

Canonical Timing Alignment of Read Speech in ALS is Correlated with Perceived Speech Impairment and Listener Effort

Hardik Kothare¹, Jackson Liscombe¹, Michael Neumann¹, Emma Leschly¹, Meredith Bartlett¹,

David Suendermann-Oeft¹ and Vikram Ramanarayanan^{1,2}

¹ Modality.AI, Inc., ² University of California, San Francisco, CA, USA

hardik.kothare@modality.ai

Introduction

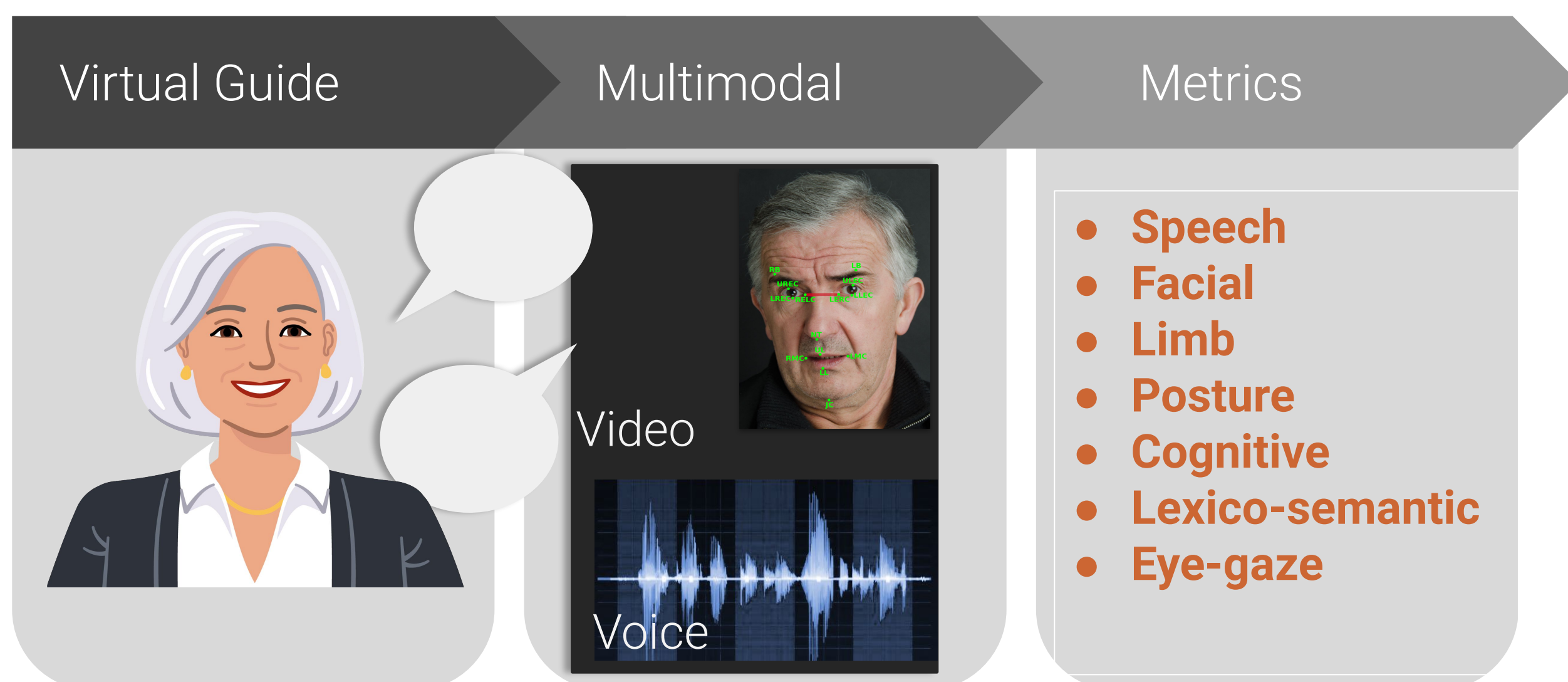


Figure 1. Schematic of the Modality.AI dialogue platform

- Remotely-collected multimodal speech-based digital biomarkers are responsive enough to detect statistically and clinically meaningful changes in **people with Amyotrophic Lateral Sclerosis (pALS)** within 4-8 weeks. These biomarkers are also **sensitive** enough to detect changes before any change on the ALSFRS-R speech score (Kothare et al., Proc. Interspeech 2023).
- Remote collection of speech-based digital biomarkers promises **cost-effective** and **geographically-distributed clinical trials**.

1. Do **objective speech-based biomarkers correlate** with **perceptual ratings of speech impairment and listener effort**?
2. **Can listener effort be predicted** using objective speech-based biomarkers?

Data and Methods

Onset	Number of participants	Mean age \pm SD (years)	Mean time span \pm SD (months)
Bulbar	28 (14 female)	64.31 \pm 8.59	10.1 \pm 7.8
Non-bulbar	97 (47 female)	61.13 \pm 9.41	14.9 \pm 10.9

Table 1: Demographics

Tasks:

- Read speech (**sentence intelligibility test (SIT)**; **Reading Passage (RP)**; Bamboo passage 99 words)
- **Oral diadochokinesis**
- **Single breath counting**
- **Picture description task**

- Data collected from 125 pALS using a **cloud-based multimodal dialogue platform** (Figure 1). Participants were recruited by EverythingALS and the Peter Cohen Foundation.
- **Tina, a virtual guide**, walked participants through **structured speaking exercises** and **objective metrics** were extracted.
- For each participant, their **first and last assessment plus a third sample** that was closest in time to the midpoint of the interval between the first and last session were selected. Six participants had only two sessions, so the final number of samples was $(119 \times 3) + (6 \times 2) = 369$ sessions.
- Using a **visual analogue scale**, three human listeners (two speech scientists and one computer science student) provided a rating of **perceived speech impairment (PSI)** on a scale of 0 (not at all) to 100 (completely impaired).
- Additionally, one clinically-trained speech language pathologist provided a rating of **listener effort (LE)** for each speech sample on a scale of 0 (no effort) to 100 (full effort).
- To ensure consistency and reduce rating time, listeners listened to only the **first 15 seconds of the participants' reading of the Bamboo passage** for every sample.
- We ran **correlation analyses** to investigate the relationship between automatically-extracted metrics and PSI as well as LE.
- Furthermore, we ran **5-fold cross validation linear regression analyses** to **predict PSI and LE values based on Canonical Timing Alignment (CTA) values** - a number between 0% (non-alignment) and 100% (perfect alignment), measured as the normalized inverse Levenshtein edit distance between words and silence boundaries - for the reading passage and sentence intelligibility test.
- **CTA is associated with intelligibility**.

Results and Discussion

- We found that **CTA** - a measure of the participant's predicted word-level timing information of a reading of the passage associated with intelligibility - **correlates moderately to strongly with visual analogue scale ratings of PSI and LE** (see Figure 2).
 - RP CTA and PSI: Pearson's $r = -0.83$, p -value = $4.03e-88$
 - SIT CTA and PSI: Pearson's $r = -0.84$, p -value = $1.23e-96$
 - RP CTA and LE: Pearson's $r = -0.62$, p -value = $1.09e-36$
 - SIT CTA and LE: Pearson's $r = -0.64$, p -value = $1.50e-43$
- **PSI** can be predicted with a mean absolute error of **11.64 percentage points** (Pearson's $r = 0.85$, p -value = $1.22e-103$).
- **LE** can be predicted with a mean absolute error of **5.86 percentage points** (Pearson's $r = 0.67$, p -value = $2.03e-48$).

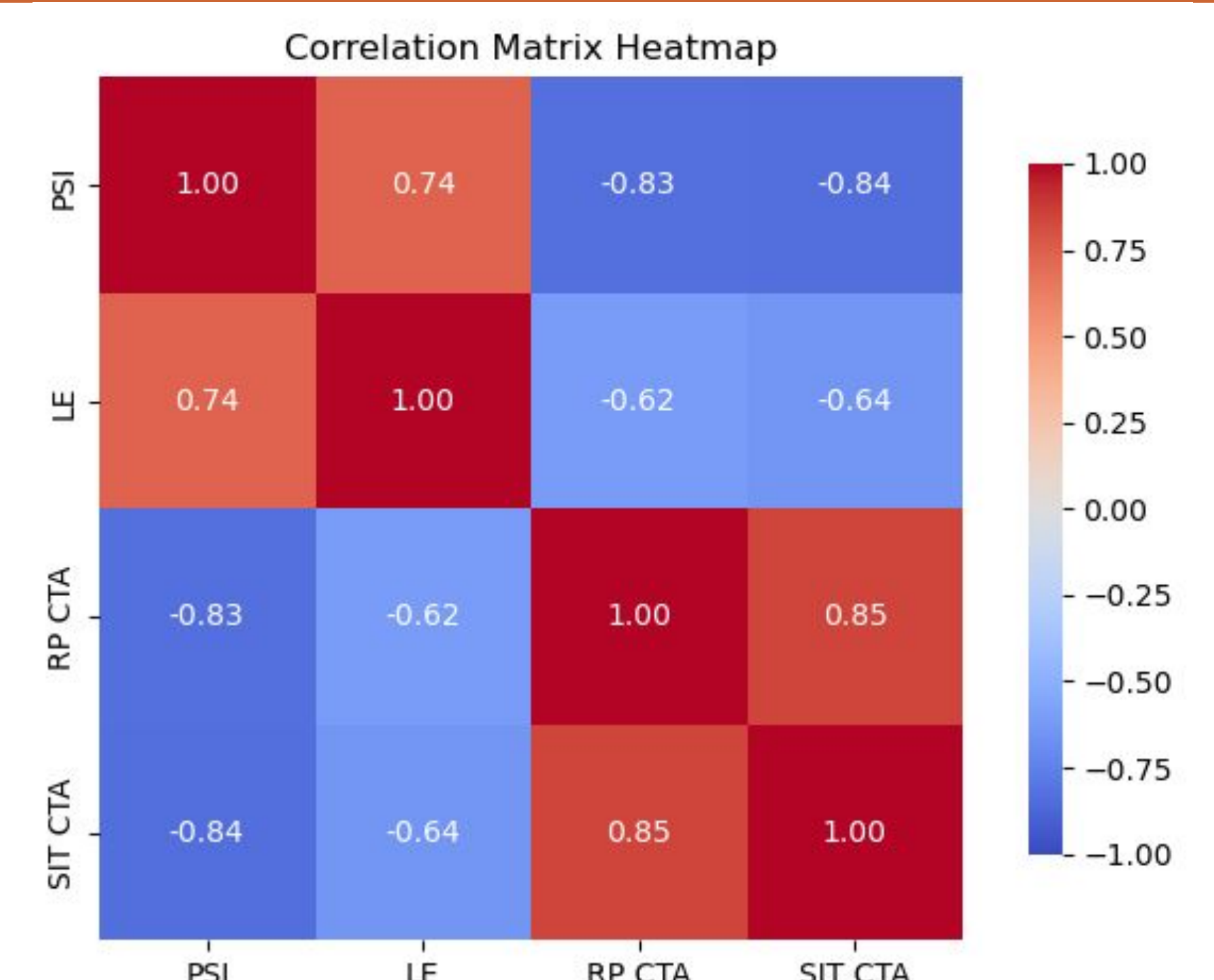


Figure 2: Correlation matrix; PSI = Perceived Speech Impairment, LE = Listener Effort, RP = Reading Passage, SIT = Sentence Intelligibility Test, CTA = Canonical Timing Alignment

Conclusions

The results suggest that speech-based digital biomarkers extracted automatically during remote monitoring of pALS can predict speech impairment and listener effort, both of which are factors contributing to intelligibility and are thus associated with disease progression and quality of life in pALS.