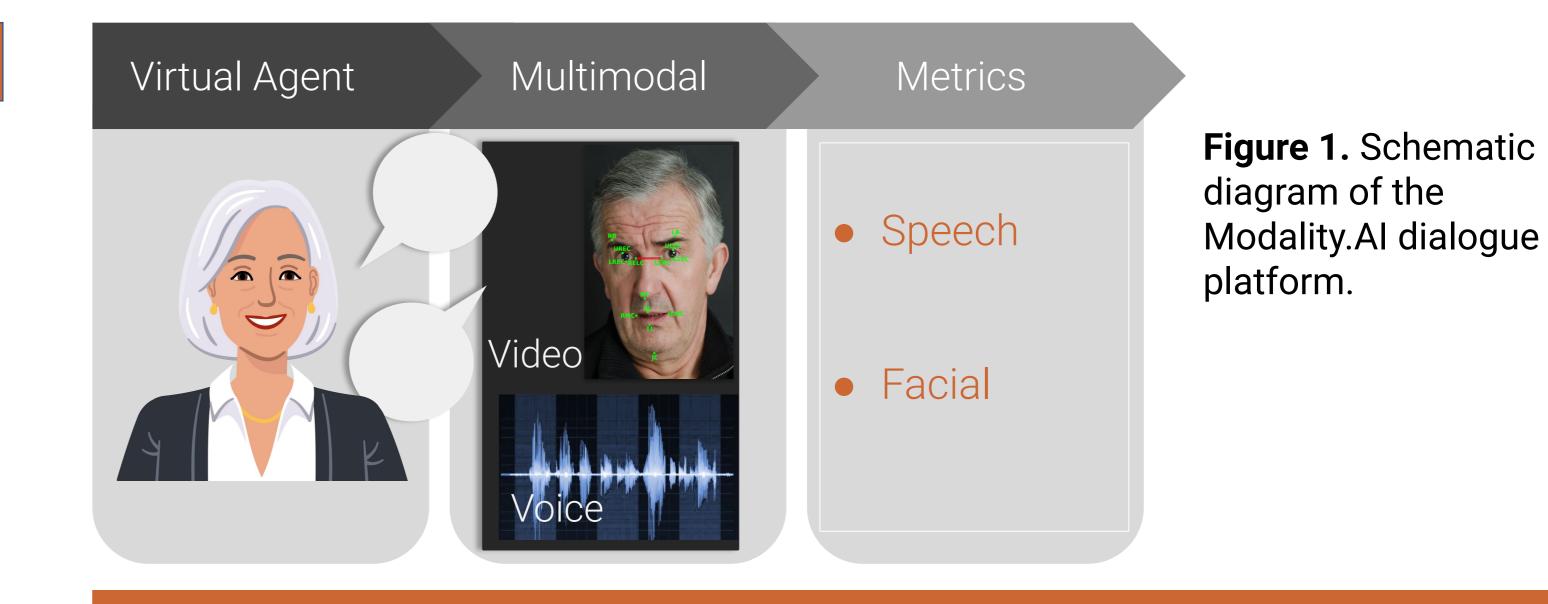
Differentiating primary and secondary expressive negative symptoms for remote digital trials using a multimodal dialogue platform

Hardik Kothare¹, Michael Neumann¹, Vanessa Richter¹, Oliver Roesler¹, Jackson Liscombe¹, Anzalee Khan², Sandy Snyder³, Christian Yavorsky⁴, Benedicto Parker², Theresa Abad², Jessica E Huber³, Jean-Pierre Lindenmayer^{2,5} and Vikram Ramanarayanan^{1,6} ¹ Modality.AI, Inc.,² Nathan S. Kline Institute for Psychiatric Research, ³ Purdue University, ⁴ Valis Bioscience, ⁵ New York University, School of Medicine, ⁶ University of California, San Francisco, CA, USA hardik.kothare@modality.ai

Introduction

- Primary expressive negative symptoms (PENS) such as alogia, anhedonia and social withdrawal are cardinal features of schizophrenia and are persistent through a patient's lifetime.
- Pharmacological side effects of antipsychotic treatment in schizophrenia also cause secondary expressive negative symptoms (SENS) that bear a resemblance to symptoms in Parkinson's disease (PD), such as reduced facial expression, bradykinesia and impaired speech.
- Clinically, PENS and SENS are indistinguishable, but SENS can be ameliorated through correct identification and timely treatment.



• A delay in identification of SENS features can result in those new symptoms becoming permanent, significantly **reducing** communication effectiveness for people with schizophrenia.

Research Question:

Can automatically-extracted speech and facial metrics through a dialogue-based remote patient monitoring platform differentiate between presentations of primary and secondary expressive negative symptoms, beyond any differences relative to healthy control participants?

Data and Methods

	Number of participants	Number of sessions	Mean sessions per participant ± SD	Mean age ± SD (years)
pSz	48 (12 female)	98	2 ± 0.7	39.2 ± 10.9
pPD	49 (23 female)	226	4.6 ± 1.7	69.1 ± 8.3
Controls	92 (50 female)	252	2.7 ± 1.5	53.2 ± 16.7

Results

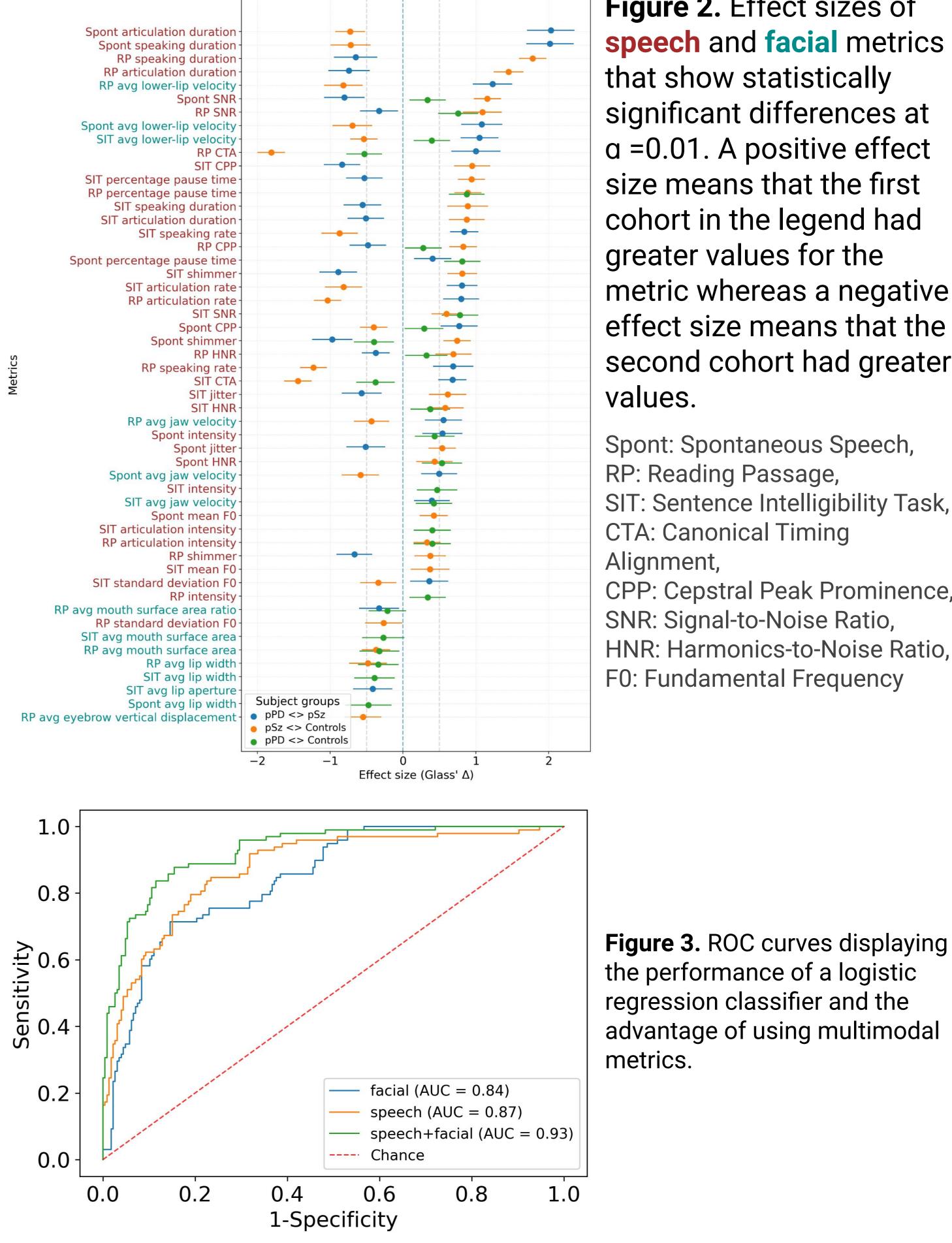


Figure 2. Effect sizes of **speech** and **facial** metrics that show statistically significant differences at a =0.01. A positive effect size means that the first cohort in the legend had greater values for the metric whereas a negative effect size means that the second cohort had greater

Spont: Spontaneous Speech, **RP: Reading Passage,** SIT: Sentence Intelligibility Task, CTA: Canonical Timing **CPP: Cepstral Peak Prominence**, SNR: Signal-to-Noise Ratio, HNR: Harmonics-to-Noise Ratio, F0: Fundamental Frequency

 Table 1: Demographics

• Using a cloud-based multimodal dialogue platform (Illustration in Fig. 1), data was collected from

48 people with Schizophrenia (pSz),

49 people with Parkinson's disease (pPD, a proxy cohort for SENS because of the large overlap in clinical presentation), and 92 healthy controls (HC)

- Participants were guided by **Tina**, a virtual agent, through structured speaking exercises and **objective speech and facial metrics** were extracted.
- Each study's protocol was designed to assess characteristics typical to the patient population of interest. For this study we chose three tasks common to both protocols:
 - Sentence Intelligibility Task (SIT): Participants were asked to read sentences when prompted by the virtual agent.
 - **Reading Passage**: Participants were asked to read a passage when prompted by the virtual agent. Participants in the schizophrenia protocol read the Bamboo Passage (99 words) and participants in the PD protocol read the Rainbow Passage (98 words). The similar lengths of these passages allow for a comparison of speaking
- People with PENS, i.e. pSz, exhibited lower range and velocity of facial motion, slower rate of speaking and poorer voice quality than people with SENS, i.e. pPD, over and above any differences from HC.
- A logistic regression classifier performed well in classifying people with PENS and SENS based on **speech metrics (AUC = 0.87)** and **facial metrics (AUC = 0.84)** alone, but the performance improved when a

duration and other measures.

- Spontaneous speech: Participants were asked to speak about a topic of their choice (with suggestions on the screen like favourite sports, books, your children, etc.) for at least 30 seconds.
- We ran **non-parametric Kruskal-Wallis tests** to identify differences between pairs of cohorts (pSz <> pPD, pSz <> HC, pPD <> HC) at a = 0.01. Effect sizes were computed as Glass's delta (false discovery) rate was controlled with Benjamini-Hochberg procedure).
- We evaluated the performance of a **logistic regression classifier**, using a leave-one-speaker-out method, in binary classification of the pPD and pSz cohorts based on speech metrics alone, facial metrics alone and a combination of speech and facial metrics.

combination of speech and facial metrics was used (AUC = 0.93).

Conclusions

- We find that **speech and facial metrics automatically extracted** through a dialogue-based RPM platform are useful in distinguishing between PENS and SENS in digital clinical trials, over and above any differences from healthy controls.
- This work serves as **proof of concept** to develop **digital phenotypes** for primary and secondary expressive negative symptoms in schizophrenia, allowing for timely identification and treatment.



