INTERSPEECH 2021
30 August – 3 September, 2021, Brno, Czechia

Investigating the Utility of Multimodal Conversational Technology and Audiovisual Analytic Measures for the Assessment and Monitoring of Amyotrophic Lateral Sclerosis at Scale

Michael Neumann[1], *Oliver Roesler*[1], *Jackson Liscombe*[1], *Hardik Kothare*[1,2], *David Suendermann-Oeft*[1], *David Pautler*[1], *Indu Navar*[3], *Aria Anvar*[3], *Jochen Kumm*[4], *Raquel Norel*[5], *Ernest Fraenkel*[6], *Alexander V. Sherman*[7,8], *James D. Berry*[7], *Gary L. Pattee*[9], *Jun Wang*[10], *Jordan R. Green*[7,8] *and Vikram Ramanarayanan*[1,2]

[1] Modality.AI, Inc., San Francisco, USA
[2] University of California, San Francisco, USA
[3] EverythingALS, Peter Cohen Foundation, Los Altos, USA
[4] Pr3vent, Inc., Palo Alto, USA
[5] IBM Thomas J. Watson Research Center, Yorktown Heights, USA
[6] Massachusetts Institute of Technology, Cambridge, USA
[7] MGH Institute of Health Professions, Boston, USA
[8] Harvard University, Cambridge, USA
[9] University of Nebraska, USA
[10] University of Texas at Austin, USA

vikram.ramanarayanan@modality.ai

Abstract

We propose a cloud-based multimodal dialog platform for the remote assessment and monitoring of Amyotrophic Lateral Sclerosis (ALS) at scale. This paper presents our vision, technology setup, and an initial investigation of the efficacy of the various acoustic and visual speech metrics automatically extracted by the platform. 82 healthy controls and 54 people with ALS (pALS) were instructed to interact with the platform and completed a battery of speaking tasks designed to probe the acoustic, articulatory, phonatory, and respiratory aspects of their speech. We find that multiple acoustic (rate, duration, voicing) and visual (higher order statistics of the jaw and lip) speech metrics show statistically significant differences between controls, bulbar symptomatic and bulbar pre-symptomatic patients. We report on the sensitivity and specificity of these metrics using five-fold cross-validation. We further conducted a LASSO-LARS regression analysis to uncover the relative contributions of various acoustic and visual features in predicting the severity of patients' ALS (as measured by their self-reported ALSFRS-R scores). Our results provide encouraging evidence of the utility of automatically extracted audiovisual analytics for scalable remote patient assessment and monitoring in ALS.

Index Terms: conversational agent, amyotrophic lateral sclerosis, computer vision, dialog systems.

1. Multimodal Conversational Agents for Health Monitoring

The development of technologies that rapidly diagnose medical conditions, recognize pathological behaviors, continuously monitor patient status, and deliver just-in-time interventions using the user's native technology environment remains a critical need today [1]. The COVID-19 pandemic has further highlighted the need to make telemedicine and remote monitoring more readily available to patients with chronic neurological disorders [2]. However, early detection or progress monitoring of neurological or mental health conditions, such as clinical depression, ALS, Alzheimer's disease, dementia, etc., is often challenging for patients due to various reasons, including, but not limited to: (i) no access to neurologists or psychiatrists; (ii) lack of awareness of a given condition and the need to see a specialist; (iii) lack of an effective standardized diagnostic or endpoint; (iv) substantial cost and transportation involved in conventional or traditional solutions; and in some cases (v) lack of medical specialists in these fields [3].

The NEurological and Mental health Screening Instrument (NEMSI) [4] was developed to bridge this gap. NEMSI is a cloud-based multimodal dialog system that can be used to elicit evidence required for detection or progress monitoring of neurological or mental health conditions through automated screening interviews conducted over the phone or via web browser. While intelligent virtual agents have been proposed in previous work for such diagnosis and monitoring purposes [5, 6], NEMSI offers three significant innovations: First, NEMSI uses readily available devices (web browser or mobile app), in contrast to dedicated, locally administered hardware. Second, NEMSI's backend is deployed in an automatically scalable cloud environment allowing it to serve an arbitrary number of end-users at a small cost per interaction. Third, the NEMSI system is equipped with real-time analytics modules that extract a variety of speech and video features of direct relevance to clinicians, such as speech and pause duration for the assessment of ALS, or geometric features derived from facial landmarks to automatically detect orofacial impairment in stroke.

This paper explores the utility of audio and video metrics collected via NEMSI for early diagnosis and monitoring of ALS. We specifically investigate two research questions. First, which metrics show statistically significant differences between (a) healthy controls and bulbar pre-symptomatic people with ALS – aka pALS – (thereby assisting in early diag-

http://dx.doi.org/10.21437/Interspeech.2021-1801

Table 1: *Participant characteristics for the three groups – controls (CON), bulbar symptomatic (BUL), and bulbar pre-symptomatic (PRE). Age and ALSFRS-R scores are presented as: median; mean (standard deviation).*

Group	Female	Male	Age (years)	ALSFRS-R	Bulbar sub score
CON	68	14	43; 41.62 (19.00)	48; 47.89 (0.94)	12; 11.94 (0.36)
BUL	17	15	63; 59.56 (10.29)	36; 33.09 (7.46)	9; 8.75 (1.57)
PRE	12	10	61; 57.18 (11.31)	40; 36.45 (8.58)	12; 12.00 (0.00)

nosis), as well as (b) bulbar-presymptomatic pALS and bulbar symptomatic pALS (thereby assisting in progress monitoring)? Second, for pALS cohorts, which metrics are most predictive of their self-reported ALS Functional Rating Scale-Revised (ALSFRS-R [7]) score? Before addressing these questions, we briefly summarize the current state of ALS research, and describe our data collection and metrics extraction process.

2. Current State of ALS Diagnosis and Monitoring

ALS is a neurodegenerative disease that affects roughly 4 to 6 people per 100,000 of the general population [8, 9]. Early detection and continuous monitoring of ALS symptoms is crucial to provide optimal patient care [10]. For instance, decline in speech intelligibility negatively affects patients' quality of life [11, 12], and continuous monitoring of speech intelligibility could prove valuable in terms of patient care. Traditionally, subjective measures, such as patient-reports or ratings by clinicians, are used to detect and monitor speech impairment. However, recent studies show that objective measures allow for earlier detection of ALS symptoms [13, 14, 15, 16, 17, 18, 19, 20]; stratification and classification of patients [21]; and can provide markers for disease onset, progression and severity [22, 23, 24, 25, 26, 27]. These objective measures can be automatically extracted, thereby allowing for more frequent monitoring, potentially improving treatment. The success of the Beiwe Research Platform [28] to track ALS disease progression demonstrates the viability of such remote monitoring solutions.

3. Methods

3.1. Collection Setup

NEMSI end users are provided with a website link to the secure screening portal and login credentials by their study liaison (physician, clinic, a referring website or patient portal). After completing microphone and camera checks, subjects participate in a conversation with a virtual dialog agent. The agent engages subjects in a mixture of structured speaking tasks and open-ended questions to elicit speech and facial behaviors relevant for the condition being screened for. Analytics modules automatically extract speech (e.g., speaking rate, duration measures, fundamental frequency (F0)) and video features (e.g., range and speed of movement of various facial landmarks) in real time and store them in a database. All this information can be accessed by the study liaison through a dashboard, which provides a summary of the interaction (including a video recording), and a detailed breakdown of the metrics by individual interaction turns.

3.2. Data

The conversational protocol elicits five types of speech samples from participants, inspired by prior work [29, 30, 31, 32]: (a) sustained vowel phonation, (b) read speech, (c) measure of diadochokinetic (DDK) rate (rapidly repeating the syllables /pɑtɑkɑ/), and (d) free speech (picture description task). For (b) read speech, the dialog contains six speech intelligibility test (SIT) sentences of increasing length (5 to 15 words), and one passage reading task (Bamboo Passage; 99 words). After dialog completion, participants filled out the *ALS Functional Rating Scale-revised (ALSFRS-R)*, a standard instrument for monitoring the progression of ALS [7]. The questionnaire consists of 12 questions about physical functions in activities of daily living. Each question provides five answer options, ranging from *normal function* (score 4) to *severe disability* (score 0). The ALSFRS-R score is the sum of all sub-scores (ranging from 0 to 48). The ALSFRS-R comprises four scales for different domains affected by the disease: bulbar system, fine and gross motor skills, and respiratory function.

Data from 136 participants (see Table 1) were collected between September 2020 and March 2021 in cooperation with EverythingALS and the Peter Cohen Foundation[1]. For this cross-sectional study we included one dialog session per subject.[2] We stratified subjects into three groups for statistical analysis: (a) Healthy controls (*CON*); (b) pALS with a bulbar subscore < 12 (first three ALSFRS-R questions) were labeled bulbar symptomatic (*BUL*); and (c) pALS with a bulbar sub-score of 12 were labeled bulbar pre-symptomatic (*PRE*). Similar to [14] we aim at identifying acoustic and visual speech measures that show significant differences between these groups.

4. Signal Processing and Metrics Extraction

4.1. Acoustic Metrics

We use measures commonly established for clinical speech analysis with regard to ALS [14], including *timing* and *frequency domain measures*, and measures specific to the DDK task, such as syllable rate and cycle-to-cycle temporal variation [33]. Table 2 shows the metrics and speech task types from which they are extracted. Additionally, speech intensity (mean energy in dB SPL excluding pauses) was extracted for all turns. The picture description task was not used for this analysis.

All acoustic measures were automatically extracted with the speech analysis software Praat [34]. Speaking and articulation rates are computed based on *expected* number of words because forced alignment is error-prone for dysarthric speech [35]. For that reason, these measures can be noisy if, for example, a patient did not finish the reading passage. Hence, we automatically removed outliers based on thresholds for the Bamboo task: speaking rates $>$ 250 words/min, articulation rates $>$ 350 words/min, and PPT $> 80\%$ are excluded.[3]

4.2. Visual Metrics

Facial metrics were calculated for each utterance in three steps: (i) face detection using the *Dlib*[4] face detector, which uses

[1] https://www.everythingals.org/research

[2] If a subject participated in multiple dialog sessions, we took the first successful one; i.e., the first *complete* call for which valid metrics were extracted and all ALSFRS-R questions were answered.

[3] The thresholds were determined by manual inspection of the data.

[4] http://dlib.net/

Table 2: *Acoustic metrics. HNR: harmonics-to-noise ratio, CPP: cepstral peak prominence, DDK: diadochokinesia, PPT: percent pause time, cTV: cycle-to-cycle temporal variation.*

Speech type	Collected metrics
Held vowel	Mean F0 (Hz), jitter (%), shimmer (%), HNR (dB), CPP (dB)
SIT and Bamboo	Speaking and articulation duration (sec), speaking and articulation rate (words/min), PPT
DDK	Speaking and articulation duration (sec), Syllable rate (syllables/sec), number of produced syllables, cTV (sec)

Table 3: *Visual metrics. Maximum (suffix _max) and average (_avg) were extracted for all metrics; and for velocity, acceleration, and jerk, minimum (_min) was also extracted.*

Category	Metrics	Description
Movement	open width LL_path JC_path eye_open eyebrow_vpos	Lips opening and mouth width; Displacement of lower lip and jaw center; Eye opening; Vertical eyebrow displacement
Velocity	vLL vJC vLL_abs vJC_abs	Velocity and speed of lower lip and jaw center (px/frames)
Acceleration	aLL aJC aLL_abs aJC_abs	Acceleration of lower lip and jaw center (px/frames^2)
Jerk	jLL jJC jLL_abs jJC_abs	Jerk of lower lip and jaw center (px/frames^3)
Surface	S (S_R, S_L) S_ratio_avg	Area of the mouth (right and left half, px^2) Mean symmetry ratio
Eye blinks	eye_blinks	Eye blinks per sec.

five histograms of oriented gradients to determine the (x, y)-coordinates of one or more faces for every input frame [36], (ii) facial landmark extraction using the Dlib facial landmark detector, which uses an ensemble of regression trees proposed in [37] to extract 68 facial landmarks according to Multi-PIE [38], and (iii) facial metrics calculation, which uses 20 facial landmarks to compute the metrics shown in Table 3 (cf. [39] for details). Finally, all metrics in pixels were normalized within every subject by dividing the values by the interlachrymal distance in pixels (measured as distance between the right corner of the left eye and the left corner of the right eye).

5. Analyses and Observations

To normalize for sex-specific differences in metrics (such as F0), we z-scored all metrics by sex group. Additionally, all metrics reported below (except speaking and articulation duration) were averaged across speech task type. A caveat to all the analyses presented here is the imbalance of sample size between the cohorts; also, future extensions to this work will need a larger sample size of the BUL and PRE cohorts to make robust and generalizable statistical claims.

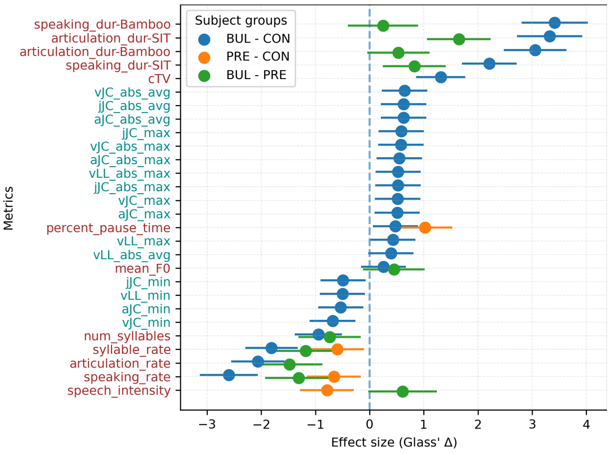


Figure 1: *Effect sizes of* ***acoustic*** *and* ***visual*** *metrics that show statistically significant differences at $p < 0.05$, shown with 95% confidence interval and ranked by the BUL–CON pair.*

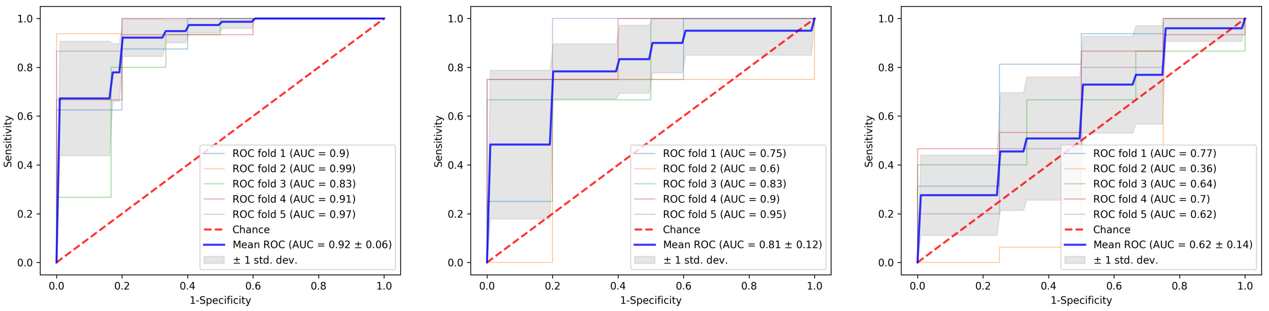
5.1. RQ1: Which metrics demonstrated statistically significant pairwise differences between controls, bulbar presymptomatic, and bulbar symptomatic pALS cohorts?

We conducted a non-parametric Kruskal-Wallis test for every acoustic and visual metric to identify those that showed a statistically significant difference between the cohorts. For all metrics with $p < 0.05$ a post-hoc analysis was done (again Kruskal-Wallis) between every group pair to find out which groups can be distinguished. Figure 1 shows effect size, measured as Glass' Δ [40], for all metrics that show statistically significant difference ($p < 0.05$) between different subject groups.[5]

In addition to the statistical tests, we conducted 5-fold cross-validation with logistic regression to investigate binary classification performances, and in turn sensitivities and specificities, of our aforementioned metrics in distinguishing the CON vs PRE (with applications to early diagnosis) and PRE vs BUL (progress monitoring) groups. We investigated using both the full feature set as well as feature selection with recursive feature elimination for classification and found that the latter performed better as the former method tends to overfit the data, given our small sample size. Receiver operating characteristics (ROC) curves for these classification experiments encapsulating sensitivities and specificities are presented in Figure 2. For the CON vs PRE case, we observed that the mean unweighted average recall (UAR) across 5 cross-validation folds was 0.63 ± 0.08 (significantly above chance), suggesting promising applications for early diagnosis. For the PRE vs BUL case (and therefore progress monitoring), the result was even better with 0.77 ± 0.05. The BUL vs CON case, unsurprisingly, produced the best results with 0.80 ± 0.08.

Looking at acoustic features, we found that timing measures (speaking and articulation duration and rate, PPT, syllable rate, cTV) exhibit strong differences between groups and that the effect sizes of these metrics are highest between the BUL group and the CON group. Mean F0 also showed a significant

[5] To investigate the extent to which sex skew towards females in the control group affects the results, we re-ran the analysis after randomly subsampling from the female control group to balance the cohorts. The results were similar to those observed in Figure 1, with the exception of (i) maximum values of all visual metrics and (ii) lower lip metrics, which did not show statistically significant effect sizes. However, given the smaller sample size in this analysis, we chose to describe results for the whole dataset in the rest of the paper.

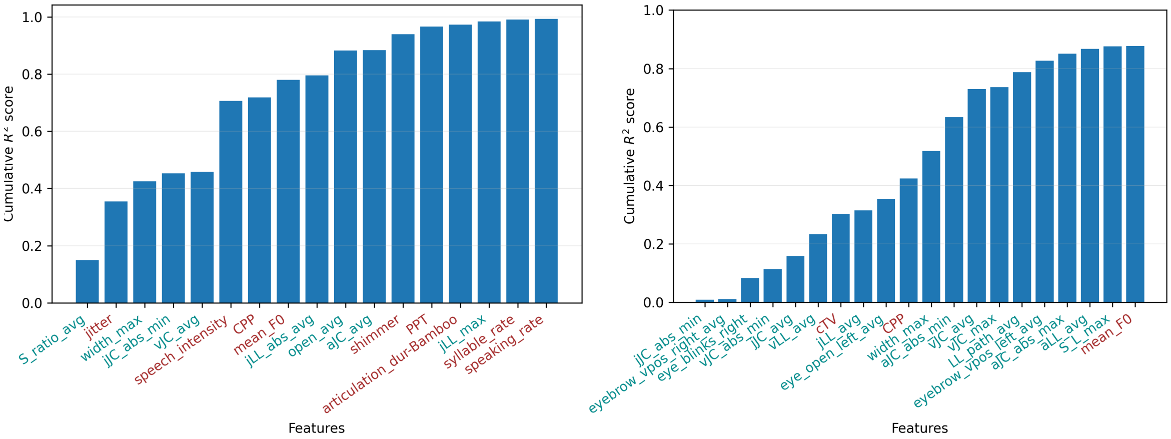


(a) *BUL vs CON*

(b) *BUL vs PRE*

(c) *PRE vs CON*

Figure 2: *ROC curves displaying the performance of binary classification with 5-fold crossvalidation for all group pairs.*



(a) *Final 17 features obtained for PRE group*

(b) *First 20 features (out of 23) for BUL group*

Figure 3: ***Acoustic*** *and* ***visual*** *features from the LASSO LARS regression path.*

difference with small effect sizes. For visual metrics, the results indicate that velocity, acceleration, and jerk measures are generally the best indicators for ALS. Additionally, while the jaw center (JC) seems to be more important than the lower lip (LL) for detecting ALS, further investigations are necessary to ensure that the difference between the JC and LL metrics is not due to a difference in facial landmark detection accuracy.

5.2. RQ2: Which metrics contribute the most toward predicting the ALSFRS-R score?

In a regression analysis, we investigated the predictive power of the extracted metrics with regard to the BUL and PRE cohorts. We employed a LASSO (least absolute shrinkage and selection operator) regression with the objective to predict the total ALSFRS-R score (implemented using least-angle regression (LARS) algorithm [41]). The algorithm is similar to forward stepwise regression, but instead of including features at each step, the estimated coefficients are increased in a direction equiangular to each one's correlations with the residual.

Figure 3a shows the final 17 features, in the order they were selected by the LASSO-LARS regression, on data from 19 PRE samples[6] along with the cumulative model R^2 at each step. We observe that both facial metrics (mouth opening and symmetry ratio, higher order statistics of jaw and lips) and acoustic metrics (particularly voice quality metrics such as jitter, shimmer, and mean F0) added useful predictive power to the model, suggesting that these might be useful in modeling severity in bulbar pre-symptomatic pALS.

For the BUL cohort, Figure 3b shows a slightly different set of 20 features obtained using LASSO-LARS (based on 26 participants). We observe that facial metrics (eye blinks and brow positions, in addition to higher order statistics of jaw and lips) add more predictive power than acoustic metrics (such as cTV, CPP and mean F0), suggesting that these might find utility in modeling severity in bulbar symptomatic pALS. These findings emphasize the benefits of a multimodal approach, which has also been shown in a similar study under controlled laboratory conditions [26], as well as the feasibility of utilizing remotely collected, non-invasive video-based measures.

6. Conclusions

Our findings demonstrate the utility of multimodal dialog technology for assisting early diagnosis and monitoring of pALS. Multiple automatically extracted acoustic (rate, duration, voicing) and visual (higher order statistics of the jaw and lip) speech metrics show significant promise in assisting with both early diagnosis of bulbar pre-symptomatic ALS vs healthy controls, as well as for progress monitoring in pALS. Moreover, using LASSO-LARS to model the relative contribution of these features in predicting the ALSFRS-R score highlights the utility of incorporating different acoustic and visual speech metrics for modeling severity in bulbar pre-symptomatic and bulbar symptomatic pALS. Future work will expand these analyses to more speakers and a more balanced age distribution across cohorts to ensure statistical robustness and generalizability of these trends.

[6] The number of samples in the classification and regression analyses differ from Table 1 because not all metrics were present for all subjects, either because system errors or the task was not performed correctly.

7. References

[1] S. Kumar, W. Nilsen *et al.*, "Mobile health: Revolutionizing healthcare through transdisciplinary research," *Computer*, vol. 46, no. 1, pp. 28–35, 2012.

[2] A. Bombaci, G. Abbadessa *et al.*, "Telemedicine for management of patients with amyotrophic lateral sclerosis through covid-19 tail," *Neurological Sciences*, pp. 1–5, 2020.

[3] R. Steven and M. Steinhubl, "Can mobile health technologies transform health care," *JAMA*, vol. 92037, no. 1, pp. 1–2, 2013.

[4] D. Suendermann-Oeft, A. Robinson *et al.*, "Nemsi: A multimodal dialog system for screening of neurological or mental conditions," in *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, 2019, pp. 245–247.

[5] D. DeVault, R. Artstein *et al.*, "Simsensei kiosk: A virtual human interviewer for healthcare decision support," in *Proceedings of the International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, Paris, France, 2014 May.

[6] C. Lisetti, R. Amini, and U. Yasavu, "Now all together: Overview of virtual health assistants emulating face-to-face health interview experience," *KI-Künstliche Intelligenz*, vol. 29, March 2015.

[7] J. M. Cedarbaum, N. Stambler *et al.*, "The alsfrs-r: a revised als functional rating scale that incorporates assessments of respiratory function," *Journal of the neurological sciences*, vol. 169, no. 1-2, pp. 13–21, 1999.

[8] D. Majoor-Krakauer, P. Willems, and A. Hofman, "Genetic epidemiology of amyotrophic lateral sclerosis," *Clinical genetics*, vol. 63, no. 2, pp. 83–101, 2003.

[9] A. Al-Chalabi and O. Hardiman, "The epidemiology of als: a conspiracy of genes, environment and time," *Nature Reviews Neurology*, vol. 9, no. 11, p. 617, 2013.

[10] S. Paganoni, E. A. Macklin *et al.*, "Diagnostic timelines and delays in diagnosing amyotrophic lateral sclerosis (als)," *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, vol. 15, no. 5-6, pp. 453–456, 2014.

[11] A. Chiò, A. Gauthier *et al.*, "A cross sectional study on determinants of quality of life in als," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 75, no. 11, pp. 1597–1601, 2004.

[12] K. Joubert, J. Bornman, and E. Alant, "Speech intelligibility and marital communication in amyotrophic lateral sclerosis: an exploratory study," *Communication Disorders Quarterly*, vol. 33, no. 1, pp. 34–41, 2011.

[13] Y. Yunusova, J. R. Green *et al.*, "Speech in als: longitudinal changes in lips and jaw movements and vowel acoustics," *Journal of medical speech-language pathology*, vol. 21, no. 1, 2013.

[14] K. M. Allison, Y. Yunusova *et al.*, "The diagnostic utility of patient-report and speech-language pathologists' ratings for detecting the early onset of bulbar symptoms due to ALS," *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, vol. 18, no. 5-6, pp. 358–366, August 2017.

[15] P. Gomez, D. Palacios *et al.*, "Articulation acoustic kinematics in als speech," in *2017 International Conference and Workshop on Bioinspired Intelligence (IWOBI)*. IEEE, 2017, pp. 1–6.

[16] R. Norel, M. Pietrowicz *et al.*, "Detection of amyotrophic lateral sclerosis (als) via acoustic analysis," *Proc. Interspeech 2018*, pp. 377–381, 2018.

[17] A. Bandini, J. R. Green *et al.*, "Kinematic features of jaw and lips distinguish symptomatic from presymptomatic stages of bulbar decline in amyotrophic lateral sclerosis," *Journal of Speech, Language, and Hearing Research*, vol. 61, no. 5, 2018.

[18] B. J. Perry, R. Martino *et al.*, "Lingual and jaw kinematic abnormalities precede speech and swallowing impairments in als," *Dysphagia*, vol. 33, no. 6, pp. 840–847, 2018.

[19] G. M. Stegmann, S. Hahn *et al.*, "Early detection and tracking of bulbar changes in als via frequent and remote speech analysis," *NPJ digital medicine*, vol. 3, no. 1, pp. 1–5, 2020.

[20] A. Wisler, K. Teplansky *et al.*, "The effects of symptom onset location on automatic amyotrophic lateral sclerosis detection using the correlation structure of articulatory movements," *Journal of Speech, Language, and Hearing Research*, 2021. [Online]. Available: https://pubs.asha.org/doi/abs/10.1044/2020_JSLHR-20-00288

[21] P. Rong, Y. Yunusova *et al.*, "A speech measure for early stratification of fast and slow progressors of bulbar amyotrophic lateral sclerosis: lip movement jitter," *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, vol. 21, no. 1-2, pp. 34–41, 2020.

[22] ——, "Predicting speech intelligibility decline in amyotrophic lateral sclerosis based on the deterioration of individual speech subsystems," *PloS one*, vol. 11, no. 5, p. e0154971, 2016.

[23] J. Wang, P. V. Kothalkar *et al.*, "Automatic prediction of intelligible speaking rate for individuals with als from speech acoustic and articulatory samples," *International journal of speech-language pathology*, vol. 20, no. 6, pp. 669–679, 2018.

[24] T. Makkonen, H. Ruottinen *et al.*, "Speech deterioration in amyotrophic lateral sclerosis (als) after manifestation of bulbar symptoms," *International journal of language & communication disorders*, vol. 53, no. 2, pp. 385–392, 2018.

[25] E. M. Wilson, M. Kulkarni *et al.*, "Detecting bulbar motor involvement in als: Comparing speech and chewing tasks," *International Journal of Speech-Language Pathology*, vol. 21, no. 6, 2019.

[26] A. Wisler, K. Teplansky *et al.*, "Speech-based estimation of bulbar regression in amyotrophic lateral sclerosis," in *Proceedings of the Eighth Workshop on Speech and Language Processing for Assistive Technologies*, 2019, pp. 24–31.

[27] C. Barnett, J. R. Green *et al.*, "Reliability and validity of speech & pause measures during passage reading in als," *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, vol. 21, no. 1-2, pp. 42–50, 2020.

[28] J. D. Berry, S. Paganoni *et al.*, "Design and results of a smartphone-based digital phenotyping study to quantify als progression," *Annals of Clinical and Translational Neurology*, vol. 6, no. 5, pp. 873–881, February 2019.

[29] 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," *Journal of Voice*, vol. 11, no. 2, pp. 222–231, 1997.

[30] 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.

[31] M. Novotny, J. Melechovsky *et al.*, "Comparison of automated acoustic methods for oral diadochokinesis assessment in amyotrophic lateral sclerosis," *Journal of Speech, Language, and Hearing Research*, vol. 63, no. 10, pp. 3453–3460, 2020.

[32] C. Agurto, M. Pietrowicz *et al.*, "Analyzing progression of motor and speech impairment in als," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2019, pp. 6097–6102.

[33] P. Rong, "Automated acoustic analysis of oral diadochokinesis to assess bulbar motor involvement in amyotrophic lateral sclerosis," *Journal of Speech, Language, and Hearing Research*, 2020.

[34] P. Boersma and V. Van Heuven, "Speak and unspeak with praat," *Glot International*, vol. 5, no. 9/10, pp. 341–347, 2001.

[35] Y. T. Yeung, K. H. Wong, and H. Meng, "Improving automatic forced alignment for dysarthric speech transcription," in *Sixteenth Annual Conference of the International Speech Communication Association*, 2015.

[36] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, San Diego, USA, June 2005.

[37] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Columbus, USA, June 2014.

[38] R. Gross, I. Matthews *et al.*, "Multi-pie," in *Proceedings of the International Conference on Automatic Face and Gesture Recognition (FG)*, vol. 28, no. 5, 2010, pp. 807–813.

[39] M. Neumann, O. Roessler *et al.*, "On the utility of audiovisual dialog technologies and signal analytics for real-time remote monitoring of depression biomarkers," in *Proceedings of the First Workshop on Natural Language Processing for Medical Conversations*, 2020, pp. 47–52.

[40] G. V. Glass, B. McGaw, and M. L. Smith, *Meta-analysis in social research*. Sage Publications, Incorporated, 1981.

[41] B. Efron, T. Hastie *et al.*, "Least angle regression," *Annals of statistics*, vol. 32, no. 2, pp. 407–499, 2004.