

RESEARCH

Open Access



Detecting bulbar amyotrophic lateral sclerosis (ALS) using automatic acoustic analysis

Leif E. R. Simmatís^{1,2,4*}, Jessica Robin³, Michael J. Spilka³ and Yana Yunusova^{1,2,4}

*Correspondence:
leif.simmatís@uhn.ca

¹ KITE-Toronto Rehabilitation
Institute, UHN, Toronto, ON,
Canada

² Department
of Speech-Language Pathology,
University of Toronto, Toronto,
ON, Canada

³ Winterlight Labs, Inc., Toronto,
ON, Canada

⁴ Sunnybrook Research Institute,
Toronto, ON, Canada

Abstract

Automatic speech assessments have the potential to dramatically improve ALS clinical practice and facilitate patient stratification for ALS clinical trials. Acoustic speech analysis has demonstrated the ability to capture a variety of relevant speech motor impairments, but implementation has been hindered by both the nature of lab-based assessments (requiring travel and time for patients) and also by the opacity of some acoustic feature analysis methods. These challenges and others have obscured the ability to distinguish different ALS disease stages/severities. Validation of automated acoustic analysis tools could enable detection of early signs of ALS, and these tools could be deployed to screen and monitor patients without requiring clinic visits. Here, we sought to determine whether acoustic features gathered using an automated assessment app could detect ALS as well as different levels of speech impairment severity resulting from ALS. Speech samples (readings of a standardized, 99-word passage) from 119 ALS patients with varying degrees of disease severity as well as 22 neurologically healthy participants were analyzed, and 53 acoustic features were extracted. Patients were stratified into early and late stages of disease (ALS-early/ALS-E and ALS-late/ALS-L) based on the ALS Functional Ratings Scale-Revised bulbar score (FRS-bulb) (median [interquartile range] of FRS-bulbar scores: 11 [3]). The data were analyzed using a sparse Bayesian logistic regression classifier. It was determined that the current relatively small set of acoustic features could distinguish between ALS and controls well (area under receiver-operating characteristic curve/AUROC = 0.85), that the ALS-E patients could be separated well from control participants (AUROC = 0.78), and that ALS-E and ALS-L patients could be reasonably separated (AUROC = 0.70). These results highlight the potential for automated acoustic analyses to detect and stratify ALS.

Introduction

Amyotrophic lateral sclerosis (ALS) is an incurable neurodegenerative disease that affects volitional motor control, visceral functions, and cognitive abilities. Survival with ALS, from disease onset, is estimated to be between 20 and 48 months [6]. Furthermore, ALS frequently causes speech impairment [37] secondary to bulbar motor system involvement. This can be devastating for patients and their families and has motivated substantial work to better understand patterns of bulbar/speech changes in people with ALS.



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Instrumental lab-based investigations of speech in ALS have demonstrated the value of speech assessment technologies for detecting and tracking ALS progression. The objective measurements afforded by the technologies provide information over and above that which can be gleaned by a clinician [32]. They can capture early signs of disease [26], be used to characterize ALS subgroups, including disease severity classifications [28], and distinguish patients from controls [39]. Detection of early signs of bulbar ALS is a substantial challenge that is very important to address for improving disease management [11]. However, lab-based systems tend to be complex and require trained personnel to operate them, even in the context of audio-only recordings. Furthermore, lab-based methods require dedicated lab space and for patients to visit a physical location outside the clinic, which requires time and effort for the patients. This creates barriers to data collection and precludes the incorporation of such tools into clinical practice or clinical trials, ultimately hindering technology adoption.

There has been great interest in developing remote, easy to use, and convenient speech assessment technologies for detection and tracking of ALS progression over time. Remote assessment systems have been developed by several groups in the recent years and have demonstrated great promise. For example, they have been used for distinguishing between ALS and control groups [21] and quantifying change over time in ALS acoustics [34]. They also have been well-tolerated by ALS patients [29]. Some recent work by Modality.AI has additionally utilized remote assessment for ALS detection as well as stratification of patients into bulbar and presymptomatic (i.e., lacking overt bulbar symptoms) patient groups [20]. However, their study focused on only a few features relating to pause timing and rate. There may be additional value in a more representative, but still compact, acoustic feature set that captures speech metrics from other domains such as voice quality. Collectively, an acoustic feature pipeline that can be utilized remotely could be of great value for stratifying patients for e.g., clinical trials or for more effective clinical decision making.

In the present study, we sought to validate an analytical pipeline developed by Winterlight Labs to detect signs of ALS from speech samples, as well as distinguish between severities of ALS-related speech impairments. Winterlight's remote assessment system has been used extensively for detecting cognitive–linguistic impairments associated with a variety of neurodegenerative and psychiatric diseases [2, 10, 14, 25], but not yet the speech motor impairment and not yet in ALS. The pipeline extracts a variety of acoustic features making it well suited for motor speech assessment in ALS. Here, we hypothesized that a core set of acoustic features derived from Winterlight's assessment pipeline could distinguish bulbar motor stages of the ALS (i.e., AUROC > 0.70) (1) ALS patients from control participants, (2) early ALS (ALS-E) patients from control participants, and (3) early ALS (ALS-E) from late ALS (ALS-L) patients. We additionally hypothesized that (4) weights given to individual features would be clinically interpretable in terms of relation to ALS and disease severity, and (5) that features influenced by sex (e.g., fundamental frequency measures) would not contribute substantially to modelling disease progression.

Results

Classification results

Classification results suggested that it was possible to separate ALS (median [interquartile range] FRS-bulbar scores: 11[3]) and control groups as well as ALS-E and ALS-L; AUROC was ≥ 0.70 for all comparisons. A plot of the 10 folds of ALS vs. control participants is shown in Fig. 1. We observed that the mean AUROC of the all-ALS vs control comparison was good (0.85), the AUROC of the ALS-E vs ALS-L comparison was somewhat lower (0.70), and that of the ALS-E vs. control comparison split the difference (0.78).

Feature coefficients

We identified across the ten train/test splits that certain groups of acoustic features tended to weight more strongly than others. See Fig. 2 for a summary of aggregated feature coefficients. It is evident that features from categories, such as speaking rate, intensity, F0 distributional characteristics (e.g., range), and shimmer tended to have higher feature weights, whereas ZCR, jitter, HNR, and pause statistics tended to have lower coefficient magnitudes. Some feature weights also reflected differences in disease severity. For example, in the ALS vs. control comparison, speech rate had a +0.36 coefficient, indicating that the speech rate in controls was higher than the ALS-E patients. In the ALS-E vs. ALS-L comparison, average word duration had a -0.63 coefficient, indicating that the ALS-E average word duration was lower than the ALS-L patients.

Impact of sex

We observed that the impact of sex as a covariate was not substantial in the majority of the ten trained and evaluated models. In all cases, the no-interaction model was either a better fit to the data, or the interaction did not fit the data substantially better. Thus, we retained the simpler model without interactions.

Discussion and conclusion

In this study, we validated an automated acoustic pipeline developed by Winterlight Labs for the purposes of stratifying ALS patients by bulbar disease severity. We observed that a relatively small set of the core acoustic features ($n=53$) derived from the automated analysis were able to detect ALS well (mean AUROC across ten test sets was 0.85) but, importantly, we were able to detect early signs of bulbar impairment at a comparable rate (mean AUROC = 0.78) and could even reasonably distinguish between ALS severities (mean AUROC = 0.70). Furthermore, acoustic features that are known to change with the disease severity in ALS (e.g., speech rate) [40] were given strong coefficients, validating the use of the pipeline for capturing speech changes in ALS. Finally, the models that included a sex-interaction term were not substantially better fits for the data than models without interaction terms. These results highlighted the substantial promise of the Winterlight system for the detection of bulbar motor changes overall as well as the detection of early bulbar changes in patients with ALS.

Additional research groups have explored the detection of ALS at various stages using acoustic features (sometimes in combination with kinematic features) and their

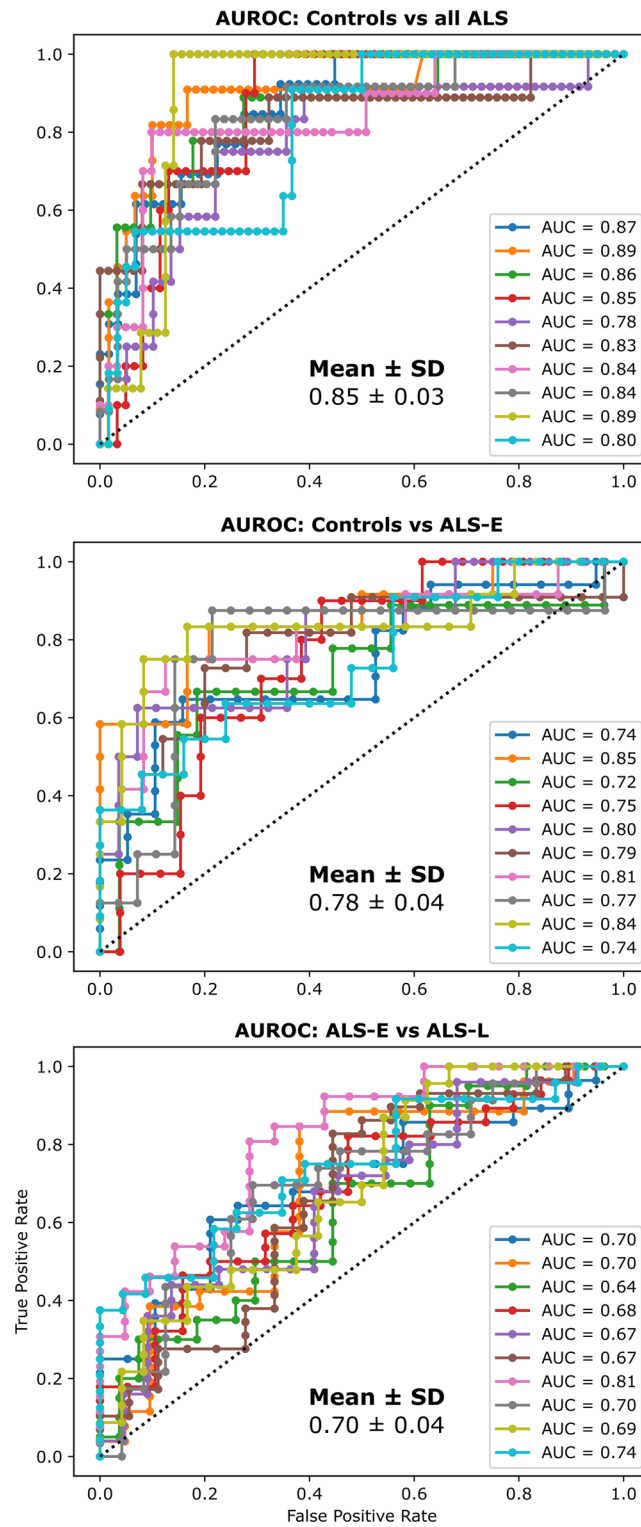


Fig. 1 Receiver-operating characteristic (ROC) plots for all 10 held-out test set folds of each comparison. AUC (i.e., AUROC) is indicated for each fold. Folds and legend labels for AUC are color-coded. Axes (true positive rate, false positive rate) are identical across all three plots. Finally, the summary statistics are depicted in each panel corresponding to mean \pm standard deviation (SD)

Fo max	0.3	0.4	-0.022
Fo mean	0.0095	-0.043	0.25
Fo median	-0.1	-0.086	0.18
Fo min	-0.046	-0.23	0.38
Fo range	0.29	0.45	-0.055
Fo var	0.39	0.34	0.28
Hnr ac max	0.11	-0.012	-0.043
Hnr ac mean	-0.19	-0.12	-0.16
Hnr ac median	-0.25	-0.16	-0.17
Hnr ac min	0.089	0.23	-0.036
Hnr ac range	0.088	-0.11	-0.015
Hnr ac var	-0.18	-0.13	-0.18
Hnr cc max	-0.096	-0.12	-0.074
Hnr cc mean	-0.13	-0.11	-0.067
Hnr cc median	-0.2	-0.13	-0.079
Hnr cc min	-0.11	-0.13	0.11
Hnr cc range	0.012	-0.11	-0.078
Hnr cc var	0.034	-0.045	0.075
Int max	-0.24	-0.36	-0.011
Int mean db	0.03	-0.007	-0.36
Int mean energy	-0.29	-0.39	0.17
Int mean sones	-0.32	-0.38	0.15
Int median	-0.41	-0.44	0.086
Int min	0.18	0.28	-0.29
Int range	-0.19	-0.32	0.29
Int var	-0.5	-0.37	0.47
Jitter ddp	-0.16	-0.18	-0.17
Jitter local	0.01	-0.11	0.075
Jitter local absolute	-0.001	0.008	-0.0045
Jitter ppq5	-0.28	-0.25	-0.08
Jitter rap	-0.14	-0.18	-0.2
Long pause ct norm	-0.087	0.051	-0.18
Long pause dur	-0.08	0.074	-0.17
Mean pause dur	0.04	0.16	-0.12
Med pause ct norm	0.0065	0.0075	0.017
Med pause dur	-0.11	0.076	-0.11
Pause word ratio	0.2	0.3	0.28
Phonation rate	0.072	-0.081	0.12
Shimmer apq11	-0.21	-0.33	0.21
Shimmer apq3	-0.32	-0.23	-0.13
Shimmer apq5	-0.26	-0.31	-0.077
Shimmer dda	-0.31	-0.23	-0.13
Shimmer local	-0.08	-0.18	0.059
Shimmer local db	0.13	-0.009	0.22
Short pause ct norm	0	-0.05	0.13
Short pause dur	-0.19	-0.032	0.12
Zcr kurtosis	-0.24	-0.13	-0.22
Zcr mean	-0.2	-0.17	0.0055
Zcr skewness	-0.14	-0.075	-0.22
Zcr var	-0.17	-0.15	0.029
Articulation rate	0.28	0.17	0.41
Avg word dur	-0.23	-0.23	-0.63
Speech rate	0.36	0.22	0.38
	ALS Control	ALS-E Control	ALS-L ALS-E

Fig. 2 Aggregated feature coefficients across the three different binary classification conditions. In general, positive values correspond to the less-impaired group having higher values. Stronger weights are indicated by either darker orange (positive weights) or darker blue (negative weights) cell colours, and weights are additionally annotated. Feature names are indicated on the vertical axis. *ac* autocorrelation, *apq* amplitude perturbation quotient (3, 5, and 11-point cases), *cc* cross-correlation, *ct* count, *db* decibels, *dda* average absolute difference in amplitudes between periods, *ddp* difference of difference of periods, *dur* duration, *Fo* fundamental frequency, *HNR* harmonic/noise ratio, *int* intensity, *max* maximum, *med* medium, *min* minimum, *norm* normalized, *ppq5* 5-point pitch perturbation quotient, *rap* relative average perturbation, *var* variance, *Zcr* zero-crossing rate

classifiers' performance was generally in line with that observed here. Modality.AI [20] used a multimodal dialogue agent to assist in the extraction of acoustic and kinematic speech features. They additionally stratified patients into bulbar symptomatic

and presymptomatic groups. Their AUC performance was comparable to that of the present study; severe patients vs control mean AUC was 0.92, followed by a mean AUC of 0.81 for bulbar vs presymptomatic, and a mean AUC of 0.62 for controls vs presymptomatic patients. Our results by comparison were 0.85 (note: all patients rather than only severe patients), 0.70, and 0.78 for the corresponding comparisons. The difference in performance between the (A) less-severe vs more-severe comparison, and the (B) less severe vs control comparison may reflect differences in stratification cutoff. Neumann et al. did not allow an FRS-Bulb score of 11/12 to count as “early” disease, whereas we did. Additional differences included our preponderance of ALS patients vs. controls (119 patients vs 22 controls) as compared to the inverse in Neumann’s work (29 patients vs 68 controls). Higher performance of voice-based classification has been reported [35, 39], but these either did not stratify patients into groups, or included mel-frequency cepstral coefficients, that we did not include in the present work because of the difficulty in their clinical interpretation.

Salient patterns were observed in the features that were given strong weights in the classification results. For example, rate-related features typically had relatively high coefficient values across all three of the binary comparisons. However, they were much stronger in ALS-L vs. ALS-E compared to, e.g., ALS-E vs. control. This reflects the greater rate of decline in speaking rate with more advanced disease [9], although it is notable that Allison et al. [1] identified rate/pause related features as important for early detection of bulbar symptoms as well, this may reflect differences in the dataset or in the determination of “early” ALS (they used a self-report threshold of < 12 on FRS-Bulb, which differs from our present criteria of $\leq 11/12$). Other measures of articulation timing and control such as voice onset time have been shown to differ between early and late stages of ALS as well [36]. Additional features from phonatory and respiratory categories may show differential effects of disease severity that could correspond to the findings from the present study. For example, previous work has identified that maximum F0 and F0 range are important features for predicting intelligibility loss [18, 27], and phonatory instability is known to increase in advanced ALS [23]. In terms of respiratory features, the previous work has identified that impairment of respiratory muscles (in particular expiratory muscles) occurs rapidly in ALS, which may correspond to the current observation of a strong weight applied to the intensity features (e.g., median intensity) [16]. Finally, it is notable that many of the features in our models, including the ones aggregated across multiple test-set repetitions, tended to be close to 0. For instance, across all three of the binary classifications, HNR features tended to have low-magnitude coefficient values, suggesting that they were not important for any of the classifications. This is likely a consequence of our choice of regularization approach, which makes interpretation of the patterns across groups more straightforward.

Some of the features that we would have most expected to be affected by sex typically had low feature weights. This was particularly the case in the F0 mean and F0 median features, which had low feature weights in the all-ALS vs. control and ALS-E vs. control comparisons. This observation supports our choice to not model interactions between sex and acoustic features in the present analysis. We acknowledge that at later stages of disease severity, there can be differential patterns of F0 change between males and females with males demonstrating higher F0 and females—lower F0 with disease

progression [17, 24]. This could explain their lower performance for distinguishing ALS severity groups. Notably, the weights for F0 features increased slightly in the ALS-E vs. ALS-L comparisons, which is unlikely to be due to a sex imbalance given that in the ALS-L group, 56% of participants were males, as compared to 63% in the ALS-E group. Potentially this is due to changes in the F0 that occur with disease progression [23].

There are many important clinical extensions to the present study that can complement our present efforts to distinguish ALS and control groups and different severities of ALS. There is rapidly growing literature highlighting the potential for speech acoustics to be used for disease differentiation and prognostication. For example, Milella et al. [19] identified that vowel space area from sustained vowel tasks, and alternating/sequential motion rates from diadochokinetic tasks, could capture differences between upper vs. lower motor neuron endotypes. Furthermore, acoustic analysis has the potential to track [34] and prognosticate [33] change over time in ALS-related speech impairments. These studies collectively highlight the value of speech analysis as a putative digital biomarker in ALS. These, along with possibly complementary modalities, such as magnetic resonance imaging (MRI) [3] may be useful for survival prediction in ALS [13, 30]. Given that speech networks in the brain can be altered in ALS [31], this or other data could be fruitful for developing a multimodal understanding of disease progression in ALS.

Our study has some limitations to be addressed in future research. An important consideration is that, of the >750 features in the Winterlight pipeline, only the 53 here pertain to acoustics, and are drawn from a relatively small number of feature domains (e.g., many related to rate and pauses). This limits the ability to represent impairments in diverse domains of speech such as the resonatory subsystem, which can be affected in ALS [8]. It is also notable that the balance of feature types in the Winterlight feature set was biased substantially towards articulatory features. Many of these features were given relatively high feature weights, but they likely captured the same overall constructs, despite measuring e.g., pauses of different durations. Potentially, the methods to address collinearity could be of benefit to this analysis in future studies. We could additionally explore methods for accounting for more covariates, such as age and education level (in addition to sex as adjusted for here), which might provide better generalization in larger real-world datasets. Classification analysis demonstrated that in cases where the number of features exceeds the number of observations, it is not possible to select more predictors than observations using a Laplace prior [38]. In practice, we did not expect to have a large number of important predictors, and empirically we observed good performance of the LASSO (i.e., Laplace) as implemented here. However, this is taken under advisory for future work to explore different coefficient shrinkage methods, particularly in cases where there may be many more than 53 acoustic features to analyze. We also had a relatively imbalanced dataset between ALS and control participants; we addressed this using robust analysis and scoring methods (Bayesian methods with AUROC evaluation) that is resistant to influence by class imbalances. However, a larger control dataset might enable a more detailed appreciation of acoustic patterns associated with healthy performance and enable comparisons between ALS and other neurodegenerative diseases as well. Furthermore, we empirically demonstrated that AUROCs were relatively consistent across the held-out test sets, suggesting that, although small, our control cohort was at least modestly dispersed. A larger control group may be able to better capture

a wider range of normative speech behaviors, which could in turn enable more granular description of ALS-related speech impairments at various stages of diseases. Finally, we explored comparisons between ALS severities and controls or each other in a binary fashion; this led to interesting results and highlighted some interesting patterns in the data; however, a more nuanced comparison would be to perform a three-way classification. We did not perform this analysis in the present study because we had a relatively mild cohort overall. Future work should explore multiclass classification after recruiting ALS patients with a wider range of ALS-related speech impairments, which may also enable a more granular commentary on changes in ALS speech function over time/ across severity levels.

The results of the present study suggest that automated acoustic analysis using a pipeline developed by Winterlight Labs can detect bulbar ALS, as well as earlier stages of the disease. These results show that even with a relatively small set of acoustic features, the Winterlight pipeline could stratify ALS patients into early and late bulbar stages, with clinically-interpretable feature importances. Future work will evaluate the detection using more participants and across a greater range of severities.

Methods

The data were collected from 141 participants (119 ALS, 22 controls). See Table 1 for a summary of relevant clinical and demographic features of the cohorts. Informed consent prior to participation was collected in accordance with the Declaration of Helsinki. Inclusion criteria were fluency in English and a diagnosis of ALS by an experienced neurologist. Exclusion criteria were the presence of any other neurological disorders (e.g., stroke), and Montreal Cognitive Assessment (MoCA) score < 26/30, indicative of a potential cognitive impairment, and the inability to read the passage fluently (e.g., due to dyslexia or impaired vision). For patients only, the ALS Functional Rating Scale-Revised bulbar scale (FRS-bulb) was used to stratify them into “early” and “late” bulbar groups using the median value in the dataset, which was 11 out of 12 maximum; i.e., < 11/12 is ALS-L and ≥ 11/12 is ALS-E. Owing to the missing data for the FRS, *n* = 93 individuals were analyzed when comparing ALS-E vs ALS-L, and *n* = 70 were analyzed when comparing control vs ALS-E. Participants read the Bamboo Passage, which is 99 words in length and assesses various aspects of articulatory and respiratory motor function [40]. The data were recorded in a speech laboratory embedded into a multidisciplinary ALS clinic. The recordings were conducted using a high-quality digital recorder at 44.1 kHz in 16-bit resolution using a cardioid lavalier microphone.

Table 1 Summary of demographic and clinical information

	ALS	Control
Total sample size (Females)	119 (47)	22 (12)
Age (median [IQR])	59 [14]	52 [8]
FRS-bulbar (median [IQR], score < 9/12)	11 [3], 13	–
Disease duration (years)	2.9 [3.3]	–

All values are formatted as median ± interquartile range

We preprocessed raw acoustic data by removing noise prior to downstream analyses, using Praat [4]. At least 0.25 s of audio data (i.e., ~10,000 samples) was used for the spectral subtraction noise reduction algorithm [5], with a window length of 0.025 s, which follows the recommendations on noise reduction in Praat. We selected sample length that was at least several times the length of the window (https://www.fon.hum.uva.nl/praat/manual/Sound_Remove_noise_.html; accessed 7 June 2023). Other settings for noise reduction included suppression range of 80 Hz to 10 kHz, and 40 Hz smoothing. The choice to use lab-based data followed from the purpose of the present study, which was to validate the Winterlight assessment pipeline for both ALS detection and ALS stratification, when the data are known to be of high quality and recorded was done under controlled conditions.

Further semi-automated quality analysis after noise suppression was performed to ensure high-quality data were analyzed. Thresholds were signal to noise ratio (SNR) > 30Db [7], clipping in fewer than 1% of data samples [15], and no unusual patterns of noise as evident by visual inspection of spectrograms (e.g., narrowband noise). These steps were performed by trained and experienced research assistants. The data with clipping or SNRs exceeding these thresholds were discarded and not analyzed further. This amounted to approximately 3 samples out of the initial set of 122 recordings, yielding the final sample size of 119.

Winterlight's automated pipeline extracts 793 features that encompass various domains of speech and language functioning. For the purposes of the present study, we chose to focus specifically on acoustic features, which were expected to reflect the *motor* speech impairment that occurs in ALS. We left the investigation of linguistic features to future work in patients who might have more pronounced cognitive deficits and those on the ALS—frontotemporal dementia (FTD) spectrum. Specifically, we focused on a total of 53 acoustic features that reflected the integrity of the respiratory, phonatory, and articulatory physiologic speech subsystems ([12]). Briefly, these features include, but are not limited to, a variety of speech/pause durations and rates (articulation and respiration), jitter/shimmer/harmonic measures (phonation), as well as additional metrics such as zero-crossing rate. See Additional file 1: Table S1 for a description of these features in detail. Briefly, feature categories included: jitter/shimmer, fundamental frequency (F0), speech/pause durations, zero-crossings, harmonic/noise ratio (HNR), and intensity.

Classification was performed using a Bayesian LASSO (i.e., the Least Absolute Shrinkage and Selection Operator) logistic regression model. See Fig. 3 for a schematic diagram of the present statistical model. Following from classical logistic regression, which is a linear operation transformed using a log link function, the present model consists of a global intercept α (i.e., between the two classes being compared at any given time) and a vector of $k \in \{1 \dots 53\}$ β coefficients (i.e., one per acoustic feature). The α parameter was drawn from a standard Normal $N(0,1)$ distribution, whereas the β_k were drawn from a Laplace $L(0.5)$ distribution, where 0.5 is the parameter controlling the width of the distribution. The latter decision was made to impose a LASSO penalty on the β_k , which is a technique for making coefficients sparse by imposing a penalty on high coefficient values. The Laplace distribution implements this in a Bayesian context [22]. Briefly, the Laplace distribution has a sharper peak

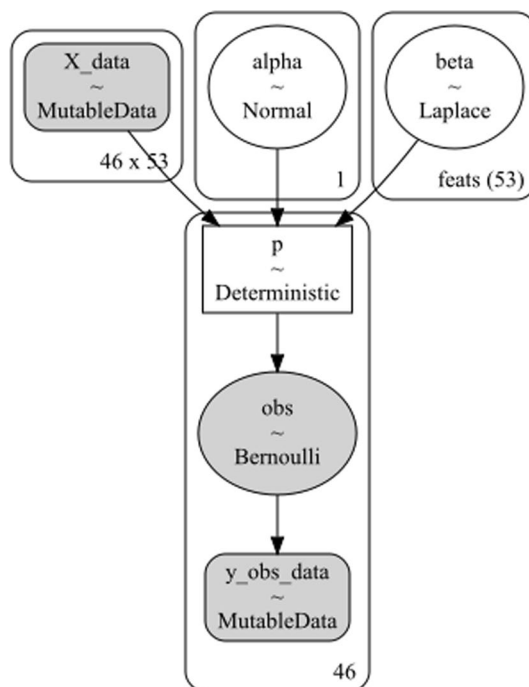


Fig. 3 The statistical model used in the present project. Image generated using PyMC v 5.4.1. Alpha is the offset parameter, beta is the vector of $J=53$ regression coefficients (one per feature), and the posterior was modelled as a Bernoulli distribution. X and y are data objects. Numerical indices reflect the number of individuals. In the figure, the number of participants is 46 (i.e., 50% of the total data in the ALS-E vs ALS-L comparison)

than a Gaussian distribution and so would be hypothesized to penalize coefficients with low values and compress them towards 0, without proportionally impacting higher coefficients.

As an empirical example of the parameter shrinkage induced by the LASSO penalty, see Fig. 4, which depicts a histogram of parameter values from one of the training folds in the present study, fitted using a Laplace distribution and a Normal distribution. It is evident that the Laplace prior forces parameters to cluster around 0, although retains a number of non-zero parameters with moderate to strong magnitudes (i.e., $\sim|0.5|$). Importantly, this enabled us to more definitively comment on features that had strong impacts on classification decisions, by forcing those with low relative contributions closer to zero.

Binary classifications were performed between: (1) control vs all ALS, (2) control vs ALS-E, and (3) ALS-E vs ALS-L. We performed ten randomized dataset splits (i.e., tenfold cross-validation), where training data (50%) and testing data (50%) were fully separated. Train and test splits were performed 10 times per comparison condition, with splits being performed pseudorandomly at each iteration. AUROC values were aggregated across the ten held-out test sets. Note that a further split of training into training/validation was not performed, because of the underlying mechanics of the Bayesian model fitting process (there is no hyperparameter tuning as in, e.g., a support vector machine, and so a grid search of hyperparameters is not needed). At each testing iteration, AUROC was evaluated using the predicted score and the ground

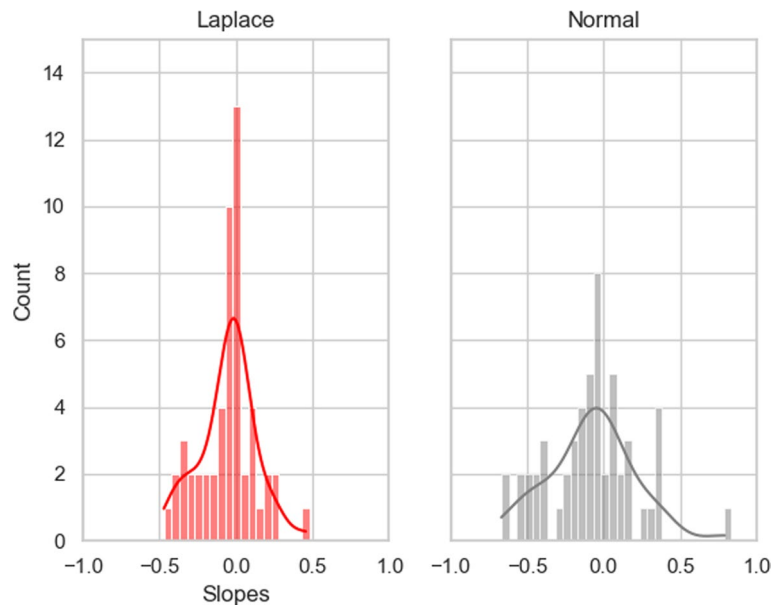


Fig. 4 Histograms and kernel density estimates (i.e., the smooth lines) demonstrating the shrinkage of regression coefficients induced by using a Laplace $L(0.5)$ prior instead of a Normal $N(0,1)$ prior. The data are from one classification fold. Note that a large number of slopes in the Laplace plot are compressed towards zero (indicated by the higher central spike in the left subplot than the right one)

truth labels. Note that train and test splits were standardized using the mean and variance of the training data.

In addition to the binary classification, we investigated the potential contribution of sex as an interactor variable in specific acoustic features, where it would be expected to play a role, given typical differences in vocal physiology between individuals born male and those born female. Specifically, sex effects were modelled in fundamental frequency and HNR features. Interactions were encoded at the data level as multiplicative interactions, and interaction vs no-interaction models were compared using the Watanabe-Akaike information criterion (WAIC).

Finally, the learned β_k for each binary comparison and for each classification fold were extracted. The median of these values was calculated for plotting purposes, to provide an indication of the relative contribution of each acoustic feature to each classification decision.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12938-023-01174-z>.

Additional file 1: Table S1. Acoustic features and their descriptions (Winterlight pipeline).

Acknowledgements

We would like to sincerely thank Dr. Madhura Kulkarni, Justin Truong, Rupinder Sran, Scotia McKinlay, and Niyousha Taati for their roles in data management and preprocessing to support this project.

Author contributions

LS conceptualized the study, performed data analyses, and prepared the manuscript. JR supported the conceptualization of the study and assisted in manuscript preparation. MS assisted in manuscript preparation. YY supported original data collection, supported the conceptualization of the study, and guided manuscript preparation.

Funding

YY is funded by an NIH R01 grant (R01DC013547), an ALS Society of Canada Project Grant, and an NSERC Discovery Grant. LS is supported by a Mitacs Elevate Postdoctoral Fellowship.

Availability of data and materials

The Research Ethics Board and Legal Services at Sunnybrook Health Sciences Centre prohibit access to data without a data sharing agreement between the authors and the interested investigator(s). Please contact Dr. Yana Yunusova to set up an agreement to access the data used in this paper.

Declarations

Ethics approval and consent to participate

This study was approved by the Sunnybrook Health Sciences Centre Research Ethics Board, approval numbers 207-2007 and 2080.

Competing interests

JR and MS are employees of Winterlight Labs, Inc. Winterlight did not influence the analytical choices made here, nor the decision to publish. LS is funded by a Mitacs Elevate fellowship in collaboration with Winterlight Labs and receives payment as an intern but is not an employee of Winterlight Labs. YY has no competing interests to disclose.

Received: 29 August 2023 Accepted: 19 November 2023

Published online: 04 February 2024

References

1. 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;18(5–6):358–66. <https://doi.org/10.1080/21678421.2017.1303515>.
2. Balagopalan A, Kaufman L, Novikova J, Siddiqui O, Paul R, Ward M, Simpson W. Early development of a unified, speech and language composite to assess clinical severity of frontotemporal lobar degeneration (FTLD). *Clin Trials Alzheimer's Dis.* 2019. <https://doi.org/10.14283/jpad.2019.48>.
3. Bede P, Murad A, Hardiman O. Pathological neural networks and artificial neural networks in ALS: diagnostic classification based on pathognomonic neuroimaging features. *J Neurol.* 2022;269(5):2440–52. <https://doi.org/10.1007/s00415-021-10801-5>.
4. Boersma P, Weenink D. Praat: doing phonetics by computer [Computer program]. Version 6.1.50; 2021.
5. Boll SF. Suppression of acoustic noise in speech using spectral subtraction. *IEEE Trans Acoust Speech Signal Process.* 1979;27(2):113–20. <https://doi.org/10.1109/TASSP.1979.1163209>.
6. Chio A, Logroscino G, Hardiman O, Swingler R, Mitchell D, Beghi E, Traynor BG. Prognostic factors in ALS: a critical review. *Amyotroph Lateral Scler.* 2009;10(5–6):310–23. <https://doi.org/10.3109/17482960802566824>.
7. Deliyski DD, Shaw HS, Evans MK. Adverse effects of environmental noise on acoustic voice quality measurements. *J Voice.* 2005;19(1):15–28. <https://doi.org/10.1016/j.jvoice.2004.07.003>.
8. Eshghi M, Connaghan KP, Gutz SE, Berry JD, Yunusova Y, Green JR. Co-occurrence of hypernasality and voice impairment in amyotrophic lateral sclerosis: Acoustic quantification. *J Speech Lang Hear Res.* 2021;64(12):4772–83. https://doi.org/10.1044/2021_JSLHR-21-00123.
9. Eshghi M, Yunusova Y, Connaghan KP, Perry BJ, Maffei MF, Berry JD, Zinman L, Kalra S, Korngut L, Genge A, Dionne A, Green JR. Rate of speech decline in individuals with amyotrophic lateral sclerosis. *Sci Rep.* 2022;12(1):1–13. <https://doi.org/10.1038/s41598-022-19651-1>.
10. Fraser KC, Meltzer JA, Rudzicz F. Linguistic features identify Alzheimer's disease in narrative speech. *J Alzheimer's Dis.* 2015;49(2):407–22. <https://doi.org/10.3233/JAD-150520>.
11. Goutman SA, Hardiman O, Al-chalabi A, Chió A, Savelieff MG, Kiernan MC, Feldman EL. Emerging insights into the complex genetics and pathophysiology of amyotrophic lateral sclerosis. *Lancet Neurol.* 2022. [https://doi.org/10.1016/S1474-4422\(21\)00414-2](https://doi.org/10.1016/S1474-4422(21)00414-2).
12. Green JR, Yunusova Y, Kuruvilla MS, Wang J, Pattee GL, Synhorst L, Zinman L, Berry JD. Bulbar and speech motor assessment in ALS: challenges and future directions. *Amyotroph Lateral Scler Frontotemporal Degen.* 2013;14(7–8):494–500. <https://doi.org/10.3109/21678421.2013.817585>.
13. Grollemund V, Le Chat G, Secchi-Buhour MS, Delbot F, Pradat-Peyre JF, Bede P, Pradat PF. Manifold learning for amyotrophic lateral sclerosis functional loss assessment: development and validation of a prognosis model. *J Neurol.* 2021;268(3):825–50. <https://doi.org/10.1007/s00415-020-10181-2>.
14. Gumus M, DeSouza DD, Xu M, Fidalgo C, Simpson W, Robin J. Evaluating the utility of daily speech assessments for monitoring depression symptoms. *Digital Health.* 2023. <https://doi.org/10.1177/20552076231180523>.
15. Hansen JHL, Stauffer A, Xia W. Nonlinear waveform distortion: assessment and detection of clipping on speech data and systems. *Speech Commun.* 2021;134:20–31. <https://doi.org/10.1016/j.specom.2021.07.007>.
16. Heiman-Patterson TD, Khazaal O, Yu D, Sherman ME, Kasarskis EJ, Jackson CE, Heiman-Patterson T, Sherman MS, Mitchell M, Sattazahn R, Feldman S, Scelsa SN, Imperato T, Shefner JM, Lou Watson M, Rollins Y, Cumming J, Newman D, Foley H, et al. Pulmonary function decline in amyotrophic lateral sclerosis. *Amyotroph Lateral Scler Frontotemporal Degen.* 2021;22(S1):54–61. <https://doi.org/10.1080/21678421.2021.1910713>.
17. Kent JF, Kent RD, Rosenbek JC, Weismer G, Martin R, Suftit R, Brooks BR. Quantitative description of the dysarthria in women with amyotrophic lateral sclerosis. *J Speech Hear Res.* 1992;35(4):723–33. <https://doi.org/10.1044/jshr.3504.723>.

18. Kent RD, Weismer G, Kent JF, Rosenbek JC. Toward phonetic intelligibility testing in dysarthria. *J Speech Hearing Disord.* 1989;54(4):482–99. <https://doi.org/10.1044/jshd.5404.482>.
19. Milella G, Sciancalepore D, Cavallaro G, Piccirilli G, Nanni AG, Fraddosio A, Errico ED, Paolicelli D, Fiorella ML, Simone IL. Acoustic voice analysis as a useful tool to discriminate different ALS phenotypes. *Biomedicines.* 2023;11:2439.
20. Neumann M, Roesler O, Liscombe J, Kothare H, Suendermann-Oeft D, Pautler D, Navar I, Anvar A, Kumm J, Norel R, Fraenkel E, Sherman AV, Berry JD, Pattee GL, Wang J, Green JR, Ramanarayanan V. Investigating the utility of multimodal conversational technology and audiovisual analytic measures for the assessment and monitoring of amyotrophic lateral sclerosis at scale. 2021; <http://arxiv.org/abs/2104.07310>.
21. Norel R, Pietrowicz M, Agurto C, Rishoni S, Cecchi G. Detection of amyotrophic lateral sclerosis (ALS) via acoustic analysis. In: Proceedings of the Annual Conference of the International Speech Communication Association, INTER-SPEECH, 2018-Sept, 2018; pp. 377–381. <https://doi.org/10.21437/Interspeech.2018-2389>.
22. Park T, Casella G. The Bayesian Lasso. *J Am Stat Assoc.* 2008;103(482):681–6. <https://doi.org/10.1198/01621450800000337>.
23. Ramig LO, Scherer RC, Klasner ER, Titze IR, Horri Y. Acoustic analysis of voice in amyotrophic lateral sclerosis: a longitudinal case study. *J Speech Hearing Disord.* 1990;55(1):2–14. <https://doi.org/10.1044/jshd.5501.02>.
24. Robert D, Pouget J, Giovanni A, Azulay JP, Triglia JM. Quantitative voice analysis in the assessment of bulbar involvement in amyotrophic lateral sclerosis. *Acta Otolaryngol.* 1999;119(6):724–31. <https://doi.org/10.1080/00016489950180702>.
25. Robin J, Xu M, Kaufman LD, Simpson W. Using digital speech assessments to detect early signs of cognitive impairment. *Front Digital Health.* 2021. <https://doi.org/10.3389/fgdth.2021.749758>.
26. Rong P, Yunusova Y, Wang J, Green JR. Predicting early bulbar decline in amyotrophic lateral sclerosis: a speech subsystem approach. *Behav Neurol.* 2015. <https://doi.org/10.1155/2015/183027>.
27. Rong P, Yunusova Y, Wang J, Zinman L, Pattee GL, Berry JD, Perry B, Green JR. Predicting speech intelligibility decline in amyotrophic lateral sclerosis based on the deterioration of individual speech subsystems. *PLoS ONE.* 2016;11(5):e0154971. <https://doi.org/10.1371/journal.pone.0154971>.
28. Rowe HP, Gutz SE, Maffei MF, Green JR. Acoustic-based articulatory phenotypes of amyotrophic lateral sclerosis and Parkinson's disease: towards an interpretable, hypothesis-driven framework of motor control. In: Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 2020; p. 4816–4820. <https://doi.org/10.21437/Interspeech.2020-1459>.
29. Rutkove SB, Narayanaswami P, Berisha V, Liss J, Hahn S, Shelton K, Qi K, Pandeya S, Shefner JM. Improved ALS clinical trials through frequent at-home self-assessment: a proof of concept study. *Ann Clin Transl Neurol.* 2020;7(7):1148–57. <https://doi.org/10.1002/acn3.51096>.
30. Schuster C, Hardiman O, Bede P. Survival prediction in Amyotrophic lateral sclerosis based on MRI measures and clinical characteristics. *BMC Neurol.* 2017;17(1):1–10. <https://doi.org/10.1186/s12883-017-0854-x>.
31. Shellikeri S, Myers M, Black S, Abrahao A, Zinman L, Yunusova Y. speech processing network regional involvement in bulbar ALS: a multimodal structural MRI study. *Amyotroph Lateral Scler Frontotemporal Degener.* 2019;20:385–95. <https://doi.org/10.1080/21678421.2019.1612920>. *Speech*.
32. Silbergleit AK, Johnson AF, Jacobson BH. Acoustic analysis of voice in individuals with amyotrophic lateral sclerosis and perceptually normal vocal quality. *J Voice.* 1997;11(2):222.
33. Stegmann G, Charles S, Liss J, Shefner J, Berisha V. Degeneration A speech-based prognostic model for dysarthria progression in ALS. *Amyotroph Lateral Scler Frontotemp Degener.* 2023. <https://doi.org/10.1080/21678421.2023.2222144>.
34. Stegmann GM, Hahn S, Liss J, Shefner J, Rutkove S, Shelton K, Duncan CJ, Berisha V. Early detection and tracking of bulbar changes in ALS via frequent and remote speech analysis. *NPJ Dig Med.* 2020;3(1):1–5. <https://doi.org/10.1038/s41746-020-00335-x>.
35. Tena A, Clarià F, Solsona F, Povedano M. Detecting bulbar involvement in patients with amyotrophic lateral sclerosis based on phonatory and time-frequency features. *Sensors.* 2022;22(3):1137. <https://doi.org/10.3390/s22031137>.
36. Thomas A, Teplansky KJ, Wisler A, Heitzman D, Austin S, Wang J. Voice onset time in early-and late-stage amyotrophic lateral sclerosis. *J Speech Lang Hear Res.* 2022;65(7):2586–93. https://doi.org/10.1044/2022_JSLHR-21-00632.
37. Tomik B, Guiloff RJ. Dysarthria in amyotrophic lateral sclerosis: a review. *Amyotroph Lateral Scler.* 2010;11(1–2):4–15. <https://doi.org/10.3109/17482960802379004>.
38. van Erp S, Oberski DL, Mulder J. Shrinkage priors for Bayesian penalized regression. *J Math Psychol.* 2019;89:31–50. <https://doi.org/10.1016/j.jmp.2018.12.004>.
39. Vashkevich M, Rushkevich Y. Classification of ALS patients based on acoustic analysis of sustained vowel phonations. *Biomed Signal Process Control.* 2021. <https://doi.org/10.1016/j.bspc.2020.102350>.
40. Yunusova Y, Graham NL, Shellikeri S, Phuong K, Kulkarni M, Rochon E, Tang-Wai DF, Chow TW, Black SE, Zinman LH, Green JR. Profiling speech and pausing in amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). *PLoS ONE.* 2016;11(1):1–18. <https://doi.org/10.1371/journal.pone.0147573>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.