Leveraging Multimodal Dialog Technologies & AI/ML for Patient Health Diagnosis, Monitoring, and Intervention

Vikram Ramanarayanan, David Suendermann-Oeft, David Pautler, Oliver Roesler, Jackson Liscombe, Michael Neumann, Doug Habberstad & Andrew Cornish

modality.ai

UCSF University of California San Francisco

ALCHEMIST ACCELERATOR

Optimize drug trials through multimodal AI

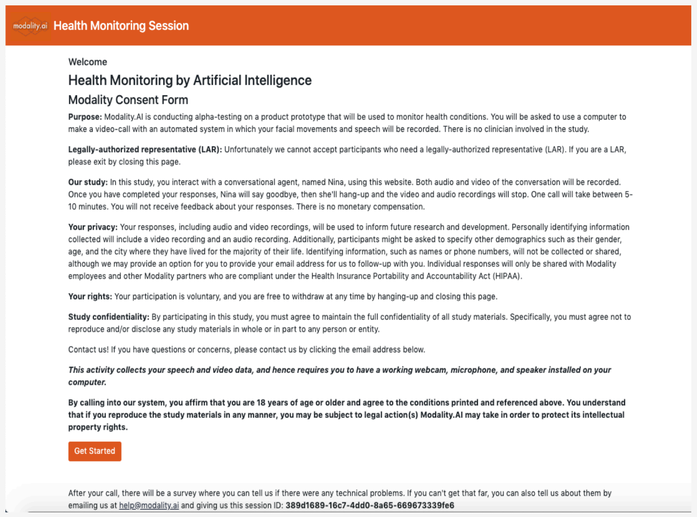
Using conversational speech and video technologies in conjunction with signal processing and machine learning algorithms can allow patients to interact with an engaging virtual agent in the comfort of their homes, producing various automatically-computed, clinically-validated measures of disease onset and progression that can help those patients, as well as their supervising physicians and caregivers in diagnosis, monitoring and understanding.

Diagnosis & monitoring of neurological health in patients remain a critical need and challenge

- lack of access to neurologists or psychiatrists
- lack of awareness of a given condition and the need to see a specialist
- lack of an effective standardized diagnostic or endpoint for many of these health conditions
- substantial transportation and cost involved in conventional or traditional solutions
- severe shortage of medical specialists in these fields to begin with

NEMSI bridges this gap!

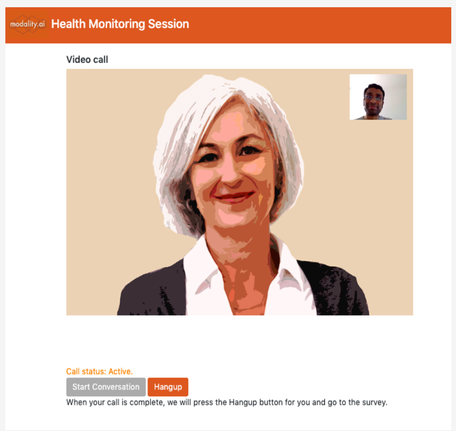
- NEurological and Mental health Screening Instrument
  - cloud-based multimodal dialog system that conducts automated screening interviews
  - over the phone, smartphone app, or web browser
  - elicits evidence required for detection or progress monitoring
- NEMSI makes novel contributions along three significant directions:
  - uses end point devices available to everyone everywhere
  - deployed in an automatically scalable cloud environment allowing it to serve an arbitrary number of end users at a very small cost per interaction
  - natively equipped with real-time speech and video analytics modules that extract a variety of features of direct relevance to clinicians

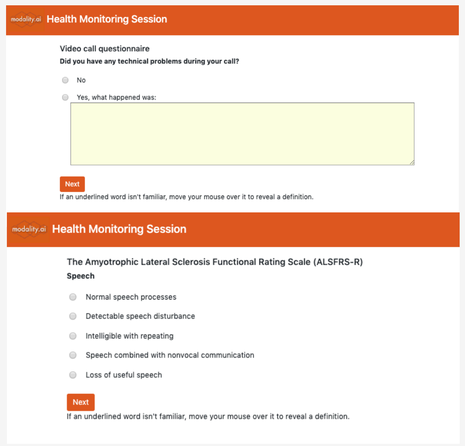


Patients/clinicians are provided with a secure web link to the assessment

Users guided to set up their webcam and microphone before interacting with agent

Users interact with an agent and fill out clinical surveys





Analytics modules extract speech and video metrics automatically

- Speech:
  - speaking rate
  - intelligibility
  - mean pausing duration
  - mean F0
  - duration
- Video:
  - opening, width, displacement, velocity, acceleration and jerk of the upper and lower lips
  - mouth surface
  - mean symmetry ratio between right and left mouth surfaces
  - vertical positions of eyebrows
  - eye opening
  - head tilt translation and rotation vectors

Verification of speech and video metrics

- Verification corpus (11 interactions) collected using AMT
- Automatic extraction of speech metrics through voice activity and pitch detection algorithms
- Automatic video brightness and blurriness checks to instruct users to fix low video quality
- Automatic face and facial landmark detection
- Calculation of facial metrics using 14 of the 68 detected facial landmarks
- Ground truth labels obtained through manual annotations of:
  - number of words and speech duration
  - facial markers required for facial metrics calculation
- Preliminary results show the automatically-computed metrics are very similar, i.e. above 80%, to the ground truth

Dashboard provides easy visualization!