LESSONS LEARNED FROM A LARGE-SCALE AUDIO-VISUAL REMOTE DATA COLLECTION FOR AMYOTROPHIC LATERAL SCLEROSIS RESEARCH

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

Background & Aims

Background: Modality.AI and EverythingALS created an open data collection platform initiative in 2020 to investigate the potential of speech and facial biomarkers to serve as endpoints in clinical trials and in assessment of ALS disease progression [1, 2]. The interactive platform, powered by Modality's virtual agent Tina, records speech, facial video and survey instrument data from participants, emulating the role of a neurologist or speech pathologist walking them through speaking exercises [1].

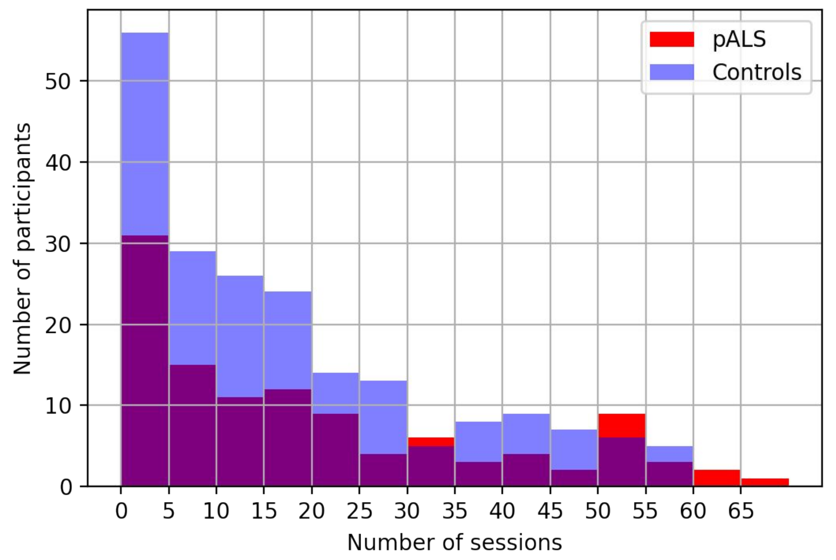
Aim: To summarize our progress and lessons learned from a 1.5 year long data collection of speech and video recordings of people with ALS and healthy controls via the platform. The research community can use this data to accelerate the development of biomarkers, diagnostics, therapies, and fundamental scientific understanding of ALS.

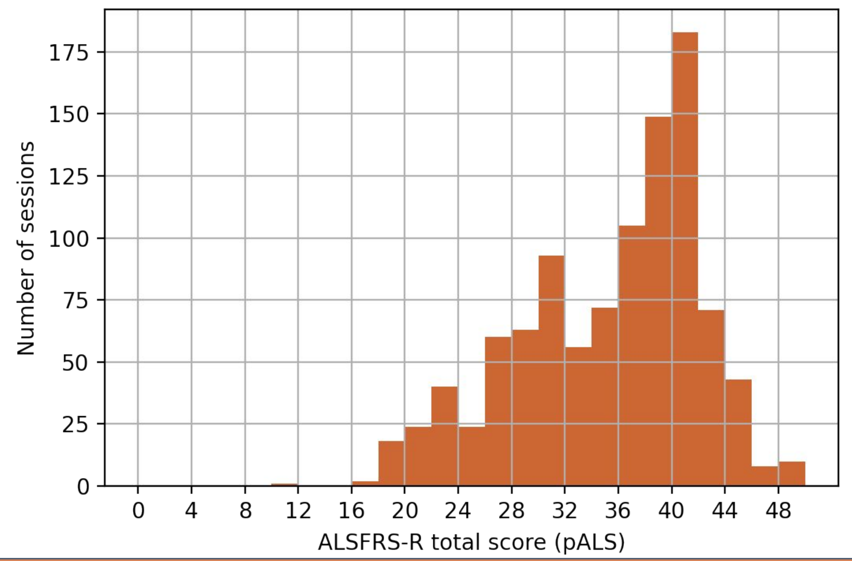
Findings: Multimodal dialog based remote patient monitoring allows us to: (i) measure changes in speech markers frequently and cost-effectively, while (ii) capturing differences between slow and fast progressors.

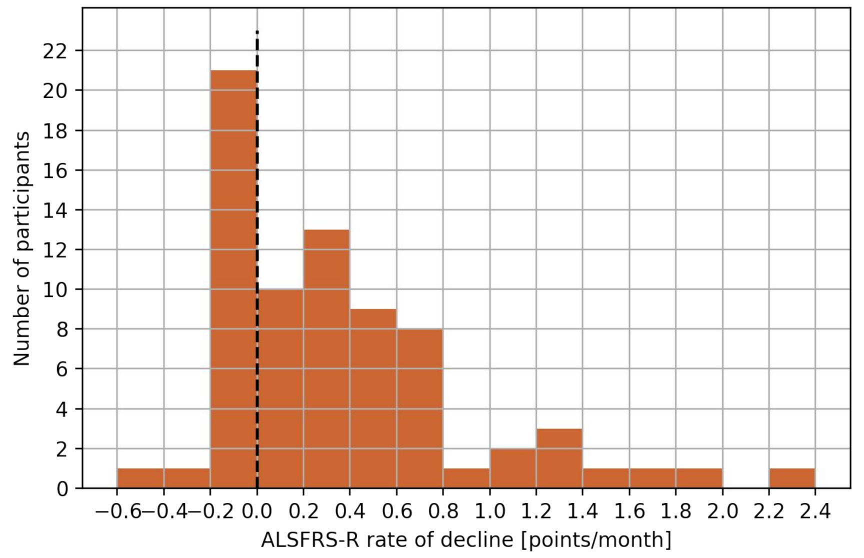
Key Stats

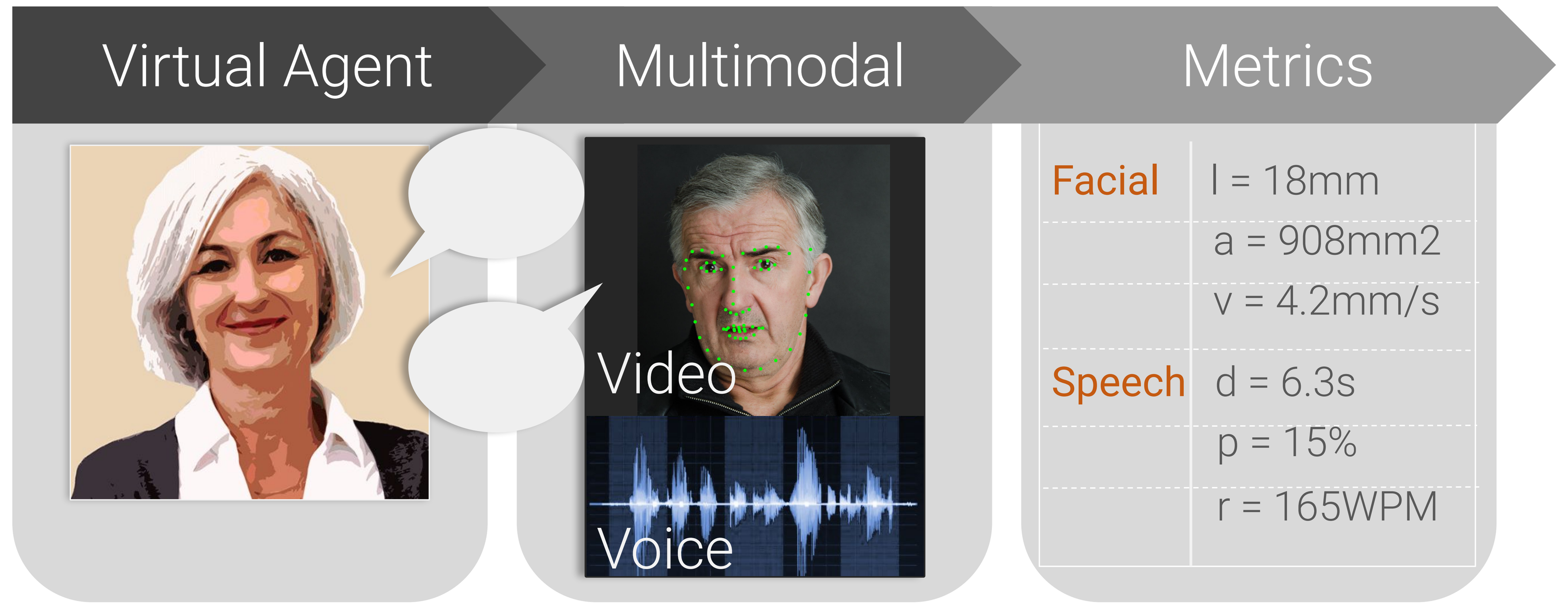
	#subjects (F: female)	total #sessions	Mean age (years)
pALS	112 (54 F)	2,188	60.2 ± 10.2
Controls	202 (137 F)	3,459	53.7 ± 13.7
Total	314 (191 F)	5,647	56.0 ± 13.0

Table 1. Participants stats as of 2022-02-22.
Mean time since onset (months) = 50.2 ± 45.9









Modality.AI dialog platform [1]

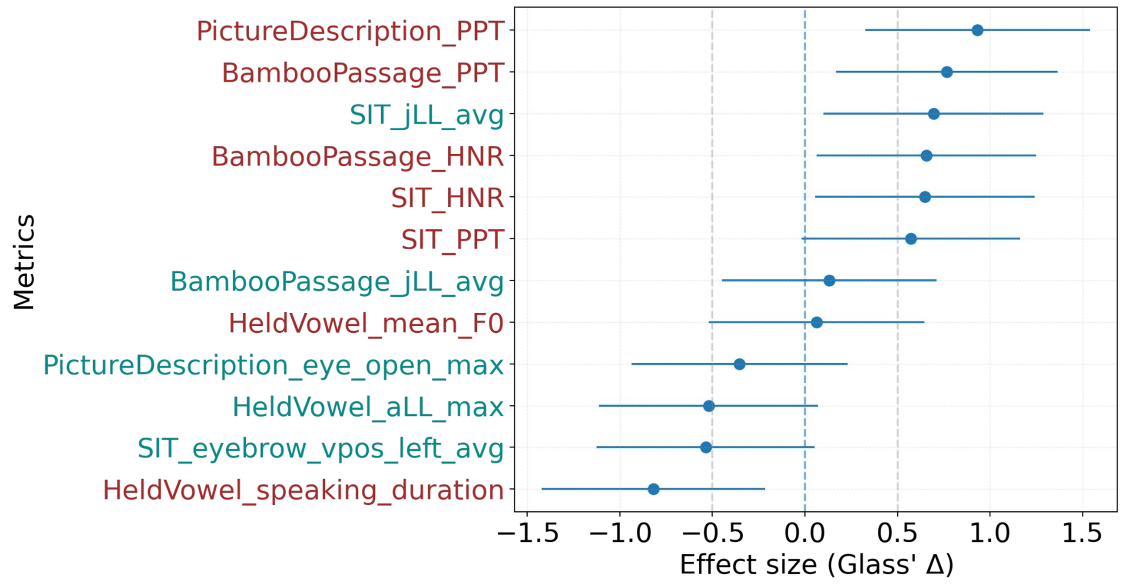
Analyses

- Each session consists of structured and spontaneous speech tasks (Table 2), followed by self-reported ALSFRS-R questionnaire.
- We earlier showed that these remotely extracted speech and facial biomarkers show promise for assisting early diagnosis (classifying healthy controls vs bulbar presymptomatic) and progress monitoring (bulbar presymptomatic vs bulbar symptomatic) of ALS. See [3] for more details.
- To investigate how well metrics capture rate of progression (calculated based on first and last ALSFRS-R score): pALS were stratified into slow and medium-to-fast progressors based on a threshold of 0.47 points/month. Statistical tests were conducted to identify acoustic & visual metrics where rates of change were significantly different between the two cohorts [4].

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 2. 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.

Findings & Lessons Learned



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 (see [4]).

Abbreviations: jLL_avg: average jerk of lower lip, aLL_max: maximum acceleration of lower lip.

Opportunities:

- Frequent measurements aid progress monitoring
- Remote setup enables scalable and inexpensive collection of large dataset
- 77% average monthly retention rate (as of Mar 5, 2022)
- 70% average completion rate (as of Mar 5, 2022)

Challenges:

- Noisy data:
  - Missing observations (participant fatigue, boredom, forgetfulness, etc.)
  - Heterogeneous w.r.t. time since onset, number of sessions per user, time between sessions, total duration from first to last session
  - Self-reported ALSFRS-R and ROADS: subjective and not always reliable (mistakes happen)
- Technical challenges due to device and browser settings (e.g., attenuation of sustained vowels) -> assessments need to be adapted to the setting

Conclusions

- **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.

REFERENCES: [1] Ramanarayanan et al. (Interspeech 2020) [2] Requardt et al. (Journal of Clinical Medicine 2021) [3] Neumann et al. (Interspeech 2021) [4] Neumann et al. (ALS MND 2021)