modality.ai

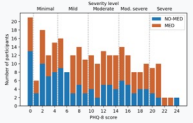
Analytical validation of Canonical Timing Alignment (CTA) and other timing-related speech biomarkers in Amyotrophic Lateral Sclerosis (ALS) extracted automatically using a remote patient monitoring platform

Jackson Liscombe, Reva Bajjuri, Hardik Kothare, **Vikram Ramanarayanan**

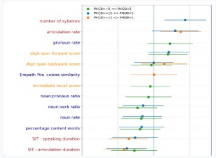
Motivation & Goals

Methods

Data



Findings



Summary

Acknowledgements

This work was supported by the **NIH NIDCD grant R42DC019877**.

We thank our collaborators at EverythingALS and the Peter Cohen Foundation for participant recruitment and data collection.

Motivation & Goals

- Several timing-related speech-based biomarkers of ALS have been validated both clinically and analytically (Barnett et al., ALSFTD 2021):

  speaking duration
  percentage pause time
  articulation rate
  articulation duration
  speaking rate

- Past research has shown the clinical usefulness of canonical timing alignment (CTA) in ALS in terms of responsiveness of bulbar decline and listener effort

What is Canonical Timing Alignment (CTA)?

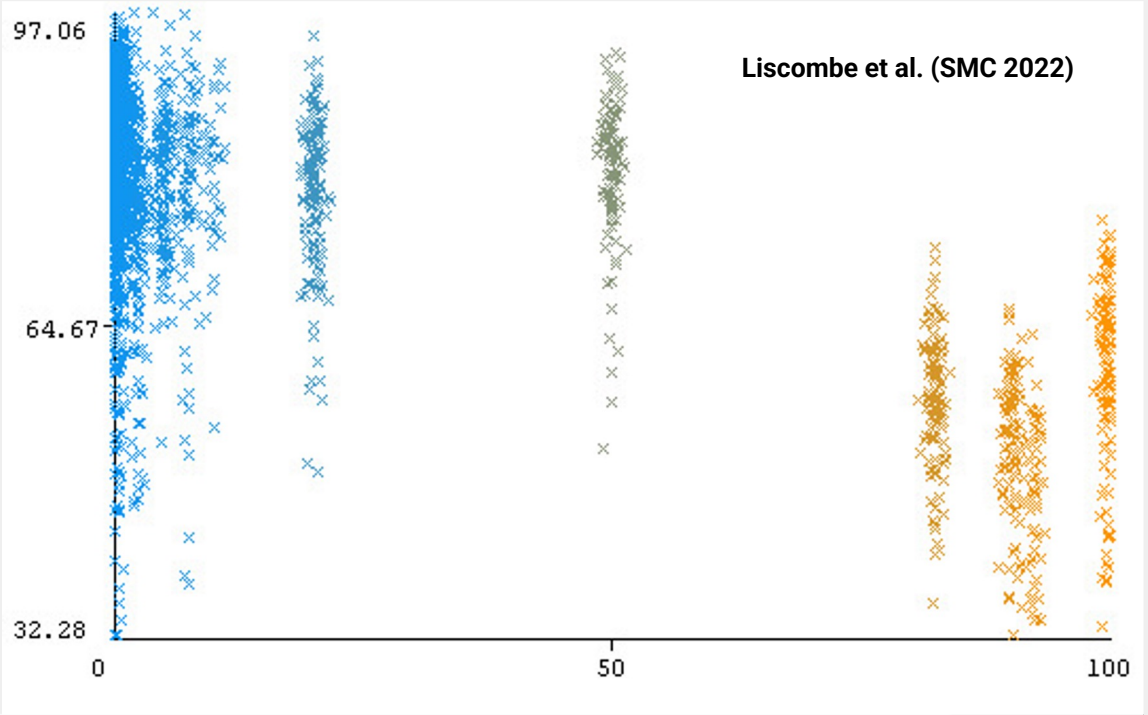
- Given two word-segmented utterances
  - Segment each into sub-word frames (30ms).
  - Calculate the difference using Levenstein edit distance, as the minimum number of frame insertions, substitutions, and deletions needed to align
  - Convert to percentage:

    (max_length - edit_distance) / max_length * 100

CTA captures cohort differences and is strongly correlated with listener effort

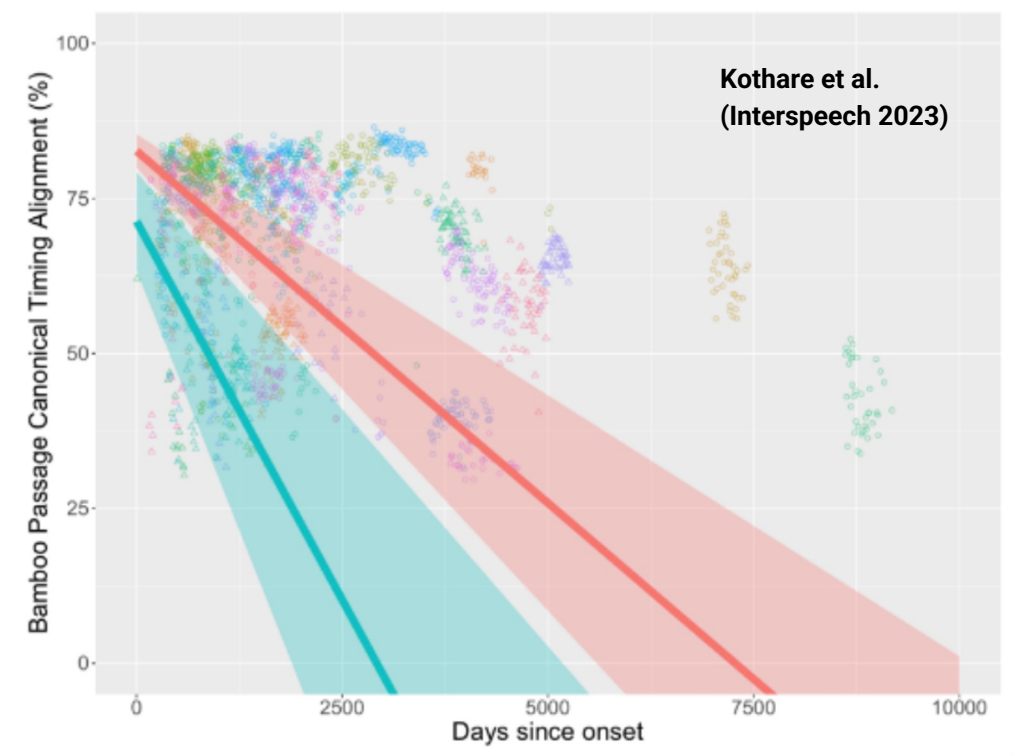
Observed significantly different CTA values on 2,174 SITs between cohorts, as determined by Mann Whitney tests (p<0.00001):

- Bulbar: 66.71%
- pre-Bulbar: 77.31%
- Control: 80.72%



CTA (y-axis) vs Listener Effort (x-axis). Correlation at -0.679

CTA tracks ALS longitudinal progression



Bulbar Onset

Slope = -0.1712 % points / week

Time to detect change > SE = 4 weeks

Time to detect a clinically-important change > 1 point on ALSFRS-R speech score = **4 weeks**

Non-Bulbar Onset

Slope (after accounting for learning effects) = -0.0793 % points / week

Time to detect change > SE = 5 weeks

Time to detect a clinically-important change > 1 point on the ALSFRS-R speech score = **9 weeks**

Research Question

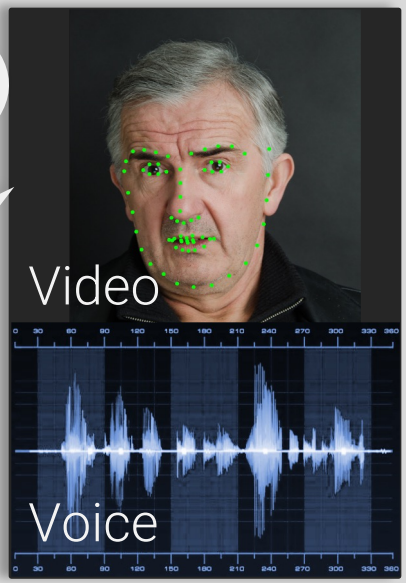
Is CTA extracted from reading tasks administered through a multimodal dialog agent analytically valid ?

Modality's Assessment Platform

Virtual Guide

Tina

Multimodal



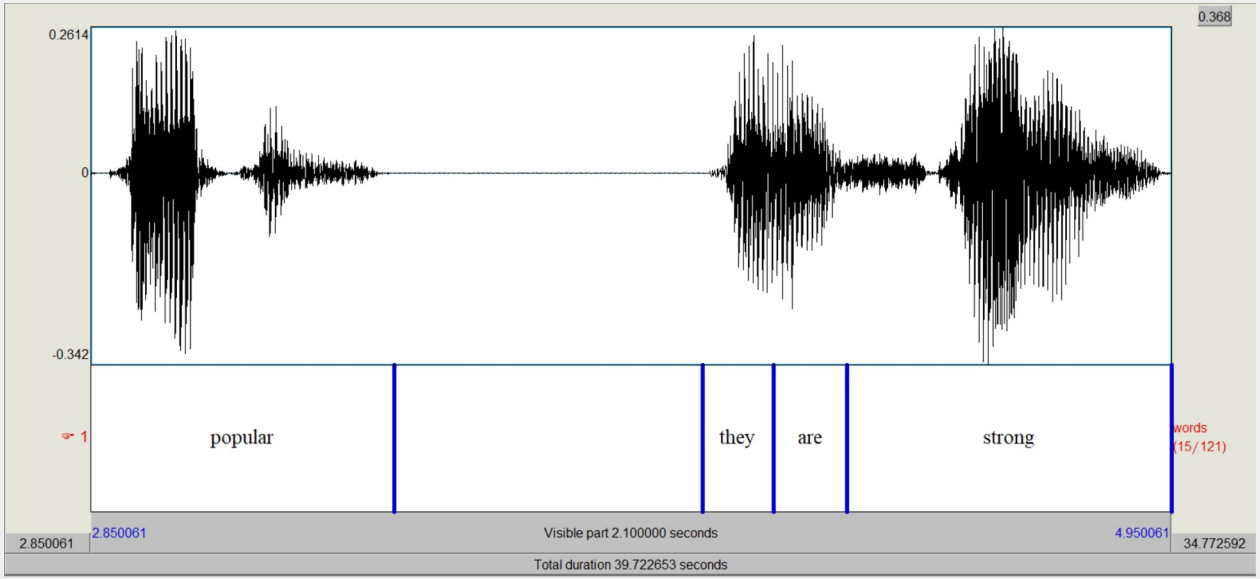
Metrics

- ❑ Speech
- ❑ Language
- ❑ Facial
- ❑ Limb
- ❑ Cognitive
- ❑ Eye gaze
- ❑ Activities of daily living
- ❑ Patient Report of Problems™

Data

- Bamboo Passage, collected in collaboration with EverythingALS
- Selected 1 session from 30 participants, 10 each from 3 cohorts:
  - Bulbar onset patients, pre-Bulbar onset patients, controls
- Three methods of deriving metrics:
  - From hand word alignments from one human annotator (H1)
  - From hand word alignment of a second human annotator (H2)
  - Automatically, including using Montreal Forced Aligner

Methods: Annotation



Annotated word and non-word boundaries using Praat.

Methods: CTA Calculation

- Given two word-segmented utterances
  - Segment each into sub-word frames (30ms).
  - Calculate the difference using Levenstein edit distance, as the minimum number of frame insertions, substitutions, and deletions needed to align
  - Convert to percentage:

    (max_length - edit_distance) / max_length * 100

Methods: Analytic Validation Statistics

- The accuracy of each metric was assessed using:
  - Spearman rank correlation,
  - Mann-Whitney U tests,
  - mean absolute error (MAE),
  - percent error (PE), where PE = MAE / (H1_max-H1_min).

- The same was done for the H1 and H2 datasets to evaluate inter-annotator agreement.

Methods: Clinical Validation Statistics

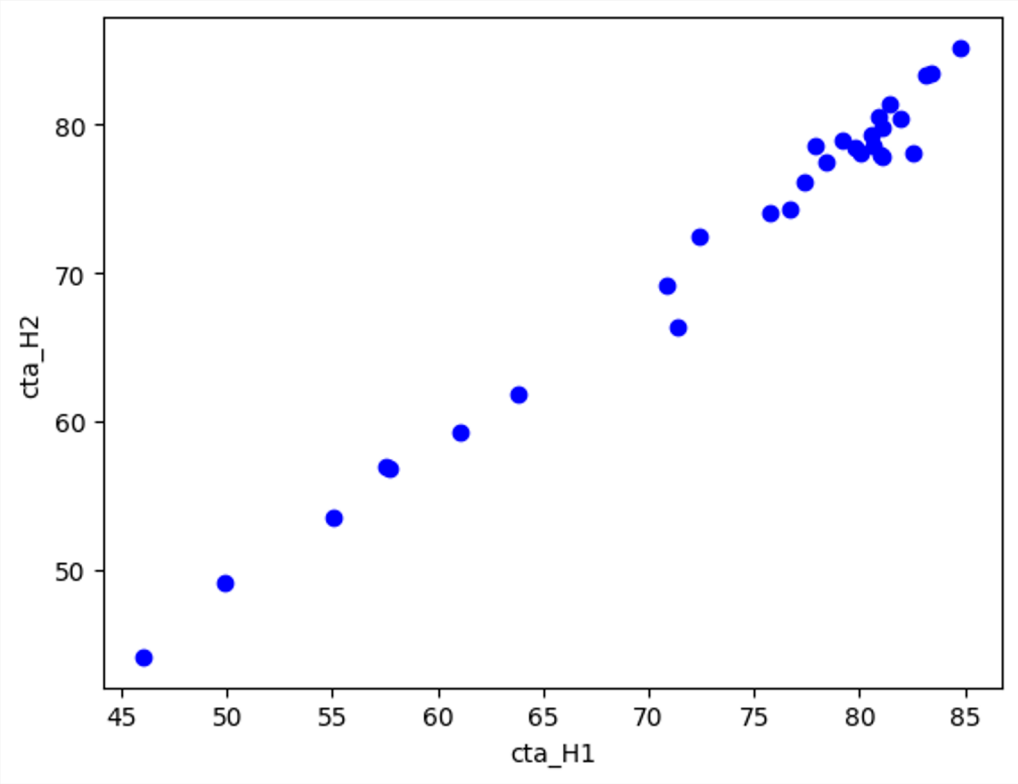
- We analyzed the effect of MAE on clinical validity by:
  - First running a Kruskal-Wallis H test with all cohorts together using the H1 dataset.
  - If significant, we ran Dunn's test for each cohort pair, and Cohen's d to calculate effect size.
  - If Cohen's d > 0.8, this would indicate that this metric has promise of being clinically relevant.

Findings: H1 vs H2

metric name	Spearman coefficient	Spearman p-value	Mann Whitney U p-value	MAE	PE
speaking duration	1	0.00E+00	0.888	0.07 seconds	0.15%
speaking rate	0.996	2.84E-30	0.807	2.51 words/second	1.90%
CTA	0.924	3.08E-13	0.329	1.48 percentage pts	3.83%
articulation rate	0.976	3.49E-20	0.483	5.23 words/second	4.02%
articulation duration	0.987	8.03E-24	0.326	1.35 seconds	4.75%
percentage pause time	0.939	1.82E-14	0.016	3.17 percentage pts	13.09%

CTA was reliable between the two human annotators. Note that percentage pause time would benefit from how certain non-speech events were annotated.

Findings: H1 vs H2, CTA scatter plot



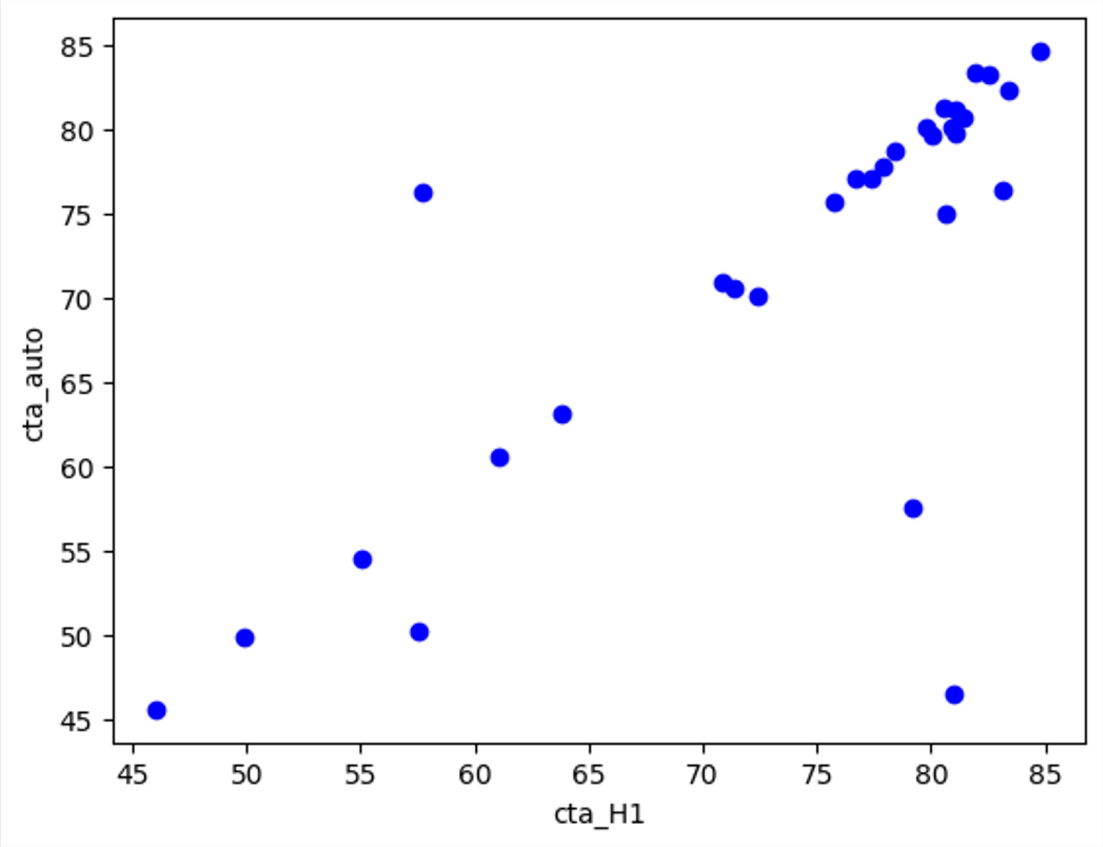
We observe excellent correlation between CTA computed off the alignments performed by the two human annotators.

Findings: H1 vs AUTO

metric name	Spearman coefficient	Spearman p-value	Mann Whitney U p-value	MAE	PE
speaking duration	0.999	4.76E-40	0.739	0.40 seconds	0.91%
speaking rate	0.991	4.53E-26	0.785	2.81 words/minute	2.13%
articulation duration	0.976	3.73E-20	0.751	0.63 seconds	2.22%
articulation rate	0.969	1.28E-18	0.684	4.52 words/minute	3.48%
percentage pause time	0.898	1.69E-11	0.947	1.40 percentage pts	5.77%
CTA	0.775	5.00E-07	0.333	3.62 percentage pts	9.34%

Here we see that CTA showed the highest percent error of all the automated metrics, though the statistics still show significant correlation.

Findings: H1 vs AUTO, CTA scatter plot



Summary

- The CTA metric was analytically validated, as were other speech timing related metrics.

- This is important because we have found that CTA is better at clinically tracking ALS progression than are those other standardized speech timing metrics.

Self-driven Neurological Assessments

Modality.AI Inc.
149 New Montgomery St, 4th Floor, San Francisco, CA 94105

Vikram Ramanarayanan, Chief Science Officer
vikram.ramanarayanan@modality.ai