

Introduction

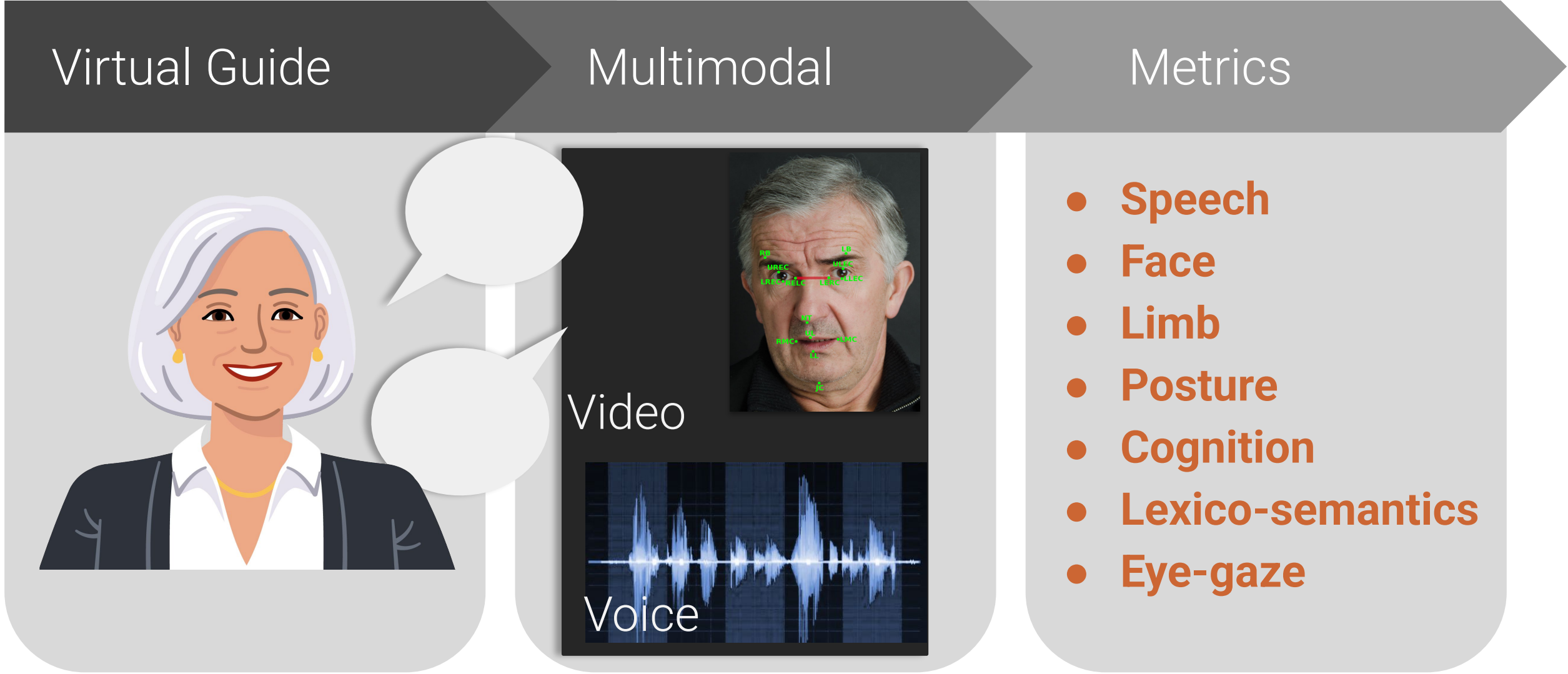


Figure 1. Schematic of the Modality.AI dialogue platform.

- As part of the **Target ALS Global Natural History study**, the **Modality.AI platform** is used to collect speech, facial, and limb motor measures
- Bi-weekly assessments are done **remotely** (at home) using a **web-based multimodal dialog system**
- Novel assessment about **activities of daily living** (ADL) introduced in this multi-site, longitudinal study to assess limb motor function

Objective: Evaluate **feasibility** and **clinical utility** of **digital multimodal biomarkers** compared to traditional clinical endpoints, such as the ALSFRS-R

Data

- Tina, a **virtual guide**, walks participants through a structured set of **speaking exercises and other assessment tasks**
- Standard speech tasks include **sustained vowel phonation**, **diadochokinesis** (DDK), **picture description**, and **reading tasks**
- Data collected (April '25): 87 recording sessions from **19 participants within Target ALS Global Natural History study**, and 58 sessions from **57 healthy controls through crowdsourcing platform Prolific**

	# Participants	# Sessions	Mean Age (SD)	ALSFRS-R* (SD)
pALS	14 (7 female)	75	62.9 (8.5)	32.5 (7.9)
HC	62 (27 female)	70	52.3 (13.6)	-
ALL	76 (34 female)	145	59.7 (11.1)	-

Table 1. Participant statistics. *ALSFRS-R at participants' baseline session.

Feature Extraction

- The Modality platform **automatically** extracts **speech, visual, and linguistic metrics** [1]
- Speech features** extracted with Praat and Montreal Forced Aligner; **Facial features** based on MediaPipe Face Mesh landmarks; **Linguistic features** computed using SpaCy based on automatic transcriptions
- ADL tasks**
 - Participants are asked to mimic **brushing their teeth**, **brushing their hair**, and **washing their face** (without holding any objects)
 - ADL features** are computed based on the geometric centroid of hand landmarks extracted using MediaPipe Hands

	Domain	Features
Speech	Energy	shimmer (%), intensity (dB), signal-to-noise ratio (dB)
	Timing	speaking and articulation duration (sec.), articulation and speaking rate (WPM), percent pause time (PPT, %), canonical timing agreement (CTA, %)
	Voice quality	cepstral peak prominence (CPP, dB), harmonics-to-noise ratio (HNR, dB)
	Frequency	mean, max., min. fundamental frequency F0 (Hz), first three formants F1, F2, F3 (Hz), slope of 2nd formant (Hz/sec.), jitter (%)
Facial	Mouth	lip aperture/opening, lip width, mouth surface area, mean symmetry ratio between left and right half of the mouth
	measurements	velocity, acceleration, jerk, and speed of lower lip and jaw center
	Movement	
	Eyes	number of eye blinks per sec., eye opening, vertical displacement of eyebrows
Text	Lexico-semantic	word count, percentage of content words, noun rate, verb rate, pronoun rate, noun-to-verb ratio, noun-to-pronoun ratio, closed class word ratio, idea density
ADL	Movement	cumulative distance, velocity, acceleration, and jerk of the dominant hand

Table 1. Overview of the extracted features.

Takeaways

- Multimodal biomarkers **distinguish pALS from healthy controls**
- Facial, speech, and linguistic biomarkers** correlate with speech, bulbar and total ALSFRS-R scores, consistent with prior results on other datasets
- Novel limb motor function measures** computed from ADL tasks correlate with ALSFRS-R fine and gross motor sub-scores

Clinical Validation

- Spearman correlations** between features and **self-reported ALSFRS-R outcomes**
- Several **facial features** showed moderate correlations with the **ALSFRS-R bulbar and total score**
- Various **speech features** showed moderate correlations with the **ALSFRS-R speech and bulbar score**
- Linguistic features** showed moderate correlations with the **ALSFRS-R speech or total score**
- All **ADL features** showed moderate correlations with the **ALSFRS-R 6 score** (dressing/hygiene)

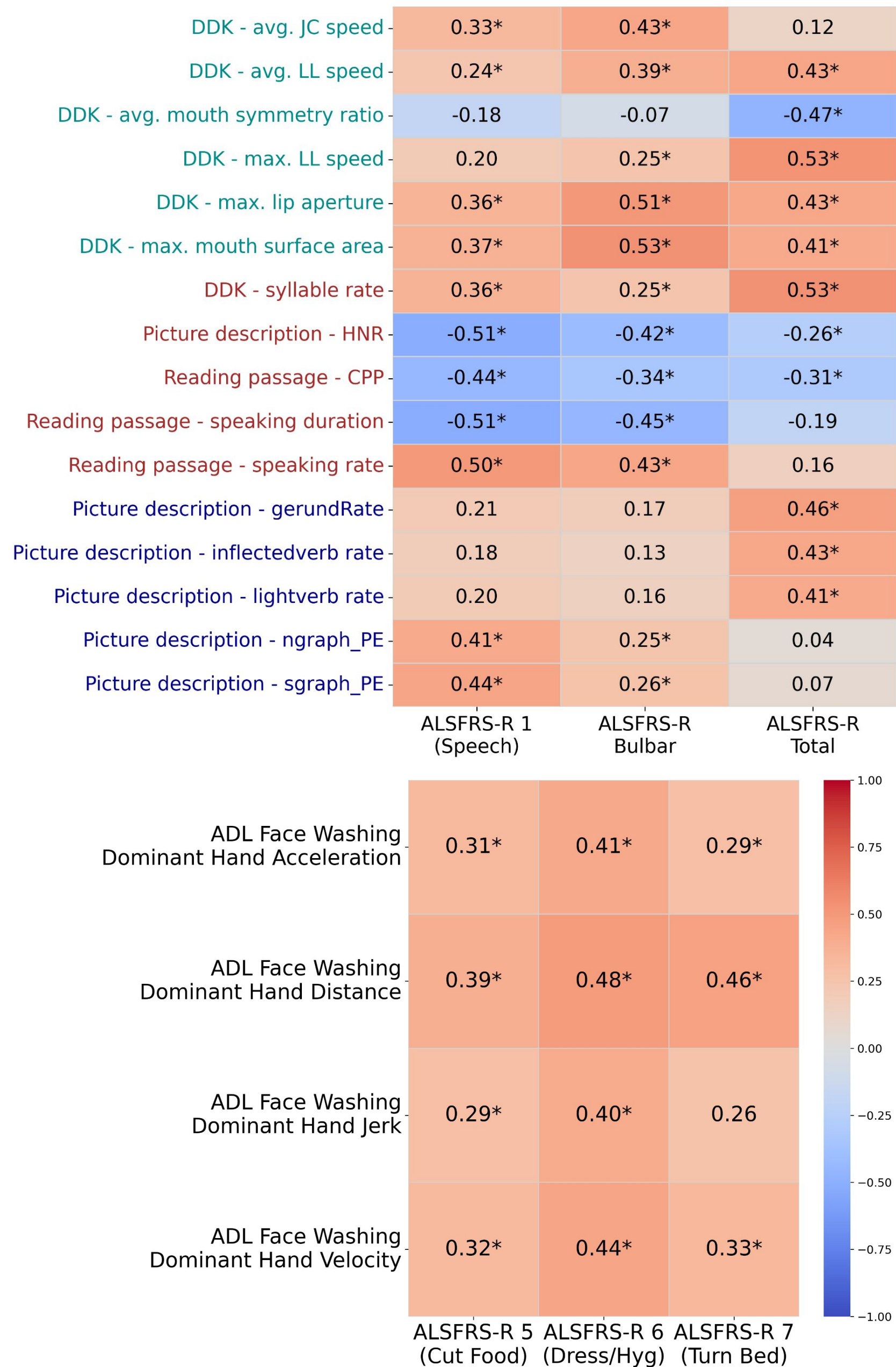


Figure 2. Correlation results for facial, speech, and linguistic metrics (top), and for ADL features (bottom) * indicates p-values below 0.05.

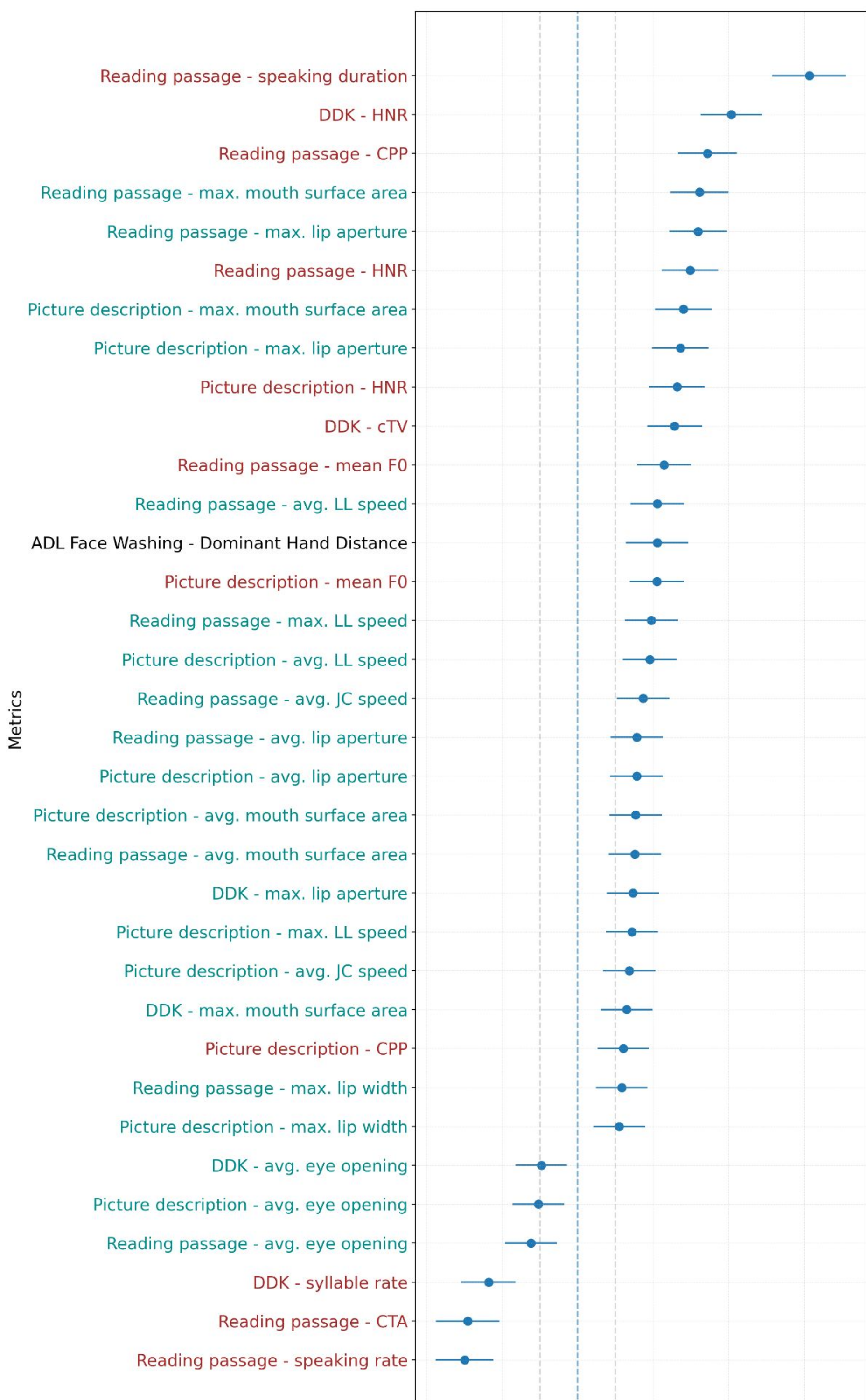


Figure 3. Effect sizes of statistically significant **speech, facial, and ADL** metrics. Positive effect sizes mean larger values for pALS.

- Non-parametric Kruskal-Wallis tests with Benjamini-Hochberg correction** (Q = 0.01) were performed for each feature to determine statistically significant differences between cohorts
- Speech features** yield highest effect sizes for the **Reading passage and DDK task**
- For the **ADL Face Washing task** the **cumulative distance** was significantly different between cohorts

