Interspeech 2024
1-5 September 2024, Kos, Greece

How Consistent are Speech-Based Biomarkers in Remote Tracking of ALS Disease Progression Across Languages? A Case Study of English and Dutch

Hardik Kothare[1], Michael Neumann[1], Cathy Zhang[1], Jackson Liscombe[1], Jordi W J van Unnik[2], Lianne C M Botman[2], Leonard H van den Berg[2], Ruben P A van Eijk[2], Vikram Ramanarayanan[1,3]

[1]Modality.AI, Inc., San Francisco, CA, USA [2]University Medical Center Utrecht, Utrecht, Netherlands [3]University of California, San Francisco, San Francisco, CA, USA

hardik.kothare@modality.ai

Abstract

Previous work has demonstrated the utility of speech-based digital biomarkers for remotely tracking longitudinal progression in people with Amyotrophic Lateral Sclerosis (pALS). Here, we investigate the responsiveness of these biomarkers across languages for consistency. We collected audiovisual data using a cloud-based multimodal dialogue platform, where pALS interacted with a virtual guide to perform several speaking exercises. We automatically extracted speech, linguistic and orofacial metrics from 143 English-speaking pALS (36 bulbar onset, 107 non-bulbar onset) and 26 Dutch-speaking pALS (10 bulbar, 16 non-bulbar onset). We used growth curve models to estimate the trajectory of these metrics over time. We observe that for most of these metrics, English-speaking pALS and Dutch-speaking pALS follow similar trajectories, i.e. the slopes are not statistically different from each other, demonstrating the potential of such speech-based biomarkers for remote monitoring across languages.

Index Terms: multimodal digital biomarkers, amyotrophic lateral sclerosis, remote patient monitoring

1. Introduction

Amyotrophic Lateral Sclerosis (ALS) or motor neuron disease is a neurodegenerative disorder in which the degeneration of upper and lower motor neurons leads to muscle weakness and paralysis [1]. Median survival time from disease onset ranges from 20 to 48 months [2]. Approximately 30% of people with ALS (pALS) present with bulbar onset, marked by rapid deterioration in speech and swallowing abilities [3]. The remaining majority present with non-bulbar onset, which manifests as muscle wasting in the limbs and torso [4]. However, a significant portion of those with non-bulbar onset eventually develop bulbar symptoms as the disease advances [5]. This rapid decline of bulbar function makes it possible to use speech-based objective biomarkers to monitor and track pALS. Indeed, objective speech and facial kinematic measures have shown utility in early detection of bulbar symptoms [6, 7, 8, 9, 10, 11, 12]. Eshghi et al. [13] showed that speaking rate and speech intelligibility can predict speech loss based on pre-defined thresholds and that these objective speech biomarkers are more responsive to functional decline than patient-reported ALSFRS-R scores. Yunusova et al. [14] reported that changes in kinematics of the jaw and lips are detectable prior to changes in vowel acoustics and speech intelligibility. Stegmann et al. [15] demonstrated that disease progression in bulbar onset and non-bulbar onset pALS can be predicted using speaking rate and articulatory precision through data collected remotely via a mobile application.

Multimodal speech-based biomarkers also show great potential in tracking the progression of bulbar decline in pALS [16, 14, 15, 13, 17]. Deployment of technology that is capable of collecting such multimodal speech-based digital biomarkers remotely shows promise in cost-effective and geographically-distributed clinical trials [18]. However, are speech-based biomarkers consistent in remote tracking of disease progression in pALS speaking different languages? Prior work suggests that certain characteristics of dysarthric speech in neurodegenerative disorders are language independent [19, 20, 21]. In this work we ask the following research questions using English and Dutch as a case study:

1. Of the speech-based biomarkers that show statistically significant differences between bulbar onset and non-bulbar onset English-speaking pALS, which metrics show consistent differences in Dutch-speaking pALS cohorts?
2. Is the rate of change of biomarker values over time consistent across English-speaking and Dutch-speaking pALS?

To our knowledge, this is the first study that systematically analyses multimodal speech biomarkers collected in pALS via a dialogue-based remote assessment platform across multiple languages.

2. Data

Data was collected from 143 English-speaking pALS and 26 Dutch-speaking pALS (see Table 1) through two ongoing studies in collaboration with EverythingALS and the Peter Cohen Foundation (English-speaking pALS) and the University Medical Center Utrecht (Dutch-speaking pALS). Both studies were approved by Institutional Review Boards[1]. Audiovisual data was collected using the Modality platform, a cloud-based multimodal dialogue system in which a virtual guide, Tina, engages participants in a semi-structured conversation to elicit speech and facial behaviours [22, 23]. The following tasks were included in both the English and the Dutch protocol: (a) read speech (sentence intelligibility test (SIT); Reading Passage (RP; English: Bamboo passage 99 words, Dutch: De Auto passage 103 words), (b) oral diadochokinesis (DDK, repeating the syllables /pɑ/, /tɑ/ and /kɑ/ in rapid succession), (c) single breath counting (SBC), and (d) free speech in form of a picture description task (PD).

3. Methods

Relevant speech acoustic, linguistic and orofacial metrics were automatically extracted from audio data collected using these tasks. Speech metrics were extracted using Praat (v6.2.17) [24] and the Montreal Forced Aligner (v2.0.0.a22) [25]. Facial video metrics were derived from facial landmarks generated using MediaPipe Face Mesh [26]. MediaPipe Face Detection, which is based on BlazeFace [27] is used to determine the (x,

[1]Advarra and Medical Research Ethics Committee (MREC) NedMec, respectively

10.21437/Interspeech.2024-1390

Table 1: *Participant demographics; age, number of sessions and time span reported as mean (standard deviation)*

		F	M	Age in years	#Sessions per participant	Time span in months
English-speaking	Bulbar onset	18	18	62.83 (8.23)	15.97 (18.48)	7.86 (8.18)
	Non-bulbar onset	52	55	62.02 (8.02)	25.93 (25.61)	13.63 (11.28)
Dutch-speaking	Bulbar onset	4	6	67.85 (3.69)	4 (3.62)	3.3 (3.74)
	Non-bulbar onset	2	14	67.84 (7.56)	5.31 (3.59)	4.5 (3.71)

Table 2: *Overview of extracted metrics. For facial metrics, functionals (minimum, maximum, average) are applied to produce one value across all video frames of an utterance. Facial metrics are measured in pixels and are normalized by dividing them by the intercaruncular distance (distance between inner corners of the eyes) for each participant. *specific to DDK task*

	Domain	**Exemplar Metrics**
Speech	Energy	shimmer (%), intensity (dB), signal-to-noise ratio (dB)
	Timing	speaking and articulation duration (sec.), articulation and speaking rate (WPM), percent pause time (PPT, %), canonical timing agreement (CTA, %), cycle-to-cycle temporal variability* (cTV, sec.), syllable rate* (syl./sec.), # of syllables*
	Voice quality	cepstral peak prominence (CPP, dB), harmonics-to-noise ratio (HNR, dB)
	Frequency	mean, max., min. fundamental frequency F0 (Hz), jitter (%)
Linguistic	Lexico-semantic	word count, percentage of content words, noun rate, verb rate, pronoun rate, noun-to-verb ratio, noun-to-pronoun ratio, closed class word ratio, idea density
Orofacial	Mouth (distances)	lip aperture/opening, lip width, mouth surface area, mean symmetry ratio between left and right half of the mouth
	Lip/Jaw Movement	velocity, acceleration, jerk, and speed of lower lip and jaw center
	Eyes	number of eye blinks per sec., eye opening, vertical displacement of eyebrows

Table 3: *Feature clusters from hierarchical clustering and the selected representative features. AUC represents the mean AUC for distinguishing bulbar onset and non-bulbar onset pALS samples across five cross validation folds. Only features with AUC>0.65 were included in the analysis. (RP: reading passage, DDK: diadochokinesis, cTV: cycle-to-cycle temporal variability, CTA: canonical timing alignment, PD: picture description, SIT: sentence intelligibility test, PPT: percentage pause time, HNR: harmonics-to-noise ratio, CPP: cepstral peak prominence, SBC: single breath count.)*

Cluster description	**Selected representative**	**AUC**
Timing: duration and rates	Speaking duration (RP)	0.84
Temporal DDK measures	cTV (DDK)	0.83
Timing alignment	CTA (RP)	0.83
Duration and word count for PD	Word count (PD)	0.83
Eyebrow displacement	Max. eyebrow vert. displ. (SIT)	0.78
Pause time	PPT (SIT)	0.77
Lip width	Max. lip width (RP)	0.72
Voice quality (read/free speech)	HNR (SIT)	0.71
Cepstral peak prominence (CPP)	CPP (RP)	0.69
Voice quality for SBC and DDK	HNR (DDK)	0.68
Lip aperture, mouth surface area	Mean lip aperture (SIT)	0.68
Eye opening measures	Max. eye opening (SIT)	0.68
Content and closed class words	Closed class word ratio (PD)	0.67
Min. and mean F0	Mean F0 (RP)	0.67
Jaw velocity for SIT	Max. jaw velocity down (SIT)	0.66
Duration measures for SBC and DDK	Syllable count (DDK)	0.65
Jaw velocity for RP	Max. jaw velocity up (RP)	0.65

y)-coordinates of the face for every video frame. We used 14 key landmarks to compute metrics like dynamics of articulators (jaw, lower lip), surface area of the mouth, and eyebrow raises. These facial features were normalised by dividing their values by the inter-caruncular distance to account for cross-participant variability due to position and movement relative to the camera [28]. Linguistic metrics were computed only for the picture description task, using the Python package spaCy [29] and automatic transcriptions of participant speech obtained using AWS Transcribe[2]. For an overview of automatically-extracted metrics, please see Table 2.

3.1. Feature Selection

To handle multicollinear features and to identify a good set of representative features, we applied hierarchical clustering on the Spearman rank-order correlations, similar to the approach in [30]. For this feature selection, data from 135 healthy English-

[2]`https://aws.amazon.com/transcribe/`

speaking controls[3] (71 female; mean age (standard deviation) = 59.9 (10.3) years) was considered in order to avoid data leakage in the experimental design and because data from healthy controls is most representative of normative feature ranges and correlations between features. Note that all subsequent analyses focus on patient data. Ward's method [31] was used for clustering and we plotted a dendrogram for visual inspection of the feature clusters. A distance threshold of 1.0[4] was chosen manually to select clusters that represented sensible feature groupings in terms of the domain. This threshold resulted in 27 clusters.

Next, for every feature, receiver operating characteristic (ROC) curve analysis was run in a 5-fold cross validation setup (`sklearn`'s *StratifiedGroupKFold* function) to determine the area under the ROC curve (AUC) for distinguishing bulbar onset participants from non-bulbar onset participants. There was no overlap of a participant's data between training and test folds. To further filter features, we imposed a minimum threshold for the ROC-AUC. Only features with an AUC$\geq$ 0.65 were considered for further analysis. These features are enumerated in Table 3 with their AUC values.

3.2. Group differences

We conducted non-parametric Kruskal-Wallis tests for the 17 metrics selected through feature selection, to test for differences between bulbar onset and non-bulbar onset pALS in both the English-speaking and Dutch-speaking cohorts. For this, we considered all data collected from the participants, irrespective of the timepoint in their disease progression. Effect sizes were measured as Glass' Δ [32] for all 17 metrics.

3.3. Test-retest reliability

For all 17 metrics, we also calculated test-retest reliability in the form of average Pearson's correlation coefficients. Pairs of sessions from the same participant, within a 31-day span of each other, were considered for this correlation. Test-retest reliability values are mentioned in parentheses after metric names in Figure 1.

3.4. Longitudinal trajectory

Growth curve models (GCMs) [33], which provide a linear fit for a non-linear mixed effects model, were run in R to estimate the trajectory of a metric over time in the English-speaking and Dutch-speaking cohorts, with random slopes and intercepts for each participant [15]. Growth curve models produce estimates of smoothed trajectories of change over time by using observed repeated measures of each individual. GCM curves for distinct cohorts can help identify differences in the longitudinal trajectory of measures in the two cohorts.

4. Results

Figure 1 shows effect sizes for all 17 metrics. For nine of the 17 metrics that showed differences between English-speaking bulbar pALS and non-bulbar pALS, the Dutch-speaking cohort also showed differences. These differences had the same directionality (bulbar > non-bulbar or vice versa) for all nine metrics in both language groups. Eight of the nine speech metrics showed differences in both language groups, whereas only one of the six facial metrics and none of the linguistic metrics showed differences in the Dutch bulbar versus Dutch non-bulbar comparison. Based on a visual inspection of Figure 1, we see that the magnitude of the effect size was similar across the English-speaking and Dutch-speaking pALS for two metrics in particular: SIT HNR and RP CTA. Most of the metrics showed similar test-retest reliability values in the Dutch and English cohorts except for jaw velocity metrics, DDK cTV and SIT PPT. When it came to the longitudinal trajectories of metrics, there were no differences in slopes for 14 of the 17 metrics. For three facial metrics — SIT maximum eyebrow vertical displacement, SIT mean lip aperture and SIT maximum eye opening — the Dutch-speaking cohort and English-speaking cohort displayed statistically different slopes or longitudinal trajectories (see Table 4). Since SIT maximum eye opening shows differences between the bulbar and non-bulbar cohorts in Dutch-speaking pALS (Figure 1), we looked at whether the differences between English-speaking pALS and Dutch-speaking pALS are driven by either the bulbar or non-bulbar cohorts. To do this, we compared longitudinal trajectories in Dutch bulbar pALS to those in English bulbar pALS, and the same for non-bulbar pALS. The three facial metrics had different slopes in the Dutch-speaking cohort and English-speaking cohort for both bulbar and non-bulbar pALS. In addition to the three facial metrics, when bulbar onset pALS were considered, PD word count, PD closed class word ratio and RP max. jaw velocity up also had different slopes. Whereas, DDK HNR had different trajectories in the English-speaking cohort and Dutch-speaking cohort when non-bulbar pALS were considered.

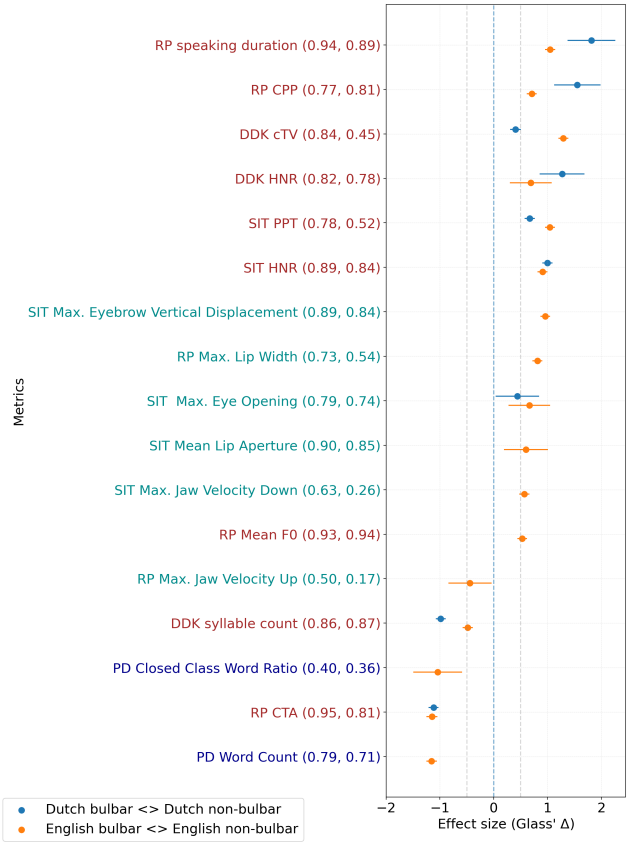

Figure 1: *Effect sizes of **speech**, **orofacial** and **linguistic** metrics shown with 95% confidence interval. The numbers in parentheses after metric names are test-retest reliability values for English and Dutch data, in that order. Positive effect size indicates greater values in bulbar pALS. RP: reading passage, CPP: cepstral peak prominence, DDK: diadochokinesis, cTV: cycle-to-cycle temporal variability, SIT: sentence intelligibility test, PPT: percentage pause time, HNR: harmonics-to-noise ratio, PD: picture description, CTA: canonical timing alignment*

[3]Data from Dutch-speaking controls was not available.

[4]The maximum distance at which all features would be combined into one single cluster was 5.4

Table 4: *Longitudinal trajectory of metrics. 143 English-speaking pALS and 26 Dutch-speaking pALS. An asterisk (*) indicates that the longitudinal trajectories of the metric differ between the English-speaking and Dutch-speaking pALS cohorts.*

Metric	p-value of difference	pALS Cohort	Intercept ± standard error	Slope ± standard error
RP speaking duration (seconds)	0.7751	English-speaking	30.10 ± 9.52	0.1188 ± 0.0881
		Dutch-speaking	29.12 ± 8.90	0.1440 ± 0.0826
DDK cTV (seconds)	0.9260	English-speaking	0.06 ± 0.01	-1.96e-05 ± 5.64e-05
		Dutch-speaking	0.07 ± 0.01	-2.48e-05 ± 5.51e-05
RP CTA (%)	0.6379	English-speaking	80.70 ± 7.77	-0.0971 ± 0.0603
		Dutch-speaking	68.87 ± 7.38	-0.0688 ± 0.0577
PD word count (words)	0.2219	English-speaking	68.51 ± 25.67	0.1653 ± 0.1805
		Dutch-speaking	10.64 ± 24.38	0.3858 ± 0.1746
SIT Max. eyebrow vert. displ.	< 0.0001*	English-speaking	2.78 ± 0.34	-0.0002 ± 0.0021
		Dutch-speaking	0.88 ± 0.32	0.0117 ± 0.0020
SIT PPT (%)	0.9688	English-speaking	2.22 ± 2.04	0.0112 ± 0.0150
		Dutch-speaking	2.56 ± 1.94	0.0106 ± 0.0145
RP Max. lip width	0.7308	English-speaking	1.64 ± 0.07	-2.69e-05 ± 4.50e-04
		Dutch-speaking	1.59 ± 0.07	1.28e-04 ± 4.43e-04
SIT HNR (dB)	0.5290	English-speaking	10.12 ± 1.42	0.0049 ± 0.0091
		Dutch-speaking	11.03 ± 1.34	-0.0009 ± 0.0088
RP CPP (dB)	0.2618	English-speaking	26.34 ± 1.25	0.0029 ± 0.0088
		Dutch-speaking	26.82 ± 1.19	-0.0070 ± 0.0086
DDK HNR (dB)	0.7355	English-speaking	7.56 ± 1.57	0.0073 ± 0.0108
		Dutch-speaking	8.58 ± 1.50	0.0036 ± 0.0105
SIT Mean lip aperture	< 0.0001*	English-speaking	0.78 ± 0.10	-0.0002 ± 0.0006
		Dutch-speaking	0.10 ± 0.10	0.0042 ± 0.0006
SIT Max. eye opening	0.0002*	English-speaking	0.35 ± 0.05	-8.43e-05 ± 2.82e-04
		Dutch-speaking	0.14 ± 0.04	9.75e-04 ± 2.73e-04
PD Closed Class Word Ratio	0.1873	English-speaking	0.42 ± 0.03	0.0001 ± 0.0002
		Dutch-speaking	0.40 ± 0.03	0.0004 ± 0.0002
RP Mean F0 (Hz)	0.6174	English-speaking	141.71 ± 13.24	0.0723 ± 0.0815
		Dutch-speaking	147.80 ± 12.62	0.0316 ± 0.0801
SIT Max. jaw velocity down	0.2334	English-speaking	0.1 ± 0.01	-3.86e-05 ± 9.64e-05
		Dutch-speaking	0.06 ± 0.01	7.63e-05 ± 9.44e-05
DDK syllable count (syllables)	0.4926	English-speaking	62.46 ± 23.29	-0.0221 ± 0.1509
		Dutch-speaking	56.53 ± 21.88	0.0814 ± 0.1447
RP Max. jaw velocity up	0.5221	English-speaking	0.11 ± 0.02	5.33e-06 ± 1.10e-04
		Dutch-speaking	0.11 ± 0.02	7.54e-05 ± 1.08e-04

5. Discussion

In this work, we set out to investigate whether speech-based objective digital biomarkers are consistent in remote tracking of ALS disease progression across languages. First, we looked at whether a selection of metrics that are useful in distinguishing English-speaking bulbar pALS from non-bulbar pALS show similar differences between Dutch-speaking bulbar pALS and non-bulbar pALS. Most of the speech acoustic and timing metrics show similar effect sizes in the English and Dutch cohorts. Facial metrics, however, do not show similar effect sizes except for maximum eye opening during sentence reading. Notably, reading passage canonical timing alignment — a number between 0% (non-alignment with a canonical production) and 100% (perfect alignment with a canonical production) as measured by the normalised inverse Levenshtein edit distance between words and silence boundaries — shows very similar effect sizes in English and Dutch. Interestingly, this measure has been found to be very responsive in tracking statistical and clinical change in pALS [17] even with small sample sizes [18]. Most metrics had similar test-retest reliability in the English-speaking and Dutch-speaking cohorts. For the metrics where test-retest reliability is lower in the Dutch cohort, it remains to be seen whether the gap is closed when a larger Dutch dataset is available. All speech metrics had similar slopes in the English and Dutch cohorts when growth curve models were fit to predict the average longitudinal trajectory of the metrics. Linguistic metrics and three of the six facial metrics also showed similar trajectories. Maximum eyebrow vertical displacement, mean lip aperture and maximum eye opening during sentence reading had steeper longitudinal trajectory slopes for the Dutch cohort than the English cohort. It is not clear why these three facial metrics should change differently in two cohorts with dissimilar linguistic backgrounds. One explanation is that the smaller sample size in the Dutch cohort makes this measure much noisier. Indeed, one limitation of this work is that the average time span of data collection for the Dutch-speaking cohort is much smaller than that in the English-speaking cohort. The Dutch-speaking data skews male. Future work with larger, more balanced sample sizes need to be conducted. This case study included languages from the same West Germanic language family [34]. Future work should test whether these observations are consistent when unrelated languages are considered.

In conclusion, speech-based digital biomarkers in pALS show good consistency and great promise in tracking ALS disease progression in cohorts of patients speaking different languages. The impact of language on disease-related changes, if any, seems minimal for the majority of the biomarkers considered.

6. Acknowledgements

This work was funded by the National Institutes of Health grant R42DC019877. We thank EverythingALS, the Peter Cohen Foundation and UMC Utrecht for participant recruitment.

7. References

[1] O. Hardiman, A. Al-Chalabi, A. Chio, E. M. Corr, G. Logroscino, W. Robberecht, P. J. Shaw, Z. Simmons, and L. H. Van Den Berg, "Amyotrophic lateral sclerosis," *Nature Reviews Disease Primers*, vol. 3, no. 1, pp. 1–19, 2017.

[2] A. Chio, G. Logroscino, O. Hardiman, R. Swingler, D. Mitchell, E. Beghi, B. G. Traynor, E. Consortium *et al.*, "Prognostic factors in als: a critical review," *Amyotrophic lateral sclerosis*, vol. 10, no. 5-6, pp. 310–323, 2009.

[3] J. R. Green, Y. Yunusova, M. S. Kuruvilla, J. Wang, G. L. Pattee, L. Synhorst, L. Zinman, and J. D. Berry, "Bulbar and speech motor assessment in ALS: Challenges and future directions," *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, vol. 14, no. 7-8, pp. 494–500, 2013.

[4] L. C. Wijesekera and P. Nigel Leigh, "Amyotrophic lateral sclerosis," *Orphanet Journal of Rare Diseases*, vol. 4, pp. 1–22, 2009.

[5] L. J. Haverkamp, V. Appel, and S. H. Appel, "Natural history of amyotrophic lateral sclerosis in a database population validation of a scoring system and a model for survival prediction," *Brain*, vol. 118, no. 3, pp. 707–719, 1995.

[6] A. Bandini, J. R. Green, J. Wang, T. F. Campbell, L. Zinman, and Y. Yunusova, "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, pp. 1118–1129, 2018.

[7] 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)*. IEEE, 2018, pp. 150–157.

[8] R. Norel, M. Pietrowicz, C. Agurto, S. Rishoni, and G. Cecchi, "Detection of Amyotrophic Lateral Sclerosis (ALS) via Acoustic Analysis," in *Proc. Interspeech 2018*, 2018, pp. 377–381.

[9] C. Barnett, J. R. Green, R. Marzouqah, K. L. Stipancic, J. D. Berry, L. Korngut, A. Genge, C. Shoesmith, H. Briemberg, A. Abrahao *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.

[10] M. Neumann, O. Roesler, J. Liscombe, H. Kothare, D. Suendermann-Oeft, D. Pautler, I. Navar, A. Anvar, J. Kumm, R. Norel, E. Fraenkel, A. V. Sherman, J. D. Berry, G. L. Pattee, J. Wang, J. R. Green, and V. Ramanarayanan, "Investigating the Utility of Multimodal Conversational Technology and Audiovisual Analytic Measures for the Assessment and Monitoring of Amyotrophic Lateral Sclerosis at Scale," in *Proc. Interspeech 2021*, 2021, pp. 4783–4787.

[11] D. L. Guarin, B. Taati, A. Abrahao, L. Zinman, and Y. Yunusova, "Video-based facial movement analysis in the assessment of bulbar amyotrophic lateral sclerosis: clinical validation," *Journal of Speech, Language, and Hearing Research*, vol. 65, no. 12, pp. 4667–4678, 2022.

[12] L. E. Simmatis, J. Robin, M. J. Spilka, and Y. Yunusova, "Detecting bulbar amyotrophic lateral sclerosis (als) using automatic acoustic analysis," *BioMedical Engineering OnLine*, vol. 23, no. 1, p. 15, 2024.

[13] M. Eshghi, Y. Yunusova, K. P. Connaghan, B. J. Perry, M. F. Maffei, J. D. Berry, L. Zinman, S. Kalra, L. Korngut, A. Genge *et al.*, "Rate of speech decline in individuals with amyotrophic lateral sclerosis," *Scientific Reports*, vol. 12, no. 1, p. 15713, 2022.

[14] Y. Yunusova, J. R. Green, M. J. Lindstrom, G. L. Pattee, and L. Zinman, "Speech in ALS: Longitudinal Changes in Lips and Jaw Movements and Vowel Acoustics," *Journal of Medical Speech-Language Pathology*, vol. 21, no. 1, 2013.

[15] G. M. Stegmann, S. Hahn, J. Liss, J. Shefner, S. Rutkove, K. Shelton, C. J. Duncan, and V. Berisha, "Early detection and tracking of bulbar changes in ALS via frequent and remote speech analysis," *NPJ Digital Medicine*, vol. 3, no. 1, p. 132, 2020.

[16] Y. Yunusova, J. R. Green, M. J. Lindstrom, L. J. Ball, G. L. Pattee, and L. Zinman, "Kinematics of disease progression in bulbar ALS," *Journal of Communication Disorders*, vol. 43, no. 1, pp. 6–20, 2010.

[17] H. Kothare, M. Neumann, J. Liscombe, J. Green, and V. Ramanarayanan, "Responsiveness, Sensitivity and Clinical Utility of Timing-Related Speech Biomarkers for Remote Monitoring of ALS Disease Progression," in *Proc. Interspeech*, 2023, pp. 2323–2327.

[18] H. Kothare, M. Neumann, and V. Ramanarayanan, "Relationship between sample size and responsiveness of speech-based digital biomarkers in ALS," in *Proceedings of: International Society for CNS Clinical Trials and Methodology (ISCTM 2024) Spring Conference, Washington, D.C.*, 2024.

[19] A. Favaro, L. Moro-Velázquez, A. Butala, C. Motley, T. Cao, R. D. Stevens, J. Villalba, and N. Dehak, "Multilingual evaluation of interpretable biomarkers to represent language and speech patterns in parkinson's disease," *Frontiers in Neurology*, vol. 14, p. 1142642, 2023.

[20] J. Rusz, J. Hlavnička, M. Novotnỳ, T. Tykalová, A. Pelletier, J. Montplaisir, J.-F. Gagnon, P. Dušek, A. Galbiati, S. Marelli *et al.*, "Speech biomarkers in rapid eye movement sleep behavior disorder and parkinson disease," *Annals of neurology*, vol. 90, no. 1, pp. 62–75, 2021.

[21] T. L. Whitehill, "Studies of chinese speakers with dysarthria: informing theoretical models," *Folia Phoniatrica et Logopaedica*, vol. 62, no. 3, pp. 92–96, 2010.

[22] V. Ramanarayanan, D. Pautler, L. Arbatti, A. Hosamath, M. Neumann, H. Kothare, O. Roesler, J. Liscombe, A. Cornish, D. Habberstad, V. Richter, D. Fox, D. Suendermann-Oeft, and I. Shoulson, "When Words Speak Just as Loudly as Actions: Virtual Agent Based Remote Health Assessment Integrating What Patients Say with What They Do," in *Proc. INTERSPEECH 2023*, 2023, pp. 678–679.

[23] V. Ramanarayanan, "Multimodal technologies for remote assessment of neurological and mental health," *Journal of Speech, Language, and Hearing Research*, pp. 1–8, 2024.

[24] P. Boersma, "Praat, a system for doing phonetics by computer," *Glot International*, vol. 5, no. 9/10, pp. 341–345, 2001.

[25] M. McAuliffe, M. Socolof, S. Mihuc, M. Wagner, and M. Sonderegger, "Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi," in *Proc. Interspeech 2017*, 2017, pp. 498–502.

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

[27] V. Bazarevsky, Y. Kartynnik, A. Vakunov, K. Raveendran, and M. Grundmann, "BlazeFace: Sub-millisecond Neural Face Detection on Mobile GPUs," *CoRR*, vol. abs/1907.05047, 2019. [Online]. Available: http://arxiv.org/abs/1907.05047

[28] O. Roesler, H. Kothare, W. Burke, M. Neumann, J. Liscombe, A. Cornish, D. Habberstad, D. Pautler, D. Suendermann-Oeft, and V. Ramanarayanan, "Exploring Facial Metric Normalization For Within- and Between-Subject Comparisons in a Multimodal Health Monitoring Agent," in *Companion Publication of the 2022 International Conference on Multimodal Interaction*, ser. ICMI '22 Companion. New York, NY, USA: Association for Computing Machinery, 2022, p. 160–165. [Online]. Available: https://doi.org/10.1145/3536220.3558071

[29] M. Honnibal and I. Montani, "spacy 2: Natural language understanding with bloom embeddings, convolutional neural networks and incremental parsing. neural machine translation," in *Proceedings of the Association for Computational Linguistics (ACL)*, 2017, pp. 688–697.

[30] D. Ienco and R. Meo, "Exploration and Reduction of the Feature Space by Hierarchical Clustering," in *Proceedings of the 2008 Siam International Conference on Data Mining*. SIAM, 2008, pp. 577–587.

[31] J. H. Ward Jr, "Hierarchical grouping to optimize an objective function," *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 236–244, 1963.

[32] G. V. Glass, B. McGaw, and M. L. Smith, "Meta-analysis in social research," *(Sage Publications)*, 1981.

[33] D. Von Rosen, "The growth curve model: a review," *Communications in Statistics-Theory and Methods*, vol. 20, no. 9, pp. 2791–2822, 1991.

[34] B. Comrie *et al.*, *The World's Major Languages*. Routledge London, 1987.