# Leveraging Remote Digital Health Technologies for Rapid Recruitment and Effective Assessment in Decentralized Clinical Trials: An MCI Case Study

Oliver Roesler[1], Jackson Liscombe[1,2], Michael Neumann[1], Hardik Kothare[1,2], Abhishek Hosamath[1], Lakshmi Arbatti[1], Doug Habberstad[1,2], Christiane Suendermann-Oeft[1,2], Meredith Bartlett[1], Cathy Zhang[1], Nikhil Sukhdev[1], Kolja Wilms[1], Anusha Badathala[2,4], Sandrine Istas[5], Steve Ruhmel[3], Bryan Hansen[3], Madeline Hannan[3], David Henley[3], Arthur Wallace[2,4], Ira Shoulson[1], David Suendermann-Oeft[1], Vikram Ramanarayanan[1,2,4]

[1]Modality.AI, Inc., [2]San Francisco Veterans Affairs Health Care System [3]Janssen Research and Development, LLC. [4]University of California, San Francisco [5]ICON Strategic Solutions, ICON plc.

oliver.roesler@modality.ai

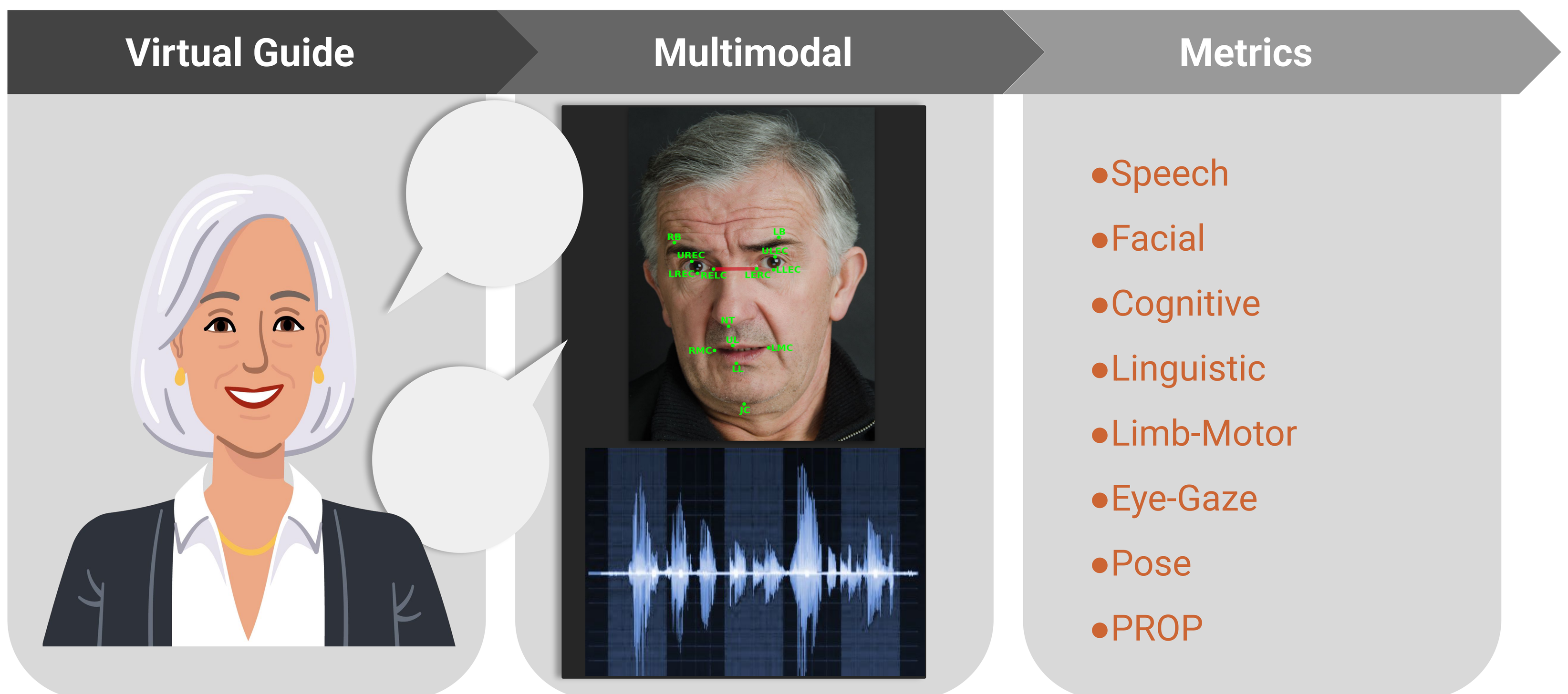


**Figure 1:** Schematic diagram of the Modality.AI dialog platform.

## Takeaways

- The use of a **remote assessment platform** allows for **rapid recruitment** and **high participant retention**.
- Combining multiple **facial, speech, and cognitive biomarkers** allows to **reliably distinguish MCI patients from healthy controls**.
- Asking patients to self-report their most bothersome problems provides **valuable insights about what matters to patients** and how MCI affects their daily life.

## Introduction

- Mild cognitive impairment (MCI) describes **cognitive decline that is stronger than the decline expected due to normal aging**.
- About **10-20% of adults who are at least 65 years** old have MCI.
- Identifying people with MCI has the potential to allow for **early pharmaceutical interventions** before strong damage to the central nervous system has occurred.

- Is it feasible to rapidly recruit an elderly population with MCI for a **remote assessment of speech and cognitive function** through a **multimodal dialog platform** and achieve high retention?
- Can the extracted biomarkers be used to **reliably distinguish MCI patients from healthy controls**?

## Data

- 200 participants (**100 people with MCI and 100 healthy controls**) were recruited via the U.S. Department of Veterans Affairs **within 5 weeks**.
- **2 assessments** (one week apart) per participant administered through the Modality platform.
- During each assessment a **virtual guide named Tina** guides participants through **23 structured exercises** to elicit speech, facial and cognitive behaviors.
- 181 participants completed both assessments leading to a **retention rate of over 90%**.

| Cohort | # Participants | Age (years) |
|---|---|---|
| MCI | 90 (9F / 81M) | 71.08 (9.10) |
| Controls | 91 (9F / 82M) | 71.30 (8.59) |

**Table 1:** Participant demographics. Age is presented as mean (standard deviation).

## Feature Extraction

- The Modality platform **automatically** extracts a variety of speech, facial, and text features in **near-real-time** during the assessment (Tab. 2).
- **Speech features** are extracted using Praat and Kaldi.
- **Facial features** are computed using facial markers extracted by MediaPipe Face Detection and Face Mesh.
- **Text features** were computed using SpaCy based on automatic transcriptions obtained through Amazon Transcribe.
- **Cognitive features** were manually extracted by human annotators after the data collection.

| | 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 measurements | lip aperture/opening, lip width, mouth surface area, mean symmetry ratio between left and right half of the mouth |
| | Movement | velocity, acceleration, jerk, and speed of lower lip and jaw center |
| | 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 |
| | Self-reported problems | reported symptoms, reported problem domains |
| Cognitive | Scores | percentage of correct words (immediate and delayed word recall), digit span forward/backward score (ranges from 0 to 2) |
| | Timing | response latency (sec.), response duration (sec.) |

**Table 2:** Overview of the extracted features across modalities.

## Feasibility Analysis

- Participants were asked to rate **different aspects of the interaction** on a 5-point Likert scale.
- The **majority of participants** rated most aspects of their interaction with the system as either "**Satisfactory**" (4) or "**Very satisfactory**" (5).

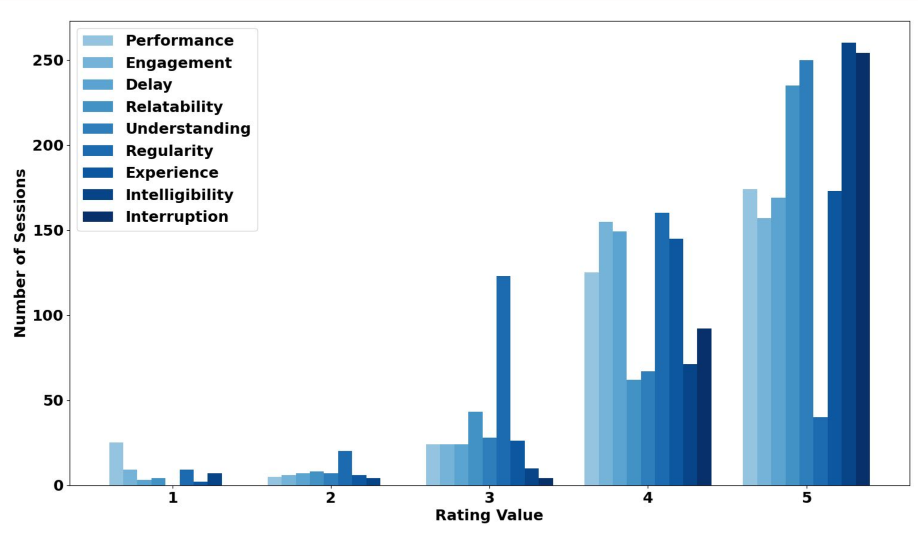


**Fig. 2:** Bar chart illustrating the results of the UX survey.

## Clinical Validation

- **Non-parametric Kruskal-Wallis tests** were performed for each individual feature to determine which of them show a statistically significant difference (α = 0.01) between cohorts.
- **Pearson correlations** were computed between features of **participants' subsequent sessions** to assess the **test-retest reliability** of the features.
- **Facial features showed the strongest signal** with mostly acceptable or good reliability (Fig. 3).
- Four different classifiers were employed for **binary classification experiments** using 5-fold cross-validation.
- The **selected feature set had a strong influence** on classification performance, while the selected classifier had only a small influence (Tab. 3).
- Using only the features that showed statistically significant differences between cohorts led to the **best AUC of 0.75** using a SVM classifier (Fig. 4).
- Each self-reported problem (PROP) was **automatically transcribed** via Amazon Transcribe.
- Transcriptions were provided to an **inference model** developed from a neural network with two hidden layers to identify which of **65 symptoms** were described by the participant.
- Symptoms were automatically grouped into **14 domains**.
- MCI patients reported more **cognition, gait, psychiatric, and sleep problems** (Fig. 5).
- MCI patients reported nearly 2 times more problems with **anxiety or worry and speech**, nearly 4 times more problems with **memory**, and 6 times more problems with **falls** than healthy controls (Fig. 6).

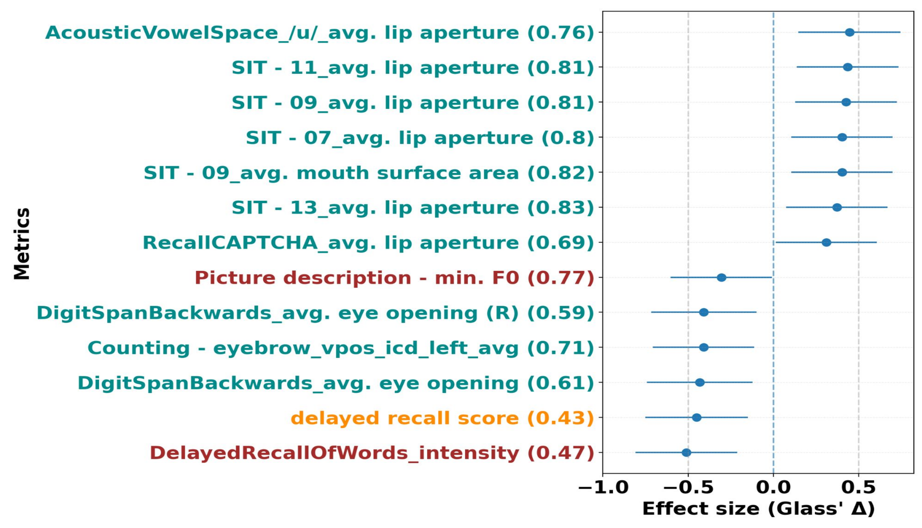


**Figure 3:** Effect sizes and test-retest reliabilities of statistically significant (α = 0.01) speech, facial, and cognitive features. Positive effect sizes mean larger values for MCI patients.

| Feature Set | Classifier LR | RF | MLP | SVM |
|---|---|---|---|---|
| speech only | 0.52 | 0.53 | 0.55 | 0.53 |
| facial only | 0.59 | 0.58 | 0.59 | 0.54 |
| text only | 0.62 | 0.59 | 0.62 | 0.61 |
| cognitive only | 0.57 | 0.58 | 0.55 | 0.55 |
| combo - all | 0.57 | 0.56 | 0.61 | 0.56 |
| combo - significant | **0.75** | **0.73** | **0.69** | **0.75** |

**Table 3:** Classification performance as measured by area under the ROC curve (AUC) across multiple classifiers (LR: Logistic Regression; RF: Random Forests; MLP: Multi-layer Perceptron; SVM: Support Vector Machine) and feature sets.

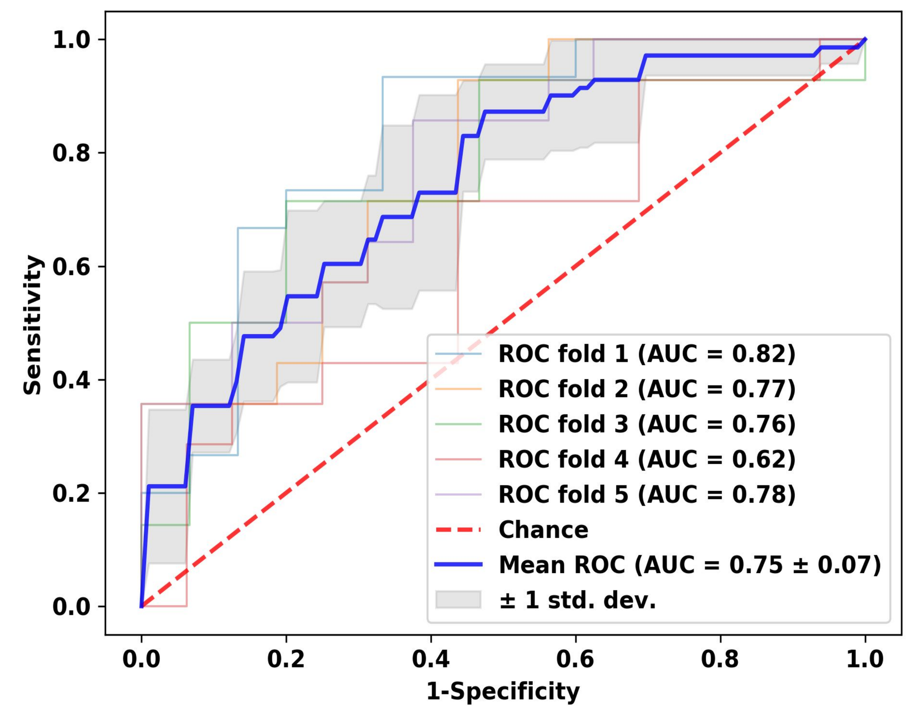


**Figure 4:** Binary classification results with 5-fold cross-validation using the 13 statistically significant feature as input to a support vector machine (SVM).

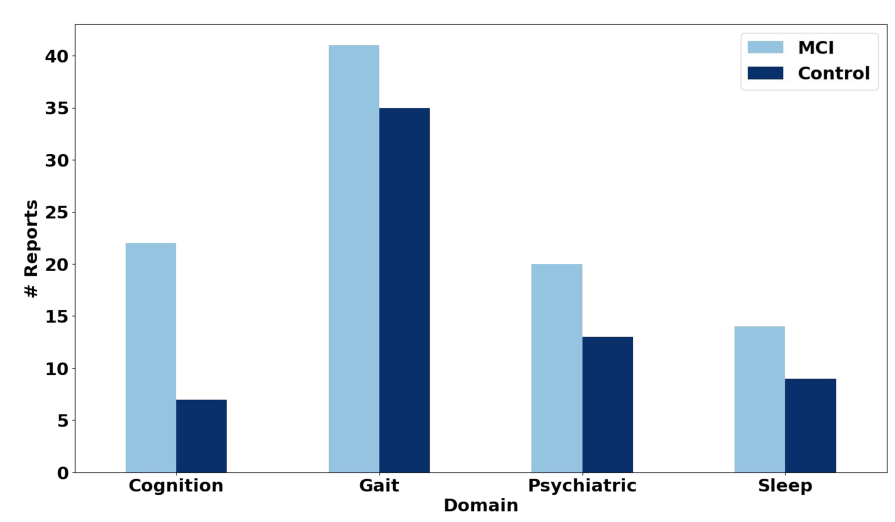


**Figure 5:** Self-reported problem domains affecting daily functioning.

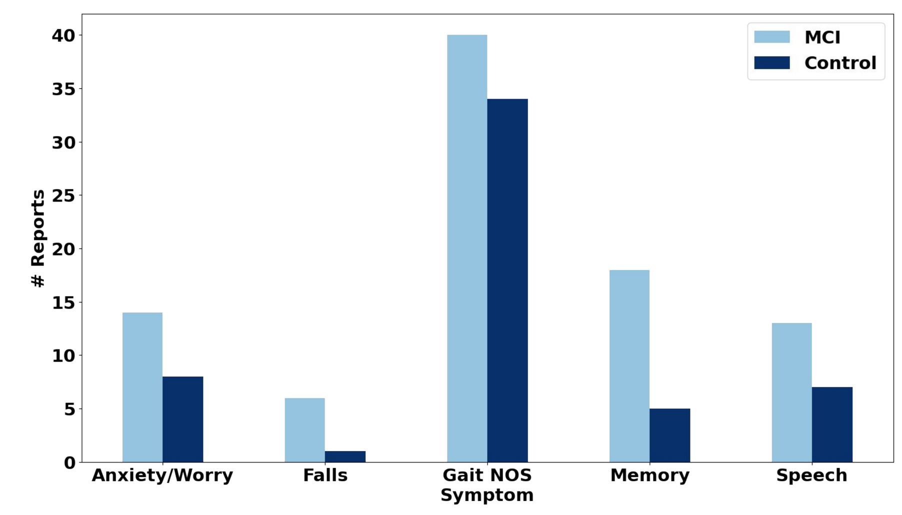


**Figure 6:** Self-reported symptoms affecting daily functioning.