hakel MDM. Sentence Intelligibility Test. Lincoln, Nebraska,: Madonna Rehabilitation Hospital,; 2007.
11. 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. [PubMed: 23898888]
12. 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 10;20(6):669–79. [PubMed: 30409057]
13. Green JR, Beukelman DR, Ball LJ. Algorithmic estimation of pauses in extended speech samples of dysarthric and typical speech. *J Med Speech Lang Pathol.* 2004;12(4):149–54. [PubMed: 20628555]
14. Yunusova Y, Graham NL, Shellikeri S, Phuong K, Kulkarni M, Rochon E, et al. Profiling speech and pausing in amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). *PLoS One.* 2016;11(1):1–19.
15. Green JR, Allison KM, Cordella C, Richburg BD, Pattee GL, Berry JD, et al. Additional evidence for a therapeutic effect of dextromethorphan/quinidine on bulbar motor function in patients with amyotrophic lateral sclerosis: A quantitative speech analysis. *Br J Clin Pharmacol.* 2018;84:2849–56. [PubMed: 30152872]
16. Smith R, Pioro E, Myers K, Sirdofsky M, Goslin K, Meekins G, et al. Enhanced bulbar function in amyotrophic lateral sclerosis: The Nuedexta treatment trial. *Neurotherapeutics.* 2017;14:762–72. [PubMed: 28070747]

17. Stipancic KL, Yunusova Y, Berry JD, Green JR. Minimally detectable change and minimal clinically important difference of a decline in sentence intelligibility and speaking rate for individuals with amyotrophic lateral sclerosis. *J Speech, Lang Hear Res.* 2018;61(11):2757–71. [PubMed: 30383220]
18. Brooks BR, Miller RG, Swash M, Munsat TL. El Escorial revisited: Revised criteria for the diagnosis of amyotrophic lateral sclerosis. *Amyotroph Lateral Scler.* 2000;1(5):293–9.
19. Wang Y-T, Green JR, Nip ISB, Kent RD, Kent JF, and Ullman C. Accuracy of perceptually based and acoustically based inspiratory loci in reading. *Behavior Research Methods, Instruments & Computers,* 2010 43(3): p. 791–797.
20. Portney Leslie Gross MPW. *Foundations of Clinical Research.* Prentice Hall; 2009 892.
21. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;1(8476):307–10. [PubMed: 2868172]
22. Beaton DE, Bombardier C, Katz JN, Wright JG. A taxonomy for responsiveness. *J Clin Epidemiol.* 2001 12;54(12):1204–17. [PubMed: 11750189]
23. Bakdash JZ, Marusich LR. Repeated measures correlation. *Front Psychol.* 2017;8:1–13. [PubMed: 28197108]
24. Harrell F Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. *J Am Stat Assoc.* 2001;
25. Team RC. R: A language and environment for statistical computing. *R Found Stat Comput* 2013;
26. Yorkston K, Beukelman D, Hakel M. *Speech Intelligibility Test (SIT) for Windows [computer software].* Lincoln, Nebraska, USA: Madonna Rehabilitation Hospital; 2007.
27. Mitchell HL, Hoit JD, Watson PJ. Cognitive-linguistic demands and speech breathing. *J Speech, Lang Hear Res.* 1996;39(1):93–104.

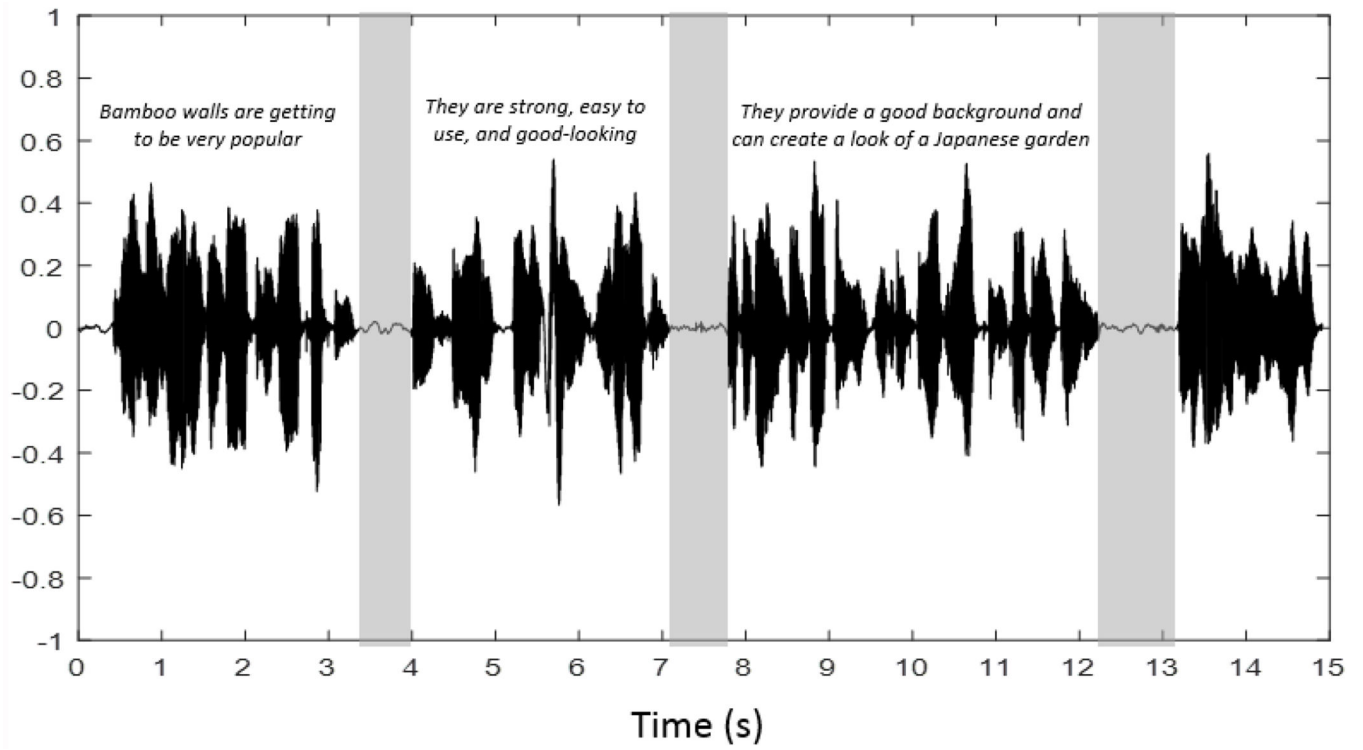


Figure 1.
The acoustic waveform recorded from a control participant with pauses marked in grey.

Table 1.

Demographic and clinical information of the dataset included in this study. Means of each group are provided with standard deviations in parentheses.

	ALS (n=272)	Controls (n=251)
Females : males	155:117	116:135
Age	58.29 (10.05)	50.89 (13.03)
Total # of Sessions	775	320
Median # of Sessions per subject	2.00 (3.00)	1.00 (1.00)
Onset, spinal : bulbar	228:44	-
Disease duration (months)	28.08 (24.17)	-
ASLFRS-R total	35.41 (7.36)	-
ALSFRS-R bulbar	10.46 (2.15)	-
% FVC	82.85 (22.15)	-
% Intelligibility, SIT	94.11 (14.01)	98.55 (1.22)
Speaking rate, SIT (WPM)	156.23 (41.07)	198.07 (21.80)

ALSFRS-R: Amyotrophic Lateral Sclerosis Functional Rating Scale-Revised (5); %FVC: % Forced vital capacity; SIT: Speech Intelligibility Test; WPM: Words per minute

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2.

Test-retest reliability and magnitude of measurement error of Speech Intelligibility Test (SIT) and speech and pause analysis (SPA) measures in a group of neurologically healthy controls.

	ICC	95% CI	Mean Difference	SEM	MDC ₉₅
% Intelligibility, SIT *	0.14	-0.15, 0.41	0.08	1.13	3.13
Speaking rate, SIT (WPM)	0.53	0.29, 0.71	1.98	14.9	41.30
Speaking rate, Passage (WPM)	0.56	0.33, 0.73	4.73	15.5	42.60
Total duration (sec)	0.58	0.35, 0.74	0.29	2.69	7.45
Speech duration (sec)	0.61	0.39, 0.76	0.70	1.93	5.35
% Pause	0.49	0.25, 0.68	1.06	3.39	9.39
Pause events (count)	0.47	0.22, 0.67	0.58	1.55	4.29
Mean phrase (sec)	0.59	0.36, 0.75	0.13	0.43	1.19
Mean pause (sec)	0.36	0.09, 0.59	0.13	0.14	0.30
CV phrase duration	0.19	-0.01, 0.44	0.08	0.14	0.39
CV pause duration	0.30	0.02, 0.54	0.03	0.12	0.28

* ICC paradox, given narrow range of observations. Test-retest agreement is 90% within 2 points

ICC: Intra-class correlation coefficient; CI: Confidence interval; SEM: Standard error of measurement calculated as follows = $SD \times \sqrt{1-ICC}$; MDC₉₅: Minimal detectable change (95% confidence) as follows = $1.96 \times \sqrt{2} \times SEM$ (22); WPM: Words per minute; CV: Coefficient of variation

Table 3.

Correlations between speech and pause measures and the clinical ALS measures of ALSFRS-R total and bulbar subscores and %FVC.

	ALSFRS-R Total	ALSFRS-R Bulbar	%FVC
% Intelligibility, SIT	0.29 (0.12, 0.39) *	0.34 (0.25, 0.43) *	0.13 (-0.01, 0.26)
Speaking rate, SIT (WPM)	0.39 (0.29, 0.47) *	0.34 (0.25, 0.43) *	0.20 (0.06, 0.33)
Speaking rate, Passage (WPM)	0.48 (0.40, 0.55) *	0.40 (0.31, 0.48) *	0.44 (0.33, 0.56) *
Total duration (sec)	-0.43 (-0.51, -0.35) *	-0.44 (-0.51, -0.35) *	-0.36 (-0.47, -0.24) *
Speech duration (sec)	-0.43 (-0.51, -0.35) *	-0.37 (-0.46, -0.28) *	-0.25 (-0.37, -0.12)
% Pause	-0.40 (-0.48, -0.31) *	-0.35 (-0.43, -0.25) *	-0.40 (-0.51, -0.28) *
Pause events (count)	-0.42 (-0.50, -0.33) *	-0.41 (-0.49, -0.32) *	-0.42 (-0.53, -0.30) *
Mean pause (sec)	-0.23 (-0.32, -0.13) *	-0.22 (-0.32, -0.13) *	-0.13 (-0.26, 0.01)
Mean phrase (sec)	0.32 (0.22, 0.40) *	0.21 (0.11, 0.31) *	0.37 (0.25, 0.49) *
CV phrase duration	-0.01 (-0.11, 0.10)	0.01 (-0.10, 0.11)	-0.01 (-0.15, 0.13)
CV pause duration	-0.02 (-0.12, 0.08)	-0.02 (-0.13, 0.08)	0.04 (-0.09, 0.18)

*
p<0.0001

95% confidence intervals are presented in parentheses.

ALSFRS-R: Amyotrophic Lateral Sclerosis Functional Rating Scale-Revised; %FVC: % Forced vital capacity; SIT: Speech Intelligibility Test; WPM: Words per minute; CV: Coefficient of variation

Table 4.

Differences between controls, pre-symptomatic and symptomatic patients for all speech and pause measures. Means for each group are presented with standard deviations in parentheses.

	Controls	ALS Pre-symptomatic	ALS Symptomatic	ANOVA <i>p</i> -value	2 SD cut-point
% Intelligibility, SIT	98.55 (1.22)	97.54 (3.33)	89.70 (20.58)	<0.0001	> 96.1
Speaking rate, SIT (WPM)	198.55 (21.80)	176.81 (33.26)	129.66 (38.91)	<0.0001	> 154.95
Speaking rate, Passage (WPM)	184.15 (23.50)	162.47 (28.78)	122.77 (35.11)	<0.0001	>137.15
Total duration (sec)	32.13 (4.16)	37.14 (7.78)	52.38 (18.28)	<0.0001	<40.45
Speech duration (sec)	26.95 (3.10)	30.22 (5.71)	40.81 (13.60)	<0.0001	<33.15
% Pause	15.81 (4.75)	18.09 (5.75)	21.75 (8.05)	<0.0001	<25.31
Pause events (count)	8.77 (2.14)	11.01 (4.09)	17.04 (9.86)	<0.0001	<13.05
Mean pause (sec)	0.58 (0.14)	0.62 (0.15)	0.70 (0.21)	<0.0001	<0.86
Mean phrase (sec)	2.89 (0.67)	2.75 (0.83)	2.55 (0.84)	<0.0001	< 4.23
CV phrase duration	0.46 (0.15)	0.52 (0.16)	0.53 (0.14)	<0.0001	<0.76
CV pause duration	0.33 (0.12)	0.36 (0.16)	0.41 (0.19)	<0.0001	<0.57

Bonferroni corrected *p* value=0.005

SD: Standard deviation; SIT: Speech Intelligibility Test; WPM: Words per minute; CV: Coefficient of variation

Table 5.

The proportion of abnormal measures in symptomatic and pre-symptomatic ALS patients. Number of participants in each group are provided with the percentage of the total sample in parentheses. AUCs are provided with 95% confidence intervals in parentheses.

	Bulbar Symptomatic (n=271)	Bulbar Presymptomatic (n=349)	p-value	ROC AUC
% Intelligibility, SIT	120 (45%)	80 (24%)	<0.001	0.62 (0.55, 0.63)
Speaking rate, SIT (WPM)	183 (70%)	86 (26%)	<0.001	0.82 (0.78, 0.85)
Speaking rate, Passage (WPM)	168 (62%)	62 (18%)	<0.001	0.81(0.77, 0.84)
Total duration (sec)	190 (70%)	85 (24%)	<0.001	0.81 (0.77, 0.84)
Speech duration (sec)	164 (61%)	67 (19%)	<0.001	0.79 (0.76, 0.83)
% Pause	76 (28%)	40 (12%)	<0.001	0.64 (0.60, 0.69)
Pause events (count)	155 (57%)	82 (24%)	<0.001	0.73 (0.69, 0.77)
Mean pause (sec)	44 (16%)	24 (7%)	<0.001	0.63 (0.58, 0.67)
Mean phrase (sec)	7 (3%)	20 (6%)	0.06	0.56 (0.52, 0.61)
CV phrase duration	16 (6%)	29 (8%)	0.27	0.53 (0.49, 0.58)
CV pause duration	43 (16%)	24 (7%)	0.0004	0.59 (0.54, 0.63)

p-value: for Chi-squared test comparing difference in the proportion of abnormal tests between bulbar symptomatic and presymptomatic patients.

Bonferroni corrected p-value=0.005.

ALSFRS-R: Amyotrophic Lateral Sclerosis Functional Rating Scale-Revised; ROC: Receiver Operating Characteristic curve; AUC: Area under the curve; SIT: Speech Intelligibility Test; WPM: Words per minute; CV: Coefficient of variation

Table 6.

Logistic regression model

Variable	Estimate	SE	P value
Intercept	4.665	1.132	<0.0001
Speaking rate, Passage (WPM)	-0.036	0.005	<0.0001
Pause events	0.008	0.024	0.75
Mean pause (sec)	0.341	0.582	0.56

Likelihood ratio chi square: 196.69, $p < 0.0001$. C: 0.81, R2: 0.37

SE: Standard error; WPM: Words per minute

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript