INTERSPEECH 2023
20-24 August 2023, Dublin, Ireland

Multimodal assessment of bulbar amyotrophic lateral sclerosis (ALS) using a novel remote speech assessment app

Leif Simmatis[1,2,3], *Timothy Pommeé*[1,2,3], *Yana Yunusova*[1,2,3]

[1]Dept. Speech-Language Path., Temerty Faculty of Medicine, Univ. Toronto, Toronto, ON, Canada
[2]KITE-Toronto Rehabilitation Institute, Toronto, ON, Canada
[3]Sunnybrook Research Institute, Sunnybrook Health Sciences Centre, Toronto, ON, Canada
leif.simmatis@uhn.ca, timothy.pommee@sri.utoronto.ca, yana.yunusova@utoronto.ca

Abstract

Speech is a valuable marker of disease onset and progression in amyotrophic lateral sclerosis (ALS). Acoustic and kinematic data have characterized speech impairments in ALS previously, and there is growing interest in combining these modalities in novel analytical platforms. We explored the use of a multimodal (audio/video) speech assessment pipeline in ALS patients with varying severities. Participants performed a passage reading task, and clinical outcomes of e.g., speech function were collected. Speech data were analyzed using a custom automated acoustic and kinematic pipeline. Sparse canonical correlation analysis (SCCA) was then used. *Both* acoustic and kinematic features loaded strongly with clinical data (loadings ≥|0.50|), indicating that multimodal features captured complementary speech function information. This reinforces the value of multimodal assessment techniques and points the way towards future remote assessment development steps.

Index Terms: Amyotrophic lateral sclerosis, multimodal, remote assessment, validation

1. Introduction

Amyotrophic lateral sclerosis (ALS) is a debilitating and incurable motor neuron disease that frequently causes speech/bulbar impairments [1]. Bulbar motor impairments are additionally associated with both increased risk of cognitive impairment as well as reduced life expectancy [2]–[4]. Therefore, it is of paramount importance to develop methods to track bulbar motor dysfunction in ALS. It is an open question as to whether or not speech information from multiple modalities provides additional utility over and above single-modality recording methods.

Both acoustic (audio) and kinematic (video) features have been independently established as important in the assessment of bulbar ALS. Previous research has suggested that acoustic measures of articulation, phonation, hypernasality and impaired breathing are important for capturing impairments associated with ALS [5], [6]. Kinematic features have also been established for detection of ALS, particularly in the early disease phases [7]. Many existing studies have focused on using lab-based equipment, including depth (i.e., 3-dimensional/3D) cameras, sensor-based electromagnetic articulography, and cardioid microphones attached to high-performance external sound cards. These technologies, while highly controllable in a laboratory setting, are infeasible for both standard clinical assessment and for remote assessment. Current research efforts are devoted to the development and validation of remote methods that employ built in computer cameras and microphones. Remote methods have gained particular interest given the COVID-19 pandemic as well as their ability to conveniently gather high-frequency data samples.

The majority of existing remote assessment systems are focused on acoustics, but there is evidence that multimodal information may be of value. For example, [8] employed a conversational artificial intelligence (AI) agent to collect data from ALS and healthy control participants, and reported excellent detection of ALS, as well as presymptomatic bulbar ALS patients from control participants. This study found that a mixture of acoustic (e.g., pause and phrase durations) and kinematic (e.g., lower lip velocity) features could detect ALS, highlighting the importance of multimodal information. Building upon this, it will be important to establish relationships between multimodal speech features and measures of disease status, such as clinical outcome measures.

To these ends, we developed a novel multimodal speech assessment app that can collect audio and video data asynchronously, with diverse acoustic and kinematic features. In the present study, we sought to determine whether *multimodal* data collected using this novel app could capture impairment across a wide range of bulbar dysfunction in a complementary manner. We used sparse canonical correlation analysis (SCCA) to characterize loading patterns across acoustic, kinematic, and clinical variables. We hypothesized that multimodal (acoustic and kinematic) features would have strong loadings on the same components as clinical measures (loadings ≥|0.50| magnitude, based on recommendations from factor analysis literature [9]), which would indicate complementary information across both modalities.

2. Methods

2.1. Participants and data collection (app)

Data from 34 individuals with ALS were recorded as part of a larger study focused on the development of a web-based app for the assessment of bulbar ALS. Participants were included on the basis of (1) a clinical diagnosis of ALS, (2) fluency in English, (3) a minimum age of 45 years, (4) no history of other neurological or speech disorders, and (5) absence of anarthria, i.e., complete loss of ability to speak (as defined by at least some remaining use of oral communication in daily life). All participants provided informed consent in accordance with the Declaration of Helsinki.

Data were gathered using an in-house developed multimodal audio/video assessment app. The app was designed to collect data asynchronously (i.e., data recording followed by upload, as opposed to streaming) in order to avoid issues arising

10.21437/Interspeech.2023-2093

from buffering/variation in internet connection speeds as might occur using a platform such as Zoom. Participants used their own computers (MacOS ≥10.10 or Windows ≥10; current version of Google Chrome ver. ≥40.0). The setup of participants was guided by experienced speech-language pathologists, and was performed to minimize variation in lighting, pose, etc.

The app assessment included a variety of speech and non-speech/orofacial tasks. Speech data gathered using the app consisted of e.g., reading a standardized passage, sentences, syllables, etc. Here only 99-word passage recordings were analyzed, (the "bamboo passage", hereafter, "Bamboo"). Audio was collected at 48 kHz, and corresponding video was captured at the maximum possible for participants' hardware (typically ~30 frames per second [FPS]).

Prior to analysis, data were subjected to manual quality control as well as quantitative checking of basic elements of data quality. The primary criteria were video frame rate >28 FPS and audio signals spared of excessive background noise. After applying these rules, we retained data from 29 individuals (see Table 1).

2.2. Clinical variables

A number of clinical/demographic variables were gathered that captured various aspects of clinical and motor function. The ALS Functional Rating Scale – Revised (ALSFRS-R, "FRS") was used as a measure of overall ALS-related functional status; specifically, the total score and the bulbar sub-score ("FRS-Bulb") were collected. We additionally gathered the Center for Neurologic Study-Bulbar Function Score (BFS) and its speech sub-score ("BFS-Speech") [10].

Table 1: *Demographics and clinical measures.*

	Values (median[IQR] unless specified)
Age (years)	68 [12.3]
N (F)	29 (7)
FRS-Bulb	10.0 [3.8]
FRS-Total	37.5 [7.0]
BFS-Speech	9.0 [12.5]
BFS-Total	27.0 [27.3]

2.3. Acoustic feature extraction

Acoustic features were extracted using a custom pipeline that in total captured 159 features. Our pipeline was focused on extracting features that had proven clinical value in previous works. In the following description of our pipeline, we subdivided features into different speech subsystems and/or their combinations. All features were extracted using Parselmouth, which is a Python-based interface to Praat functionality [11] and, unless otherwise specified, Praat default values (e.g., audio window lengths for formants) were used.

For many of the features calculated by our acoustic pipeline, it was first necessary to divide the speech sample into voiced- and voiceless components. Voiced segments were extracted using custom, automated Praat scripts and used to estimate the performance of the phonatory subsystem. These included several measures of jitter and shimmer, as well as harmonic/noise ratio (HNR) and fundamental frequency (F_o) [12], [13], and measures from the Acoustic Voice Quality Index (AVQI) [14]. This encompassed 32 features.

Respiratory subsystem measures primarily focused on pause and phrase durations, as well as overall speaking rate. These features have been explored extensively in the context of ALS [15]–[18] and have been found to relate to disease severity as well as disease progression. This encompassed 7 features.

Articulatory features have been extensively validated in ALS and other neurological diseases. Formant ranges and trajectories (i.e., derivatives) have been utilized for dysarthric speech detection in a variety of clinical contexts. For example, the trajectory of the first and second formants have been noted as indices of vocal tract velocity [19], [20]. Here, we extracted the first five formants (F1-F5) and calculated the following measures: 5th percentile, 95th percentile, the range between 5th and 95th percentile, mean, and standard deviation (SD). We also extracted the first and second derivatives of the first three formants, and then calculated their respective descriptive statistics. We also extracted the mean and variance of the first thirteen MFCCs as well as of their first derivatives. Additionally, Slis et al. (2021) described the utility of condensing multiple MFCCs into a single value called the total squared change in MFCCs (*tsc_mfcc*) [21], which we calculated here. Finally, articulatory entropy [22] was extracted as a means to capture the working articulatory space without having to resort to manually calculating vowel space area using corner vowels. In total, this encompassed 108 features.

Finally, features of coordination between subsystem measures were extracted [23], [24]. These consisted of cross-correlations between the first three formants, cepstral peak prominence, intensity, and F_o. However, to reduce our feature space where possible, we captured the index of the eigenvalue that corresponded to 95% of the variance in the cross-correlation signal. Based on the work of [23], [24], we expected that a lower index would correspond to pathological speech patterns (i.e., the area under the eigenvalue distribution being shifted to the left; see previous papers for further detail on the expected correspondence between eigenspectrum complexity and health or disease status). This encompassed 12 features.

2.4. Kinematic feature extraction

Facial landmark tracking was performed using the Google Mediapipe framework [25], which extracted 478 3D facial landmarks from each frame of video. This model was chosen because we established that it subjectively tracked facial landmarks well in its opensource configuration across a variety of ages and skin tones, and it is also capable of running at real time (~30FPS) or faster on regular CPUs.

We focused on tracking a single landmark at the median of the lower vermilion border of the mouth (LL) which has been used in clinical studies previously for tracking oral movements [26]. Additionally, the medial canthi of the eyes were tracked, as was a reference point in the centre of the

forehead. Canthal landmarks were used for distance standardization, i.e., the computation of the intercanthal distance (ICD) (to account for different distances between the face and the camera over time). The reference landmark (REF) was used for determining the movement of the LL point. Essentially, the distance that was measured throughout the video was computed as (REF-LL)/ICD, where REF-LL is the Euclidean distance between each of the two 3D points over time.

Prior to subsequent processing, facial landmarks were smoothed using singular spectrum analysis (SSA) [27], in which each of the 478*3 timeseries of landmark trajectories was decomposed into components, the first of which was retained as the cleaned version of the original signal. Filtering window length was set at 60ms in order to avoid over-smoothing of the data, which corresponded to 2 video frames. SSA has properties of being a data-adaptive finite impulse response (FIR) filter with zero-phase characteristics.

Features that captured the clinical domains of speed and range of motion (ROM) were extracted from the smoothed data. Kinematic features included 5th, 25th, 50th, and 95th percentiles of absolute values of the first and second derivatives of position (i.e., speed and acceleration), as well as the ROM, which was the difference between the 95th and 5th percentiles of the normalized REF-LL distance. This encompassed 9 features in total.

2.5. Sparse canonical correlation analysis (SCCA)

Because we extracted diverse types of data from a relatively small dataset using multiple modalities (acoustic, kinematic, and clinical), we used a robust statistical procedure to understand relationships between different data modalities. We chose the canonical correlation analysis (CCA), which identifies common factor loadings across multiple datasets that are related to corresponding underlying variability patterns. Owing to the small number of samples compared to the number of features, we employed a penalized matrix decomposition (PMD) formulation of CCA [28] called sparse CCA (SCCA). Prior to SCCA, we rescaled features using standard normalization (i.e., mean of 0 and variance of 1) to ensure that scaling effects did not negatively impact the interpretation of the loadings that were found.

Because we used a relatively small dataset, we also decided to restrict our considered "substantial" loadings to those that were ≥|0.50|. This has previously been demonstrated to be a conservative threshold for use in classical exploratory factor analysis [29]. In order to select how many latent dimensions to extract during analysis, we first fitted a model with a large number of latent variables (10) and observed the scores corresponding to each dimension. The first 4 dimensions had scores ≥0.5 (comparable interpretation to a correlation coefficient; scores 0 to 1, with 1 being better), and the 7th to 10th had scores <0.35, indicating strongly that 4 latent dimensions represented the optimal structure of the data for further analysis. After extracting the loadings from each of the acoustic/kinematic and clinical datasets, we then concatenated them and performed clustering analysis in order to extract patterns of loadings across features, to see which groups of features "behaved" similarly. Hierarchical agglomerative clustering was used for visualization. Analyses were performed in Python (v 3.8.9) using a combination of custom scripts and packages, such as the CCA-zoo package [30].

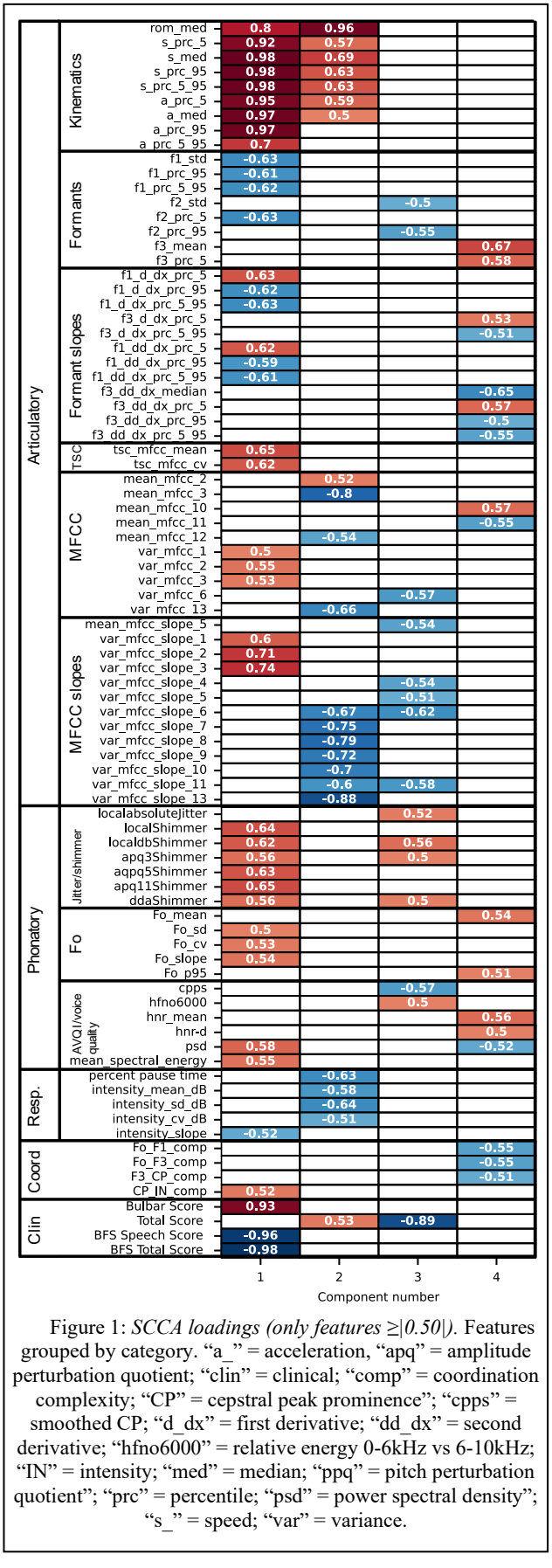

Figure 1: *SCCA loadings (only features ≥|0.50|).* Features grouped by category. "a_" = acceleration, "apq" = amplitude perturbation quotient; "clin" = clinical; "comp" = coordination complexity; "CP" = cepstral peak prominence"; "cpps" = smoothed CP; "d_dx" = first derivative; "dd_dx" = second derivative; "hfno6000" = relative energy 0-6kHz vs 6-10kHz; "IN" = intensity; "med" = median; "ppq" = pitch perturbation quotient"; "prc" = percentile; "psd" = power spectral density"; "s_" = speed; "var" = variance.

3. Results

Results of the SCCA analysis are summarized in Figure 1, and they highlight important cross-modality patterns in the kinematic, acoustic, and clinical data. Note that for brevity, Figure 1 only shows features that had loading strengths ≥|0.50| in at least one dimension.

The first component captured loadings across diverse acoustic and kinematic features. This included all kinematics, jitter/shimmer, MFCC features such as *tsc_mfcc*, features from the AVQI, and some additional measures of F_0 variability and coordination complexity, as well as FRS-Bulb and the BFS scores. The second component had strong loadings with MFCC mean and variance features, intensity features, and kinematics, as well as FRS-Total. The third component had strong loadings for diverse features spanning subsets of jitter/shimmer, formants, AVQI, and MFCC/MFCC slope categories, as well as FRS-Total. Finally, the fourth component captured relationships between formant/formant slopes and coordination features. The fourth component had no strong clinical loadings.

As a specific example of the patterns of loadings between features, Figure 2 depicts a biplot of loadings from the first and third components. From the Figure, it is clear that the kinematic features tended to strongly associate with one another. Also of note is that the kinematic features tended to strongly align along with the BFS measures, as well as *tsc_mfcc*, jitter/shimmer, and formant slopes.

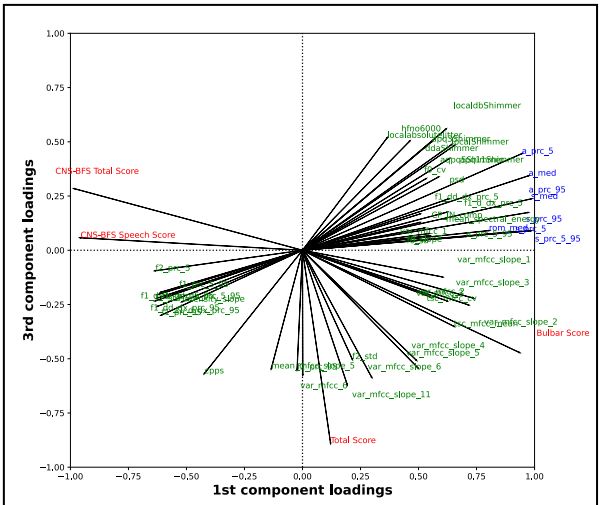

Figure 2: *Loading patterns across the first and third canonical components from the SCCA analysis.* Blue indicates kinematic features, green indicates acoustic features, and red indicates clinical measures.

4. Discussion

In this study, we evaluated the complementary contributions of multimodal speech data (acoustic/audio and kinematic/video) to capture clinical impairments in patients with ALS, as indexed by strong shared loadings onto canonical components. We found that our features from both acoustic and kinematic domains had strong relationships with clinical measures (shared component loadings ≥|0.50|), indicating shared contribution to a latent bulbar impairment construct. These results, although exploratory in nature, highlight the importance of multimodal assessments for capturing motor speech impairments in ALS.

We observed in this study that groups of features from different domains associated with each other (i.e., they loaded onto the same canonical components strongly), and associated strongly with clinical measures. For example, in the first component, kinematic features were associated with various measures of jitter and shimmer, as well as clinical measures of bulbar function (FRS-bulbar, BFS-Total, and BFS-Speech), *tsc_mfcc*, and coordination measures. These feature sets have previously been found to relate to speech function in other populations [21], [23]. Importantly for the context of ALS, the first canonical component seemed to capture an overall bulbar impairment pattern characterized by perturbations to the motor control of the phonatory and articulatory subsystems. A previous study identified these subsystems as specifically important for early detection of ALS [31], which lends validity to the findings of our current study and reinforces the importance of the multimodal approach.

Many of the loadings had strongly opposing directions, which simply indicated the directionality of association – features that were truly unassociated were orthogonal. This effect is best appreciated in Figure 2; features with opposing vector directions were strongly *anti*correlated, whereas those at close to 90-degree angles are *un*correlated. An example of where this is salient is FRS-Total, which loaded with some acoustic and kinematic features along component 3 (-0.89) but did not share their across-component loading patterns (e.g., jitter/shimmer features).

Our study had some limitations. Our sample size was relatively small, and so the findings demonstrated must be considered as exploratory. However, analyses using other SCCA methods found similar patterns of loadings (data not shown for brevity) and so we can be reasonably confident that the findings depicted here represented real effects given the current dataset. Additionally, SCCA being an unsupervised measure of association, we did not perform the clinically-relevant task of disease prediction using multimodal information; we leave this to future, larger, studies. Furthermore, we did not have an adequate sample size to differentiate possible patterns across male and female participants. This will be explored in future with more in-depth analyses of larger cohorts, which will be collected with less direct oversight of clinicians. Finally, some features such as formant slopes and those representing the resonatory subsystem will require phoneme-level automated processing. This will be addressed in future work.

5. Conclusions

We identified that a multimodal assessment of speech could capture kinematic and acoustic feature patterns that corresponded intelligibly to each other, and to clinical outcome measures that are of interest in ALS research. These findings highlight the substantial value of multimodal speech assessment systems in ALS, and provide justification for future studies of multimodal, remote speech analysis systems in ALS as well as various other neurological and neurodegenerative diseases.

6. Acknowledgements

We would like to sincerely thank Dr. Madhura Kulkarni, Tara Lazetic, and Porsha Taheri for their contributions to data management and processing during this project.

7. References

[1] B. Tomik and R. J. Guiloff, "Dysarthria in amyotrophic lateral sclerosis: A review," *Amyotrophic Lateral Sclerosis*, vol. 11, no. 1–2. pp. 4–15, 2010. doi: 10.3109/17482960802379004.

[2] S. S. Gubbay, E. Kahana, N. Zilber, G. Cooper, S. Pintov, and Y. Leibowitz, "Amyotrophic lateral sclerosis. A study of its presentation and prognosis," *J. Neurol.*, vol. 232, no. 5, pp. 295–300, 1985, doi: 10.1007/BF00313868.

[3] H. Schreiber *et al.*, "Cognitive function in bulbar- and spinal-onset amyotrophic lateral sclerosis: A longitudinal study in 52 patients," *J. Neurol.*, vol. 252, no. 7, pp. 772–781, 2005, doi: 10.1007/s00415-005-0739-6.

[4] L. E. Sterling *et al.*, "Association between dysarthria and cognitive impairment in ALS: A prospective study," *Amyotroph. Lateral Scler.*, vol. 11, no. 1–2, pp. 46–51, 2010, doi: 10.3109/17482960903207997.

[5] P. Rong *et al.*, "Predicting speech intelligibility decline in amyotrophic lateral sclerosis based on the deterioration of individual speech subsystems," *PLoS One*, vol. 11, no. 5, May 2016, doi: 10.1371/journal.pone.0154971.

[6] M. Eshghi, K. P. Connaghan, S. E. Gutz, J. D. Berry, Y. Yunusova, and J. R. Green, "Co-occurrence of hypernasality and voice impairment in amyotrophic lateral sclerosis: Acoustic quantification," *J. Speech, Lang. Hear. Res.*, vol. 64, no. 12, pp. 4772–4783, 2021, doi: 10.1044/2021_JSLHR-21-00123.

[7] A. Bandini, J. R. Green, L. Zinman, and Y. Yunusova, "Classification of bulbar ALS from kinematic features of the jaw and lips: Towards computer-mediated assessment," *Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH*, vol. 2017-Augus, pp. 1819–1823, 2017, doi: 10.21437/Interspeech.2017-478.

[8] M. Neumann *et al.*, "Investigating the utility of multimodal conversational technology and audiovisual analytic measures for the assessment and monitoring of amyotrophic lateral sclerosis at scale," *Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH*, vol. 4, pp. 3061–3065, 2021, doi: 10.21437/Interspeech.2021-1801.

[9] A. B. Costello and J. Osborne, "Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis," *Res. Eval. Pract. Assessment, Res. Eval.*, vol. 10, p. 7, 2005, doi: 10.7275/jyj1-4868.

[10] R. Smith *et al.*, "Enhanced Bulbar Function in Amyotrophic Lateral Sclerosis: The Nuedexta Treatment Trial," *Neurotherapeutics*, vol. 14, no. 3, pp. 762–772, 2017, doi: 10.1007/s13311-016-0508-5.

[11] Y. Jadoul, B. Thompson, and B. de Boer, "Introducing Parselmouth: A Python interface to Praat," *J. Phon.*, vol. 71, pp. 1–15, Nov. 2018, doi: 10.1016/j.wocn.2018.07.001.

[12] R. D. Kent, R. L. Sufit, J. C. Rosenbek, J. F. Kent, R. E. Martin, and B. R. Brooks, "Speech Deterioration in Amyotrophic Lateral Sclerosis: A Case Study.," *J. Speech Hear. Res.*, vol. 34, no. December, pp. 1269–1275, 1991.

[13] A. K. Silbergleit, A. F. Johnson, and B. H. Jacobson, "Acoustic Analysis of Voice in Individuals with Amyotrophic Lateral Sclerosis and Perceptually Normal Vocal Quality," 1997.

[14] Y. Maryn, P. Corthals, P. Van Cauwenberge, N. Roy, and M. De Bodt, "Toward improved ecological validity in the acoustic measurement of overall voice quality: Combining continuous speech and sustained vowels," *J. Voice*, vol. 24, no. 5, pp. 540–555, 2010, doi: 10.1016/j.jvoice.2008.12.014.

[15] A. A. Waito *et al.*, "Validation of Articulatory Rate and Imprecision Judgments in Speech of Individuals With Amyotrophic Lateral Sclerosis Ashley," *Am. J. Speech-Language Pathol.*, vol. 30, pp. 137–149, 2021.

[16] A. S. Mefferd, G. L. Pattee, and J. R. Green, "Speaking rate effects on articulatory pattern consistency in talkers with mild ALS," *Clin. Linguist. Phonetics*, vol. 28, no. 11, pp. 799–811, Nov. 2014, doi: 10.3109/02699206.2014.908239.

[17] C. Barnett *et al.*, "Reliability and validity of speech & pause measures during passage reading in ALS," *Amyotroph. Lateral Scler. Front. Degener.*, vol. 21, no. 1–2, pp. 42–50, Jan. 2020, doi: 10.1080/21678421.2019.1697888.

[18] Y. Yunusova *et al.*, "Profiling speech and pausing in amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD)," *PLoS One*, vol. 11, no. 1, Jan. 2016, doi: 10.1371/journal.pone.0147573.

[19] R. L. Horwitz-Martin *et al.*, "Relation of automatically extracted formant trajectories with intelligibility loss and speaking rate decline in amyotrophic lateral sclerosis," *Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH*, vol. 08-12-Sept, no. February 2018, pp. 1205–1209, 2016, doi: 10.21437/Interspeech.2016-403.

[20] H. P. Rowe, K. L. Stipancic, A. C. Lammert, and J. R. Green, "Validation of an acoustic-based framework of speech motor control: Assessing criterion and construct validity using kinematic and perceptual measures," *J. Speech, Lang. Hear. Res.*, vol. 64, no. 12, pp. 4736–4753, 2021, doi: 10.1044/2021_JSLHR-21-00201.

[21] A. Slis, N. Lévêque, C. Fougeron, M. Pernon, F. Assal, and L. Lancia, "Analysing spectral changes over time to identify articulatory impairments in dysarthria," *J. Acoust. Soc. Am.*, vol. 149, no. 2, pp. 758–769, 2021, doi: 10.1121/10.0003332.

[22] Y. Jiao, V. Berisha, J. Liss, S. C. Hsu, E. Levy, and M. McAuliffe, "Articulation entropy: An unsupervised measure of articulatory precision," *IEEE Signal Process. Lett.*, vol. 24, no. 4, pp. 485–489, 2017, doi: 10.1109/LSP.2016.2633871.

[23] T. Talkar *et al.*, "Acoustic indicators of speech motor coordination in adults with and without traumatic brain injury," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2021, vol. 1, pp. 426–430. doi: 10.21437/Interspeech.2021-1581.

[24] T. F. Quatieri, T. Talkar, and J. S. Palmer, "A Framework for Biomarkers of COVID-19 Based on Coordination of Speech-Production Subsystems," *IEEE Open J. Eng. Med. Biol.*, vol. 1, pp. 203–206, 2020, doi: 10.1109/OJEMB.2020.2998051.

[25] Y. Kartynnik, A. Ablavatski, I. Grishchenko, and M. Grundmann, "Real-time Facial Surface Geometry from Monocular Video on Mobile GPUs," Jul. 2019, [Online]. Available: http://arxiv.org/abs/1907.06724

[26] L. Simmatis, C. Barnett, R. Marzouqah, B. Taati, M. Boulos, and Y. Yunusova, "Reliability of Automatic Computer Vision-Based Assessment of Orofacial Kinematics for Telehealth Applications," *Digit. Biomarkers*, vol. 6, no. 2, pp. 71–82, Jul. 2022, doi: 10.1159/000525698.

[27] F. J. Alonso, J. M. Del Castillo, and P. Pintado, "Application of singular spectrum analysis to the smoothing of raw kinematic signals," *J. Biomech.*, vol. 38, no. 5, pp. 1085–1092, May 2005, doi: 10.1016/j.jbiomech.2004.05.031.

[28] D. M. Witten, R. Tibshirani, and T. Hastie, "A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis," *Biostatistics*, vol. 10, no. 3, pp. 515–534, Jul. 2009, doi: 10.1093/biostatistics/kxp008.

[29] A. B. Costello and J. W. Osborne, "Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. - Practical Assessment, Research & Evaluation," vol. 10, 2005.

[30] J. Chapman and H.-T. Wang, "CCA-Zoo: A collection of Regularized, Deep Learning based, Kernel, and Probabilistic CCA methods in a scikit-learn style framework," *J. Open Source Softw.*, vol. 6, no. 68, p. 3823, Dec. 2021, doi: 10.21105/joss.03823.

[31] P. Rong, Y. Yunusova, J. Wang, and J. R. Green, "Predicting early bulbar decline in amyotrophic lateral sclerosis: A speech subsystem approach," *Behav. Neurol.*, vol. 2015, 2015, doi: 10.1155/2015/183027.