NON-INVASIVE SPEECH ANALYSIS FOR DYSPHAGIA DETECTION IN AMYOTROPHIC LATERAL SCLEROSIS

F. Pierotti[1], D. Gasperini[2], S. Capobianco[3], L. Becattini[4], F. Bianchi[5], A. Nacci[3], A. Santoro[3], B. Fattori[3], G. Siciliano[4], A. Bandini[1,2,6]

[1] The BioRobotics Institute and Department of Excellence in Robotics and AI, Scuola Superiore Sant'Anna, Pisa, Italy
[2] Health Science Interdisciplinary Research Center, Scuola Superiore Sant'Anna, Pisa, Italy
[3] ENT, Audiology and Phoniatrics Unit, Pisa University Hospital, Pisa, Italy
[4] Neurology Unit, Department of Clinical and Experimental Medicine, University of Pisa, Pisa, Italy
[5] Neurology Unit, Department of Neuroscience, Azienda Ospedaliera Universitaria Pisana, Pisa, Italy
[6] KITE Research Institute, University Health Network, Toronto, ON, Canada
francesco.pierotti@santannapisa.it; daniela.gasperini@santannapisa.it; silviacapobianco.md@gmail.com; lu.becattini@gmail.com; francicr86@gmail.com; a.nacci@med.unipi.it; asantoro_dr@hotmail.com; bruno.fattori@unipi.it; gabriele.siciliano@unipi.it; andrea.bandini@santannapisa.it

Abstract: **The acoustic and kinematic analyses of speech represent non-invasive and cost-effective tools to support clinicians in the assessment of bulbar dysfunction, which is particularly challenging in amyotrophic lateral sclerosis (ALS). In this study, we applied multimodal speech analysis to identify suitable acoustic and kinematic biomarkers potentially able to detect bulbar impairment, particularly the presence of dysphagia in its early stages. Our results revealed clear distinctions between dysphagic and non-dysphagic individuals with ALS during connected speech, with the third formant and speech timing metrics being the most significant features ($p < 0.001$). Our findings suggest that a reduced F3 and an increased duration of pauses during connected speech may serve as important biomarkers for detecting the onset of dysphagia in ALS.**

Keywords: **Amyotrophic lateral sclerosis, bulbar impairment, dysphagia assessment, speech biomarkers, audio-video**

I. INTRODUCTION

Amyotrophic lateral sclerosis (ALS) is a rare neurodegenerative disease, with a survival time ranging from 3 to 5 years, depending on several factors, including age, type of onset (spinal or bulbar), and disease progression rate [1]. ALS is characterized by the degeneration of the motor neurons, causing paralysis of the limbs, trunk, and orofacial muscles.

ALS is frequently accompanied by bulbar impairments, including speech and swallowing disorders [2], which are hallmarks of decreased quality of life, shorter survival time, and overall progression [3]. Among bulbar symptoms, dysarthria and dysphagia are the two most prevalent and clinically significant. Specifically, dysarthria is a motor speech disorder caused by loss of articulatory strength and control [4]. Dysphagia is associated with difficulty in safely and efficiently swallowing food and liquids, predominantly affecting the oral and pharyngeal phases of swallowing [5]. Dysphagia reduces the ability to eat normally, increasing the risk of malnutrition, aspiration pneumonia, and death [6]. Therefore, early and accurate detection of dysphagia and its decline is fundamental to gaining insights into disease progression, preventing severe clinical complications, and rapidly accessing tailored therapeutic strategies [7]. Given that the anatomical structures responsible for speech production and swallowing largely overlap [8], dysarthria and dysphagia frequently co-occur in individuals with ALS [9].

The gold standard instrumental assessment methods for swallowing are the fiberoptic endoscopic evaluation of swallowing (FEES) and the videofluoroscopic swallowing study (VFSS) [10]. However, these methods require specialized personnel, in-clinic visits, and can cause discomfort, pain, and even severe complications, such as laryngospasm [11], or require radiation exposure.

In recent years, the acoustic and video analyses of speech production have emerged as promising pathways towards more accessible methods of bulbar function assessment in ALS [12], [13]. Preliminary results have shown that acoustic and kinematic features of speech and orofacial functions may be correlated with bulbar decline in ALS [14], [15]. Recently, a growing number of studies have applied voice analysis for detecting dysphagia in various neurological conditions [16], [17], identifying acoustic parameters that may provide useful insight into dysphagia, including frequency perturbation (e.g., relative average perturbation), amplitude perturbations (e.g., shimmer), and harmonic-to-noise ratio [18]. However, these studies mainly rely on sustained phonations (e.g., vowel /*a*/) or syllable repetitions and do not exploit kinematic analysis of

Referee List (DOI 10.36253/fup_referee_list)
FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)
Lorenzo Frassineti, Antonio Lanatà, Claudia Manfredi (edited by), *Models and Analysis of Vocal Emissions for Biomedical Applications. 14th International Workshop, December, 16-17, 2025*, © 2025 Author(s), CC BY 4.0, published by Firenze University Press, ISBN 979-12-215-0821-5, DOI 10.36253/979-12-215-0821-5

speech. Also, to the best of our knowledge, none of the previous studies investigated dysphagia through speech acoustics in ALS.

We fill this gap by exploiting acoustic and kinematic analysis of speech to objectively describe dysphagia in ALS. Our goal is to identify objective and non-invasive metrics capable of detecting early signs of bulbar impairment, specifically swallowing difficulties.

II. METHODS

A. Data collection

Fifteen individuals with ALS, aged 52-81 years (mean = 64.2 ± 8.5 years, 7 females, 4 bulbar onset, total ALSFRS-r = 39.5 ± 5.3), were recruited at the Neurology Unit of the Pisa University Hospital. The inclusion criteria required participants to have been diagnosed with ALS, with symptom onset within 18 months from the screening visit, without a previous history of speech, swallowing, or oro-facial impairments.

Each participant was asked to perform several speech tasks, including sustained phonation of vowels, diadochokinetic repetitions of /pa/ and /pataka/, a connected speech task involving picture description, and reading aloud a passage , which was proven to provide informative metrics of ALS bulbar dysfunctions [14]. The passage was the Italian phonetically balanced text "*Il deserto*" [19].

Speech signals were recorded at 44.1 kHz using two high-quality microphones (Sennheiser MKE 200 for voice recording and Rode SmartLav+ for environmental noise), while orofacial movements were simultaneously video-recorded via a front-facing webcam (Razer Kiyo 4MP) positioned approximately 50 cm from the participant's face. A custom Python GUI was developed to streamline the recordings, ensuring synchronization between audio and video streams. Data collection was carried out in a quiet room of the Cisanello hospital in Pisa, and participants were asked to stay seated during the test.

Clinical assessments included: ALS functional rating scale revised (ALSFRS-R), Penn Upper Notor Neuron score (PUMNS), Medical Research Council (MRC) scale for the assessment of muscle strength, and instrumental assessment via FEES. In this study, we considered the pooling score [20] obtained from FEES exam, which indicates the level of bolus residue and swallowing dynamics.

B. Audio and video analyses

For each frame of the video recordings, the facial region of the participant was identified using the single-shot scale-invariant face detection model [21]. Subsequently, the 3D coordinates of 68 facial landmarks were extracted using the face alignment network [22], therefore, estimating the positions of the mouth, jawline, nose, eyes, and eyebrows. From these landmarks, we computed a set of features describing the range of motion, speed, symmetry, and shape of different parts of the face [2]. Specifically, we compute cumulative path travelled by the lower lip and by the jaw, the mean value and the range of the mouth area with respect to the rest position, the maximum and minimum mouth speed, lower lip, and jaw, the absolute difference between right and left mouth areas, the correlation between right and left mouth corners movements, the mean value of mouth eccentricity, and the range of mouth eccentricity. All the features were normalized by the intercanthal distance to ensure consistency across frames.

From the audio recordings, we extracted several acoustic features related to phonation, prosody, articulation, speech timings, and intelligibility. These features were computed using Python programming language, by exploiting Parselmouth [23], a Python library implementing Praat functions [24], and MATLAB 2024b. First, we transcribed the audio by using the WhisperX automated speech recognition (ASR) model [25], which provided both the transcription of speech recognized and word-level timestamps, with its own level of confidence. Therefore, we derived several speech timing metrics, such as inter-word interval, speech and articulation rate, articulation entropy, pause percentage, confidence score metrics, and word error rate. These features reflect several aspects of speech timings, velocity, and articulatory difficulty. Additionally, to face the difficult recognition of ASR in the case of a dysarthric speaker, we identified speech activity by applying a voice activity detection (VAD) algorithm [26], and we used these timings to compute metrics of speech/silent timings.

We also utilized Praat functions to extract additional acoustic features, including metrics of fundamental frequency (F0), jitter, shimmer, harmonics-to-noise ratio (HNR), and cepstral coefficients, as well as first, second, and third formants (F1, F2, F3, respectively). We followed the Praat guidelines [24] to impose minimum and maximum pitch frequencies and formants depending on the participant's sex. Furthermore, we derived articulatory features from vowel formant measures. In particular, we identified words with corner vowels from the ASR transcript, and then we aligned them at the phoneme level using the Wav2Vec2 force alignment (FA) algorithm. Then, we extracted F1, F2, and F3 centered at the vowel estimated time. Finally, we computed the vowel space area (VSA), formant centralization ratio (FCR), and the ratio between F2 of /*i*/ and F2 of /*u*/.

Overall, 132 features were extracted, with 106 coming from audio and 26 from video.

C. Statistical analysis

Participants were divided into two groups based on the pooling score with a solid-dry bolus [20]: participants with a pooling score lower than 6 (non-dysphagic group); and participants with a pooling score higher than or equal to 6 (i.e., associated with mild, moderate, or severe dysphagia) [20]. We chose to investigate the solid-dry score, since dysphagia for solids is generally one of the first symptoms of swallowing impairment in ALS, due to reduced tongue propulsion resulting from lingual muscle atrophy [27].

We applied the maximum relevance minimum redundancy (mRMR) algorithm to rank the features based on the outcome obtained with the pooling score. For the statistical analysis, we applied the Shapiro-Wilk test to assess normality of the data distribution, and we applied a t-test or a Mann-Whitney test to find any statistically significant difference for each feature across groups.

III. RESULTS

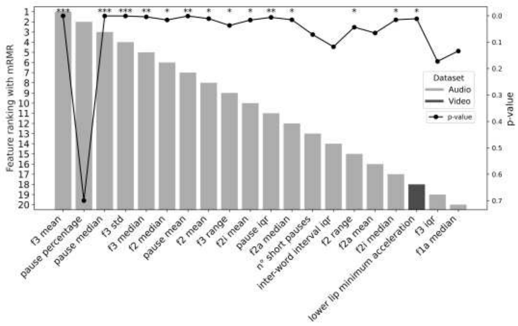

Fig. 1: top-ranked acoustic and kinematic features with the mRMR algorithm and their p-value.

Among all the participants, 14 completed all assessments, whereas one withdrew before the audio-video recording. Based on the pooling score, 5 participants belonged to the first group (no dysphagia, 66.40 ± 10.67 years, 5 females, 2 bulbar onset), with a duration of the recording equal to 131.02 ± 59.45 seconds, while the remaining 9 were included in the second group (62.00 ± 6.90 years, 2 females, 2 bulbar onset), having a recording duration of 110.65 ± 58.77 seconds. Fig. 1 shows the features ranking and the highest statistical difference. Notably, the number of statistically different acoustic features is much higher than the kinematic features, which are significant only for the minimum acceleration of the lower lip. Interestingly, the most important features according to the mRMR ranking are not necessarily those showing the strongest statistical significance between groups.

The results in Fig. 2 showed that highly significant differences exist between dysphagic and non-dysphagic groups for F3 metrics, with individuals with dysphagia showing higher values. In contrast, non-dysphagic participants showed statistically significantly lower values for pause time metrics.

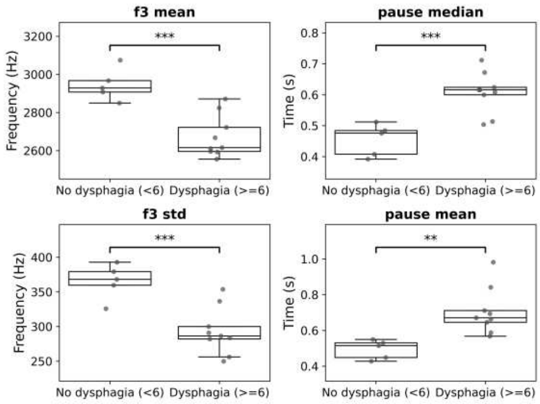

Fig. 2: boxplots of the features with the largest statistically significant differences.

IV. DISCUSSION

In this study, we analyzed audio and video recordings of ALS individuals during a speech task. We aimed to identify any acoustic and kinematic features statistically different between participants with and without dysphagia. The acoustic and kinematic features were directly extracted, considering the entire audio, which differs from all the previous studies focused predominantly on phonation [17]. Metrics of F3 showed the largest statistical difference and importance, with reduced mean values and standard deviations in the group with dysphagia.

This behavior of F3, associated with the position of the soft palate and velopharyngeal closure, aligns with the finding of [28]. Although the association between velopharyngeal deficit and solid-dry dysphagia is not yet fully understood, a reduced F3 could indicate an incomplete velopharyngeal closure, potentially contributing to less efficient swallowing. Moreover, we found pause duration metrics to be statistically different, with increased values in the case of dysphagia. These results reflect the findings of studies on bulbar impairment [29], [30], which highlighted higher pause time and lower speech rate in cases of speech impairment, especially with the bulbar onset and with the disease progression. Therefore, our work contributes to supporting and promoting future research on the identification of features that may reveal difficulties in swallowing before they become clinically obvious.

This study is limited by the small sample size and the restricted set of analyzed speech tasks. Moreover, since the audio was analyzed entirely, some important speech acoustic metrics may be averaged and are not easy to interpret. However, connected speech is a more comprehensive task, which includes also more complex tasks than simple vowel phonation, allowing the analysis of more complete articulatory movements.

Future work should explore other tasks, speech parameters, more detailed analysis, and a larger dataset. Additionally, a longitudinal study may confirm these findings, allowing for making predictions of the bulbar decline.

V. CONCLUSION

This study highlights that acoustic and kinematic analysis of connected speech can provide valuable biomarkers for the assessment of dysphagia in individuals with ALS. Measures of F3 and pause timings emerged as the most sensitive indicators. These findings aim to support research in the use of speech-based approaches as a tool to support clinicians in the assessment and monitoring of swallowing impairment, offering a more accessible, non-invasive, and cost-effective solution to traditional assessment methods.

ACKNOLEDGEMENTS

The financial support of AriSLA – Fondazione Italiana di ricerca per la SLA is acknowledged (Project MIMOSA - Multimodal Intelligent Methods for Orofacial and Speech Assessment to predict ALS bulbar decline)

REFERENCES

[1] A. Chiò *et al.*, ‘Prognostic factors in ALS: A critical review’, *Amyotroph. Lateral Scler.*, vol. 10, no. 5–6, pp. 310–323, Jan. 2009, doi: 10.3109/17482960802566824.

[2] A. Bandini, J. R. Green, B. Taati, S. Orlandi, L. Zinman, and Y. Yunusova, ‘Automatic Detection of Amyotrophic Lateral Sclerosis (ALS) from Video-Based Analysis of Facial Movements: Speech and Non-Speech Tasks’, in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, Xi’an: IEEE, May 2018, pp. 150–157. doi: 10.1109/FG.2018.00031.

[3] S. H. Felgoise, V. Zaccheo, J. Duff, and Z. Simmons, ‘Verbal communication impacts quality of life in patients with amyotrophic lateral sclerosis’, *Amyotroph. Lateral Scler. Front. Degener.*, vol. 17, no. 3–4, pp. 179–183, May 2016, doi: 10.3109/21678421.2015.1125499.

[4] H. P. Rowe, S. Shellikeri, Y. Yunusova, K. V. Chenausky, and J. R. Green, ‘Quantifying articulatory impairments in neurodegenerative motor diseases: A scoping review and meta-analysis of interpretable acoustic features’, *Int. J. Speech Lang. Pathol.*, vol. 25, no. 4, pp. 486–499, Jul. 2023, doi: 10.1080/17549507.2022.2089234.

[5] A. Sasegbon and S. Hamdy, ‘The anatomy and physiology of normal and abnormal swallowing in oropharyngeal dysphagia’, *Neurogastroenterol. Motil.*, vol. 29, no. 11, p. e13100, Nov. 2017, doi: 10.1111/nmo.13100.

[6] P. E. Marik and D. Kaplan, ‘Aspiration Pneumonia and Dysphagia in the Elderly’, *Chest*, vol. 124, no. 1, pp. 328–336, Jul. 2003, doi: 10.1378/chest.124.1.328.

[7] K. M. Allison, Y. Yunusova, T. F. Campbell, J. Wang, J. D. Berry, and J. R. Green, ‘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. Front. Degener.*, vol. 18, no. 5–6, pp. 358–366, Jul. 2017, doi: 10.1080/21678421.2017.1303515.

[8] J. R. Duffy, ‘Motor Speech Disorders: Clues to Neurologic Diagnosis’.

[9] B. J. Wang, F. L. Carter, and K. W. Altman, ‘Relationship between Dysarthria and Oral-Oropharyngeal Dysphagia: The present evidence’, *Ear. Nose. Throat J.*, p. 014556132095164, Oct. 2020, doi: 10.1177/0145561320951647.

[10] D. K.-H. Lai *et al.*, ‘Computer-aided screening of aspiration risks in dysphagia with wearable technology: a Systematic Review and meta-analysis on test accuracy’, *Front. Bioeng. Biotechnol.*, vol. 11, p. 1205009, Jun. 2023, doi: 10.3389/fbioe.2023.1205009.

[11] A. Nacci, F. Ursino, R. La Vela, F. Matteucci, V. Mallardi, and B. Fattori, ‘Fiberoptic endoscopic evaluation of swallowing (FEES): proposal for informed consent’, *Acta Otorhinolaryngol. Ital. Organo Uff. Della Soc. Ital. Otorinolaringol. E Chir. Cerv.-facc.*, vol. 28, no. 4, pp. 206–211, Aug. 2008.

[12] J. R. Green *et al.*, ‘Bulbar and speech motor assessment in ALS: Challenges and future directions’, *Amyotroph. Lateral Scler. Front. Degener.*, vol. 14, no. 7–8, pp. 494–500, Dec. 2013, doi: 10.3109/21678421.2013.817585.

[13] L. E. R. Simmatis, J. Robin, M. J. Spilka, and Y. Yunusova, ‘Detecting bulbar amyotrophic lateral sclerosis (ALS) using automatic acoustic analysis’, *Biomed. Eng. Online*, vol. 23, no. 1, p. 15, Feb. 2024, doi: 10.1186/s12938-023-01174-z.

[14] M. Neumann, H. Kothare, and V. Ramanarayanan, ‘Multimodal speech biomarkers for remote monitoring of ALS disease progression’, *Comput. Biol. Med.*, vol. 180, p. 108949, Sep. 2024, doi: 10.1016/j.compbiomed.2024.108949.

[15] Y. Yunusova, J. R. Green, M. J. Lindstrom, L. J. Ball, G. L. Pattee, and L. Zinman, ‘Kinematics of disease progression in bulbar ALS’, *J. Commun. Disord.*, vol. 43, no. 1, pp. 6–20, Jan. 2010, doi: 10.1016/j.jcomdis.2009.07.003.

[16] J. S. Ryu, S. R. Park, and K. H. Choi, ‘Prediction of Laryngeal Aspiration Using Voice Analysis’:, *Am. J. Phys. Med. Rehabil.*, vol. 83, no. 10, pp. 753–757, Oct. 2004, doi: 10.1097/01.PHM.0000140798.97706.A5.

[17] K. W. Dos Santos *et al.*, ‘Using Voice Change as an Indicator of Dysphagia: A Systematic Review’, *Dysphagia*, vol. 37, no. 4, pp. 736–748, Aug. 2022, doi: 10.1007/s00455-021-10319-y.

[18] I. Hwang, J.-M. Kim, J. S. Ryu, and K. Lee, ‘Voice-Based Dysphagia Detection: Leveraging Self-Supervised Speech Representation’, in *Interspeech 2025*, ISCA, pp. 5683–5687. doi: 10.21437/Interspeech.2025-761.

[19] A. Romano, U. Cesari, M. Mignano, O. Schindler, and I. Vernero, ‘LA QUALITÀ DELLA VOCE’, presented at the Atti dell’VIII Convegno dell’associazione Italiana di Scienze della Voce, Roma, Jan. 2012, pp. 1–34.

[20] D. Farneti *et al.*, ‘The Pooling-score (P-score): inter- and intra-rater reliability in endoscopic assessment of the severity of dysphagia’, *Acta Otorhinolaryngol. Ital. Organo Uff. Della Soc. Ital. Otorinolaringol. E Chir. Cerv.-facc.*, vol. 34, no. 2, pp. 105–110, Apr. 2014.

[21] S. Zhang, X. Zhu, Z. Lei, H. Shi, X. Wang, and S. Z. Li, ‘S^3FD: Single Shot Scale-Invariant Face Detector’, in *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice: IEEE, Oct. 2017, pp. 192–201. doi: 10.1109/ICCV.2017.30.

[22] A. Bulat and G. Tzimiropoulos, ‘How Far are We from Solving the 2D & 3D Face Alignment Problem? (and a Dataset of 230,000 3D Facial Landmarks)’, in *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice: IEEE, Oct. 2017, pp. 1021–1030. doi: 10.1109/ICCV.2017.116.

[23] 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.

[24] Boersma, Paul, ‘Praat, a system for doing phonetics by computer’ *Glot International,* 2001, 5:9/10, 341-345..

[25] M. Bain, J. Huh, T. Han, and A. Zisserman, ‘WhisperX: Time-Accurate Speech Transcription of Long-Form Audio’, in *INTERSPEECH 2023*, ISCA, Aug. 2023, pp. 4489–4493. doi: 10.21437/Interspeech.2023-78.

[26] Jongseo Sohn, Nam Soo Kim, and Wonyong Sung, ‘A statistical model-based voice activity detection’, *IEEE Signal Process. Lett.*, vol. 6, no. 1, pp. 1–3, Jan. 1999, doi: 10.1109/97.736233.

[27] D. C. Wolf, ‘Dysphagia’, in *Clinical Methods: The History, Physical, and Laboratory Examinations*, 3rd ed., H. K. Walker, W. D. Hall, and J. W. Hurst, Eds., Boston: Butterworths, 1990. Accessed: Sep. 20, 2025. [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK408/

[28] Y. Liang, F. A. Numan, K. Li, and G. Liao, ‘Spectrum analysis of Chinese vowels formant in patients with tongue carcinoma underwent hemiglossectomy’, *Int. J. Clin. Exp. Med.*, vol. 8, no. 2, pp. 2867–2873, 2015.

[29] 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, p. e0154971, May 2016, doi: 10.1371/journal.pone.0154971.

[30] M. Eshghi *et al.*, ‘Rate of speech decline in individuals with amyotrophic lateral sclerosis’, *Sci. Rep.*, vol. 12, no. 1, p. 15713, Sep. 2022, doi: 10.1038/s41598-022-19651-1.