Multimodal dialog based speech and facial biomarkers capture differential disease progression rates for ALS remote patient monitoring

Michael Neumann[1], Oliver Roesler[1], Jackson Liscombe[1], Hardik Kothare[1], David Suendermann-Oeft[1], James D. Berry[2], Ernest Fraenkel[3], Raquel Norel[4], Aria Anvar[5], Indu Navar[5], Alexander V. Sherman[2,6], Jordan R. Green[2,6] and Vikram Ramanarayanan[1,7]
[1]Modality.AI, [2]MGH Institute of Health Professions, [3]Massachusetts Institute of Technology, [4]IBM Thomas J. Watson Research Center, [5]EverythingALS, Peter Cohen Foundation, [6]Harvard University, [7]University of California, San Francisco
Contact: michael.neumann@modality.ai

Introduction

Objective: identify audiovisual speech markers that are responsive to clinical progression of amyotrophic lateral sclerosis (ALS)

Methods: longitudinal analysis of acoustic and visual speech measures extracted from a web-based conversational assessment of people with ALS (pALS)

Implications: multimodal remote patient monitoring allows to
(i) measure changes in speech markers frequently and cost-effectively, while
(ii) capturing differences between slow and fast progressors

Methods and Materials

- 54 pALS (Table 2) completed at least three sessions between October 2020 and July 2021 using a cloud-based multimodal dialog platform[1] (Illustration on Figure 1).
- Each session consists of structured and spontaneous speech tasks (Table 1), followed by self-reported ALSFRS-R questionnaire.
- Rate of progression was calculated based on first and last ALSFRS-R score. pALS were stratified into slow and fast progressors based on a threshold of 0.47 points/month[2].
- Statistical tests were conducted to identify acoustic and visual metrics for which the rates of change are significantly different between the two cohorts.

Speech task	Acoustic measures	Visual measures
Held vowel phonation	mean F0, HNR, jitter, shimmer, CPP , duration	velocity, acceleration, and jerk of lower lip and jaw center, eye opening, vertical eyebrow displacement, eye blinks, area of the mouth, symmetry ratio of the mouth area
DDK	duration, syllable rate, cTV	
Bamboo reading passage, SIT, Picture description	duration, speaking and articulation rate, PPT, HNR, mean F0, CPP	

Table 1. Stimuli and corresponding extracted acoustic & visual speech measures. DDK-AMR: diadochokinesis; SIT: speech intelligibility test; HNR: harmonics-to-noise ratio; CPP: cepstral peak prominence; cTV: cycle-to-cycle temporal variation; PPT: percent pause time.

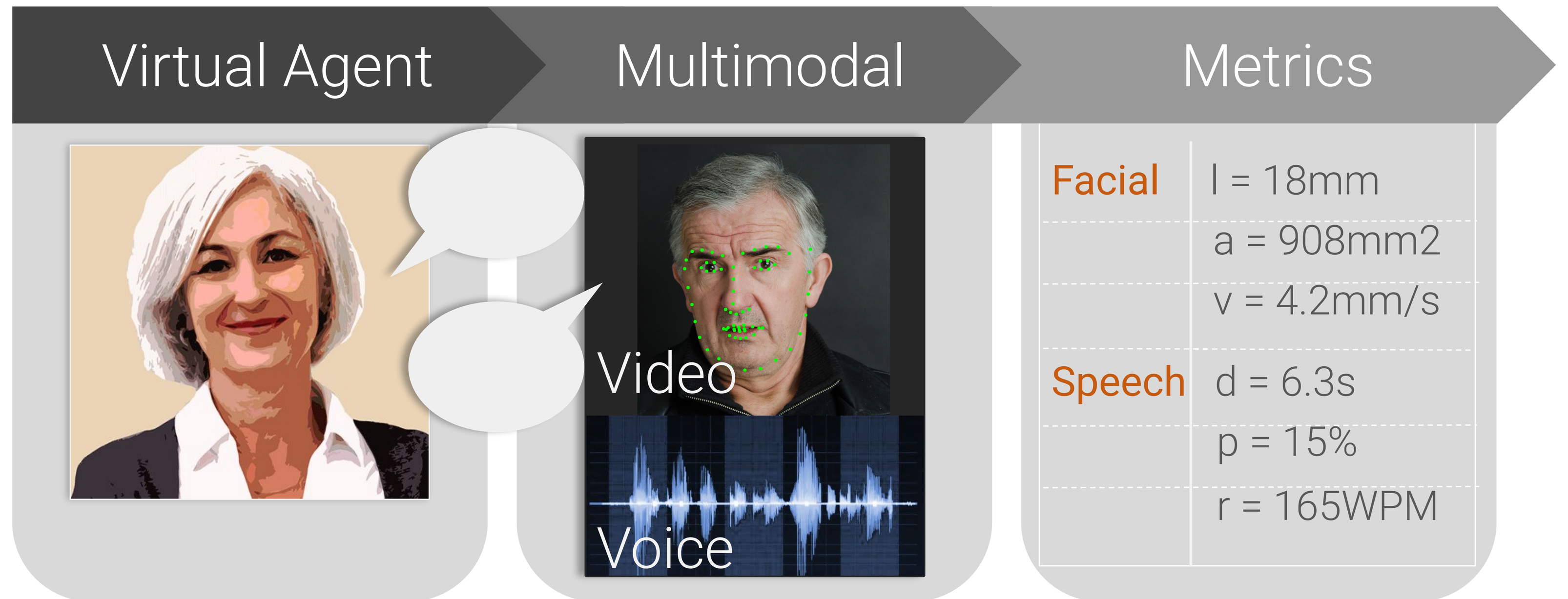

Figure 1. Modality.AI dialog platform.

	#subjects	median ALS-FRS-R progression rate	Mean age (years)*	Mean bulbar subscore*	Mean time since onset (months)*	Mean #sessions
Slow P	17 F, 19 M	0.0 F, 0.0 M	60.3 ± 8.6	10.5 ± 1.6	68.7 ± 57.9	16
Fast P	9 F, 9 M	0.74 F, 1.5 M	61.3 ± 9.4	10 ± 2.1	39.1 ± 16.7	9.5

Table 2. Participants. P: progressors; F: female; M: male. Progression in points/month. *at first sessions

Results and Discussion

- A variety of metrics showed statistically significant differences in their rate of change between slow and fast progressors (Figure 2).
  In particular, metrics related to **timing, voice quality, fundamental frequency**, and **higher order kinematics of the lower lip**, **eye opening and eyebrow positioning** showed differences.
- These differences in trajectories signify changes in **articulatory motor control** and their **acoustic implications** as well as **extraocular and periocular motor control**.
- For example, PPT in the Bamboo passage was **increasing** by 0.09%/month (median slope) for fast progressors, while it was **decreasing** by -0.07%/month for slow progressors (which might be attributed to a training effect of the repeated task performance).
- To confirm robustness of the findings, data from more participants over a longer period of time will be analyzed in a future study.

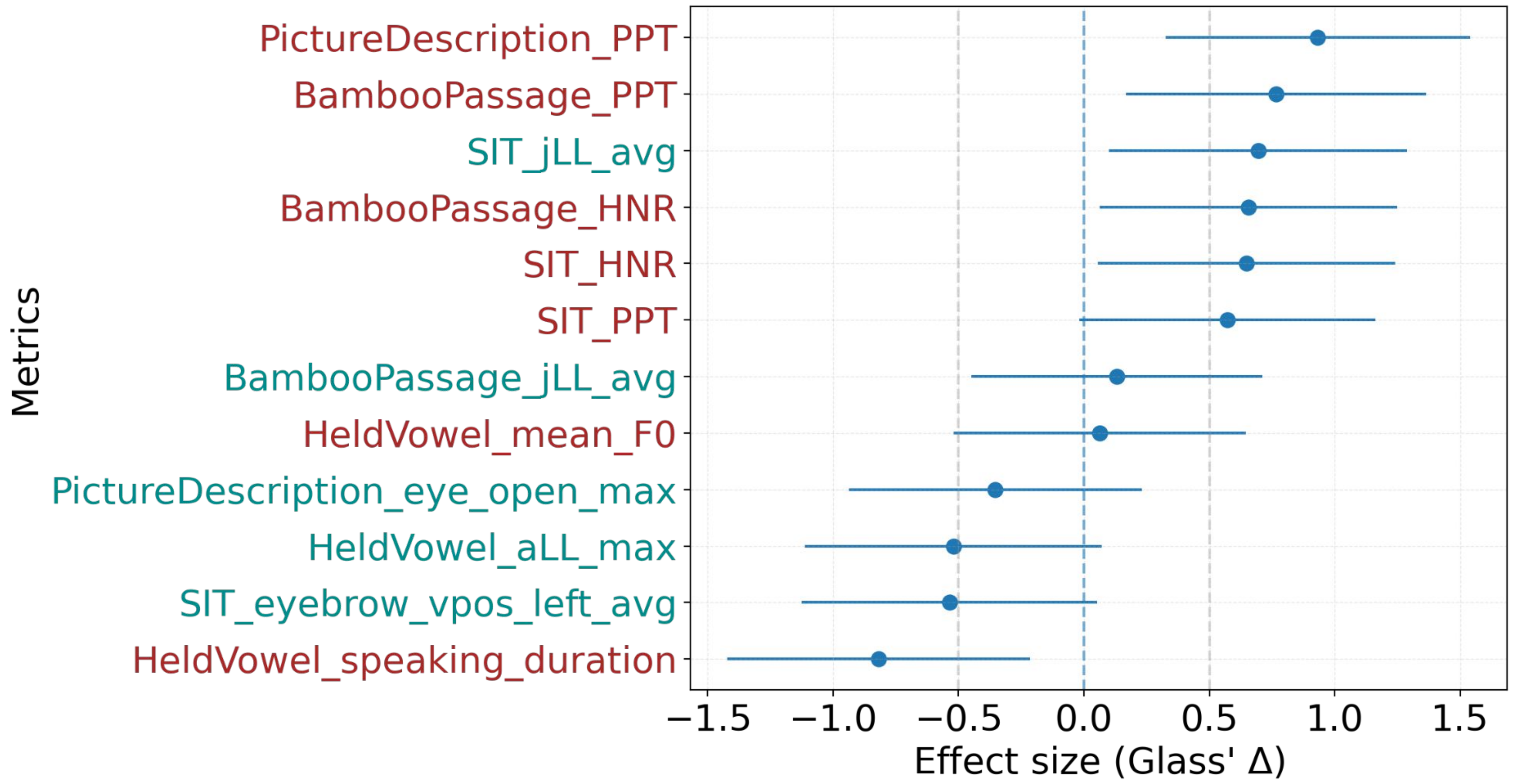

Figure 2. Effect sizes of **acoustic** and **visual** metrics that show statistically significant differences in the rate of change between cohorts at $p < 0.05$, shown with 95% confidence interval. Positive effects mean that the quantity is increasing at a larger rate in fast progressors. Abbreviations: jLL_avg: average jerk of lower lip, aLL_max: maximum acceleration of lower lip.

Limitations:

- Duration between participants first and last sessions varies (on average 135 ± 70 days), as well as the number of sessions from each participant.
- Sustained phonation speech samples exhibited browser-driven systematic attenuation of loudness, which can affect feature extraction.

Conclusions

- As hypothesized, **frequent and continuous monitoring of acoustic and visual speech markers** can **capture objective physiological changes** that may not be captured by subjective scales like the ALSFRS-R instrument, the current clinical standard to track progression in ALS.
- Changes in these audiovisual metrics could serve as **potential digital biomarkers**, which could contribute towards patient stratification and tracking of outcomes following pharmaceutical interventions.

1 Ramanarayanan, Vikram, et al. "Toward Remote Patient Monitoring of Speech, Video, Cognitive and Respiratory Biomarkers Using Multimodal Dialog Technology." INTERSPEECH. 2020.
2 Requardt, Maria Viktoria, et al. "Clinical Determinants of Disease Progression in Amyotrophic Lateral Sclerosis—A Retrospective Cohort Study." Journal of clinical medicine 10.8 (2021): 1623.