Using AI-driven platform to Detect Negative Symptoms of Schizophrenia Through Facial and Acoustic Analysis

Anzalee Khan[1,2]; Jean-Pierre Lindenmayer[1,3]; Vikram Ramanarayanan[5]; Hardik Kothare[5]; David Paulter[5]; Mohan Parak[1,2]; Benedicto Parker[1,2];Christian Yavorsky[4]; David Suendermann-Oeft[5]

Affiliations: 1. Nathan S. Kline Institute for Psychiatric Research, Orangeburg, NY; 2. Manhattan Psychiatric Center, NY, NY; 3. New York University, School of Medicine, NY, NY; 4. Valis Bioscience, Berkeley, CA; 5. Modality.AI Inc., San Francisco, CA

METHODOLOGICAL QUESTION

- The automatic analysis of facial and acoustic expressions is an evolving field that finds several clinical applications. One of these applications is the study of facial and speech productions in individuals with schizophrenia, which is a major indication of negative symptoms of this illness. One feature of Negative symptoms is the reduction of facial movements and emotional facial expressions and is a transdiagnostic feature of schizophrenia.
- Current methods of assessing negative symptoms depend on verbal report from patients and/or caregivers and a clinical interview. These interviews can be insensitive to change in treatment, subjective, requires extensive training and subject to cultural disparities.
- Facial and speech changes in negative symptom may be difficult to track and quantify with only interview methods. With passive (real-world) data collection, combined with patented algorithms and machine learning, we are starting to see examples of AI-driven biosensors that can predict early signs of diseases in advance of an actual event. Similarly, digital therapeutics platforms can be used to supplement clinical interviews for more objective and precise measurements.
- Would a novel artificial intelligence (AI) system analyzing facial and acoustic features improve measurement of negative symptoms in schizophrenia?

AIMS

- **Aim 1**: To investigate whether negative symptoms can be meaningfully measured using AI-enabled vocal and facial analysis software called Neurological and Mental Health Screening Instrument (NEMSI) by comparing speech metrics (e.g., prosody, rate, intelligibility, pausing duration etc.) and video metrics (e.g., specific facial and head movements) to clinician-rated psychometric assessments for negative symptoms.
- **Aim 2**: To investigate the feasibility and user experience (patient) of NEMSI through system acceptability, usability, engagement, and benefits; and to identify if participants' negative symptoms, and levels of persecutory ideation would impact their use of the system.

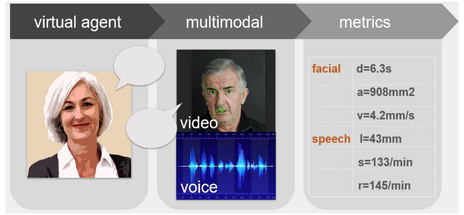
METHOD

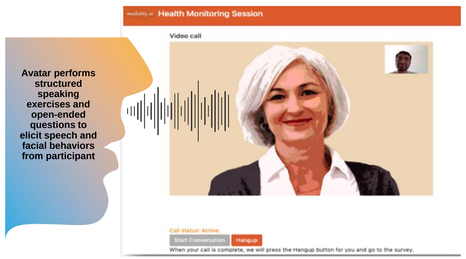
- Experimental Approach: At the first visit, the following instruments are administered: sociodemographic and clinical questionnaire, PANSS, BNSS, CDSS, CGI-S, AIMS, SAS, BARS and NEMSI. The second visit occurs within a one-week period and is done by the same clinician to assess for test-retest reliability and intra-rater reliability. The second visit includes the same instruments in addition to the CGI-I (severity of illness, improvement, and degree of change). Healthy controls only performed the NEMSI.
- Patient Eligibility: Inpatients with diagnosis of schizophrenia, age 18 - 60, English speaking, WRAT-IV Reading Score ≥ 8th grade, Negative symptoms as evidenced by score of ≥ 18 on PANSS Marder Negative Symptom Factor
- Healthy Control Eligibility: Individuals with no prior history of mental illness, age 18 - 60, English speaking.
- Analysis: Reliability (ICC), concurrent, convergent, divergent and discriminative validity of NEMSI speech and facial metrics to the BNSS, PANSS Marder Negative factor and the CDSS

SPEECH, VOCAL AND FACIAL AI PROGRAM

Computer-based negative symptom measure: For NEMSI, participants interact with an avatar that provides a series of emotionally-ambiguous, valence-neutral tasks including a series of reading aloud tasks composed of sentences and a passage; an eyebrow raising task, an image description task, and a free speech task related to a topic of interest from the list provided.

The session takes 8–10 minutes to complete, during which the software produces facial and vocal metrics.



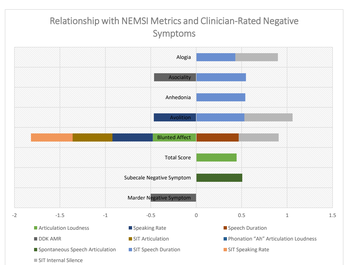


Speech and Facial Data from the program includes:

- Phonation
- Cepstral Peak Prominence (level of noise in vocal signal, measures dysphonia)
- Speech Intelligibility (SIT), Duration, and Rate (with and without pauses)
- Articulation Rate and Loudness
- DDK - also known as syllable alternating motion rate (AMR), assesses repetitive movements of oral articulators
- Internal Silence (pauses)
- Syllable Rate and Count
- Lip Aperture
- Mouth Surface Area
- Jaw Velocity and acceleration
- Lower Lip Velocity and Acceleration
- Eye Opening and Eyebrow vertical position
- Head Tilt

RESULTS: SPEECH AND VOCAL METRICS

	PANSS		BNSS					
	Marder Negative Symptom	Subscale Negative Symptom	Total Score	Blunted Affect	Avolition	Anhedonia	Asociality	Alogia
Speech Metrics								
Articulation Loudness			0.445	-0.478				
Speaking Rate				-0.444	-0.466			
Speech Duration				0.467				
DDK AMR	-0.502						-0.463	
SIT Articulation				-0.437				
Phonation "Ah" Articulation Loudness								0.486
Spontaneous Speech Articulation		0.506						
SIT Speech Duration					0.53	0.54	0.548	0.431
SIT Speaking Rate				-0.459				
SIT Internal Silence				0.44	0.53			0.468



Relationships with NEMSI Facial and Speech Metrics and Clinician-Rated Assessments

- Speech Articulation is how clearly the speaker pronounces words. When some sounds are slurred together or dropped out of a word, the word may not be understood
  - The loudness of speech articulation was positively related to the BNSS Total Score
  - The phonation of articulation loudness was positively related to Alogia (poverty of speech)
  - Spontaneous Speech was positively related to the PANSS Negative Symptom Subscale
- Speech Intelligibility (SIT) refers to how well someone can be understood when they're speaking
  - Speaking Rate was negatively correlated with Blunted Affect (better speaking rate, less blunted affect)
  - Speech Duration was positive correlated with Avolition, Anhedonia, Asociality and Alogia
  - Internal Silence was positive correlated with Blunted Affect, Avolition and Alogia

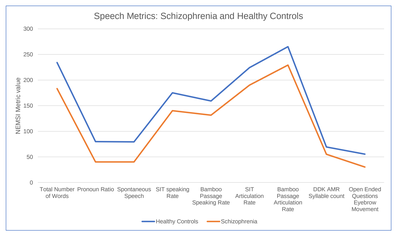
BASELINE DEMOGRAPHICS

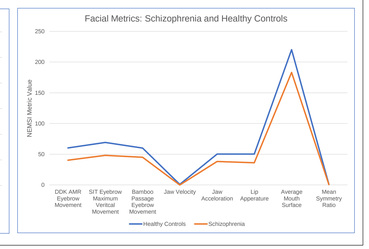
Characteristics	Schizophrenia		Healthy Controls	
	Mean	SD	Mean	SD
Age (in years)	39.95	11.69	42.14	12.19
	n	%	n	%
Gender				
Male	29	82.85	19	61.29
Female	6	17.14	12	38.71
Race				
Black	25	71.43	19	61.29
White	10	28.57	10	32.25
Asian	0	0	0	0
Other	0	0	2	6.45
Ethnicity				
Hispanic	17	48.57	20	64.52
Non-Hispanic	18	51.42	10	32.25
Not reported	0	0	1	3.22

PANSS	n	Mean	SD
Positive Subscale	35	19.23	4.34
Negative Subscale	35	24.38	3.12
General Psychopathology	35	39.99	6.10
PANSS Total	35	82.23	11.00
Marder Positive Symptom	35	26.11	3.93
Marder Negative Symptom	35	22.67	2.67
Marder Disorganized Symptom	35	18.22	3.91
Marder Hostility Symptom	35	7.90	2.37
Marder Anxiety Symptom	35	6.23	2.33

Schizophrenia compared to HC

A significant difference ($p < 0.05$) was observed between patients and HCs for most NEMSI metrics. Significant metrics are presented in the charts below





RESULTS: FACIAL EXPRESSION AND GESTURES

	PANSS		BNSS					
	Marder Negative Symptom	Subscale Negative Symptom	Total Score	Blunted Affect	Avolition	Anhedonia	Asociality	Alogia
Facial Metrics								
Jaw Velocity	-0.445	-0.433	0.481					
Articulatory: Total Mouth Surface	-0.439	-0.418		-0.349				
Eyebrow Movement	-0.449	-0.341	-0.445	-0.558				
Lip Aperature	-0.441	-0.334	-0.44			-0.339		

Relationships with AI Facial Metrics and Clinician-Rated Assessments

- Facial Features
  - Jaw velocity (mean and maximum speed of the jaw) was negative correlated with PANSS Negative symptoms and BNSS Total Score
  - Total Mouth Surface area and eyebrow movement was also negatively correlated with PANSS Negative Symptoms, the BNSS Total Score and Blunted Affect
  - Lip Aperature was negatively correlated with PANSS Negative Symptoms, BNSS total and Anhedonia

RESULTS: RELIABILITY AND VALIDITY

- Reliability NEMSI AI (Time 1 and Time 2): ICC = 0.981
- Reliability PANSS Marder Negative Symptoms (Time 1 and Time 2): ICC = 0.954
- Reliability BNSS Total Score (Time 1 and Time 2): ICC = 0.954
- Validity of NEMSI with 1. BNSS Total Score = 0.804, 2. PANSS Marder Negative Symptom = 0.803, BNSS Alogia = 0.812, BNSS Avolition = 0.844
- Internal Consistency of NEMSI: 0.877 Test-Retest Reliability NEMSI: $p < 0.01$

CONCLUSIONS

- Speech and facial AI technology could aid in negative symptoms assessments
- The NEMSI showed adequate reliability, validity, and internal consistency
- Additional testing on larger sample sizes, reproducibility, and generalizability of the software is warranted.