8573L Screen 177 - Assessing the Utility of Vowel Space Characteristics in Remotely Recorded Speech for ALS Progress Monitoring

Michael Neumann[1], Hardik Kothare[1], Jackson Liscombe[1], Vikram Ramanarayanan[1,2]
[1]Modality.AI, Inc., [2]University of California, San Francisco
michael.neumann@modality.ai

Background & Objective

- **Amyotrophic lateral sclerosis (ALS)** is a **progressive neurodegenerative disorder** that affects motor neurons
- People with ALS (pALS) experience speech impairment
- Vowel features can characterize speech motor control [1]
- **Vowel space area (VSA)** is a commonly utilized measure [2,3]
- **Formant centralization ratio (FCR)** is less sensitive to inter-speaker variability compared to VSA [4]
- **Previous studies** are based on **recordings in clean lab conditions** using specific recording hardware
- To our knowledge, this is the **first analysis of vowel characteristics for identifying dysarthric speech that is based on remotely recorded speech** with diverse consumer-grade hardware

Objective: Assess the feasibility and usefulness of automatically measured vowel characteristics – based on remotely recorded speech samples in varying conditions – for progress monitoring in ALS

Methods and Materials

- **72 pALS** (33 females, mean age 62.7, SD: 8.6 years) participated in a **self-administered speech assessment** using a web-based multimodal dialog system [5] (561 recordings sessions total)
- A virtual guide, Tina, guided participants through a set of speech exercises
- Stimuli used for this study: **consonant-vowel-consonant (CVC) words 'peep' (/pi:p/), 'pep' (/pɛp/), 'poop' (/pu:p/), and 'pop' (/pɑp/)**
- The bilabial /p/ was used at onset and coda to reduce co-articulation effects
- We **extracted formant frequencies F1 and F2** using Praat and calculated three vowel features (Table 1)
- Features were z-score normalized separately for female and male speakers

Feature	Formula	Description
Vowel space area (VSA)	$\lvert (F2u \cdot F1i)-(F1u \cdot F2i)+(F2i \cdot F1\varepsilon)-(F1i \cdot F2\varepsilon)+(F2\varepsilon \cdot F1\alpha)-(F1\varepsilon \cdot F2\alpha)+(F2\alpha \cdot F1u)-(F1\alpha \cdot F2u) \rvert \,/\, 2$	Area of quadrilateral formed by four corner vowels on F1-F2 space, in Hz^2; calculated with the shoelace formula
Formant centralization ratio (FCR)	(F2u + F2α + F1i + F1u) / (F2i + F1α)	Vowel centralization measure that reduces sensitivity to inter-speaker variability
F2i/F2u ratio	F2i / F2u	Sensitive to changes in articulatory movements (anterior-posterior tongue movement and lip rounding)

Table 1: Vowel features based on first two formants F1 and F2.

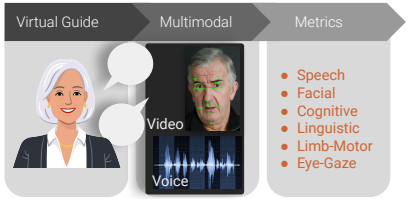

Fig. 1: Schematic diagram of the Modality.AI dialog platform.

- Stratification: following [6], **samples were grouped into two cohorts based on self-reported ALSFRS-R [7] bulbar sub score** (ranges from 0 to 12)
  - **Bulbar symptomatic (BUL)**, 273 sessions
  - **Bulbar pre-symptomatic (PRE)**, 288 sessions
- At baseline visit, 31 participants in PRE, 41 participants in BUL cohort
- We ran non-parametric **Kruskal-Wallis tests to identify statistically significant differences in vowel characteristics** between the cohorts
- Receiver operating characteristics **(ROC) analysis** was performed, both with static feature values and with features' rate of change over time, **to investigate usefulness for monitoring progression in ALS**

Results

- **VSA on average lower in BUL group** (229,948 Hz^2 for females, 150,569 Hz^2 for males) than in PRE (331,077 Hz^2 for females, 182,152 Hz^2 for males)
- **FCR on average higher in BUL group** (f: 1.19, m: 1.19) than in PRE (f: 1.03, m: 1.13) → stronger vowel centralization in bulbar symptomatic pALS
- **F2i/F2u on average lower in BUL group** (f: 1.78, m: 1.78) than in PRE (f: 2.10, m: 1.99) → front vowel /i/ and the back vowel /u/ tend to move closer to each other in terms of F2 with stronger articulatory deficits [4]

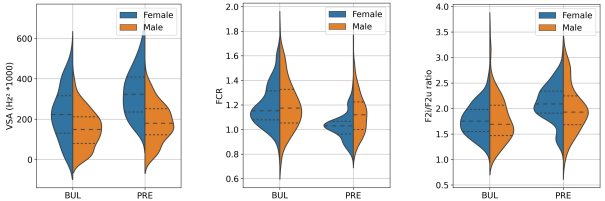

Fig. 2: Distribution of the three vowel features in the BUL and PRE cohorts.

- Kruskal-Wallis test (at baseline): **FCR statistically significantly different between the BUL and PRE cohort** ($p<0.05$, Glass' $\Delta=1.16$)
- ROC analysis (discriminate between the two cohorts): best results in terms of area under the curve (AUC) for **FCR and F2i/F2u ratio (AUC=0.68** for both)
- In the longitudinal analysis, the **change rate of the VSA** yielded the best result (AUC=0.70)

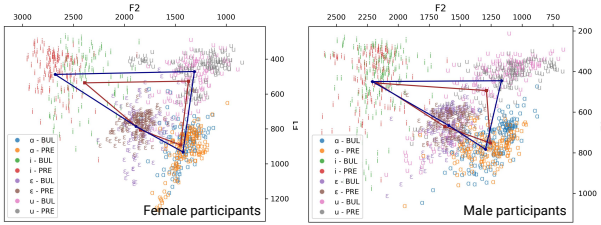

Fig. 3: Bulbar symptomatic pALS exhibit on average a smaller vowel space area (red lines) than pre-symptomatic pALS (blue lines). Quadrilaterals are drawn based on the centroids of each corner vowel for each cohort.

Conclusions

- We **investigated the clinical utility of** automatically extracted **vowel features for progress monitoring in ALS**
- findings indicate that **vowel features reflect differences in speech motor control** between bulbar symptomatic and pre-symptomatic people with ALS
- In this dataset, **differences** are **more pronounced in female participants**
- Observations underscore the **utility of remotely collected speech samples** recorded with consumer-grade hardware, **using an interactive dialog system** for naturalistic speech elicitation

References

[1] Berisha, V., Sandoval, S., et al. (2014). Characterizing the distribution of the quadrilateral vowel space area. The Journal of the Acoustical Society of America, 135(1), 421-427.
[2] Weismer, G., Jeng, J. Y., et al. (2001). Acoustic and intelligibility characteristics of sentence production in neurogenic speech disorders. Folia Phoniatrica et Logopaedica, 53(1), 1-18.
[3] Gómez-Vilda, P., Londral, A. R. M., et al. (2015). Monitoring amyotrophic lateral sclerosis by biomechanical modeling of speech production. Neurocomputing, 151, 130-138.
[4] Sapir, S., Ramig, L. O., et al. (2010). Formant centralization ratio: A proposal for a new acoustic measure of dysarthric speech. Journal of speech, language, and hearing research: JSLHR, 53(1), 114.
[5] Suendermann-Oeft, D., Robinson, A., et al. (2019). Nemsi: A multimodal dialog system for screening of neurological or mental conditions. In Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents (pp. 245-247).
[6] Allison, K. M., Yunusova, Y., et al. (2017). 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, 18(5-6), 358-366.
[7] Cedarbaum, J. M., Stambler, N., et al. (1999). The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. Journal of the neurological sciences, 169(1-2), 13-21.

This work was funded by the National Institutes of Health grant R42DC019877. We thank all study participants for their time and we gratefully acknowledge the contribution of the Peter Cohen Foundation and EverythingALS towards participant recruitment and data collection.