Multimodal AI of Facial and Acoustic Biomarkers of Negative Symptoms in Schizophrenia

Anzalee Khan[1,2]; Jean-Pierre Lindenmayer[1,2,3]; Sebastian Prokop[1,2,4]; Saqib Bashir[1,2]; Beverly Insel[1,2]; Mohan Parak[1,2], Benedicto Parker[1,2], Christian Yavorsky[5]; David Pautler[6], David Suendermann-Oeft[6], Vikram Ramanarayanan[6]

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. St. George's University, Grenada; 5. Valis Bioscience, Berkeley, CA; 6. Modality.AI Inc., San Francisco, CA

BACKGROUND

- Can negative symptoms in schizophrenia be meaningfully measured using AI-enabled vocal and facial analysis? If reliability and validity are adequate, results can lead to contact-free, non-invasive, cost-effective assessment and monitoring of negative symptoms.
- Many individuals with schizophrenia present with negative symptoms including abnormalities in vocal expression, such as altered vocal production (i.e., alogia, pressured speech) and intonation/emphasis (i.e., blunted affect; affective lability). This is reflected in communication via coupled mechanisms: vocal articulation, facial gesturing and dialogue content.
- One barrier in understanding and measuring vocal abnormalities in negative symptoms is a reliance on clinician-based rating scales - these scales can be subjective, insensitive to change in a treatment, require extensive training, be subjected to cultural disparities, and have abstruse operational definitions.
- Speech behaviors and facial movements can inform clinicians about negative symptoms and include monotone and monosyllabic speech, few gestures, pausing, speech rates, speed of movement of certain facial areas. Facial and speech changes in negative symptom patients are difficult to track and quantify with conventional techniques. A rising number of conversational agents or chatbots are equipped with artificial intelligence (AI) architecture.

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.

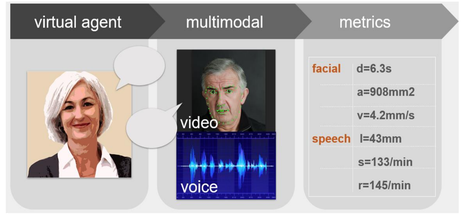
METHOD

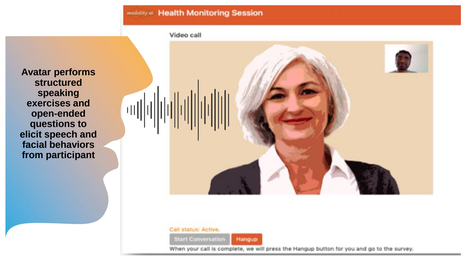
- Experimental Approach: At the first visit, the following instruments are administered: sociodemographic and clinical questionnaire, PANSS, BNSS, CDSS, CGI-S, AIMS, SAS, BARS and NEMSI. The second visit occurs within a one-week period and is done by the same clinician to assess for test-retest reliability 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.



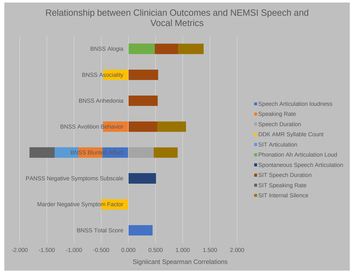


Speech and Facial Data from the program includes:

- 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

	BNSS Total Score	Marder Negative Symptom Factor	PANSS Negative Symptoms Subscale	BNSS Blunted Affect	BNSS Avolition Behavior	BNSS Anhedonia	BNSS Asociality	BNSS Alogia
Articulation Loudness	0.445			-0.478				
Speaking Rate				-0.444	-0.466			
Speech Duration				0.467				
DDK AMR Syllable Count		-0.502					-0.463	
SIT Articulation				-0.437				
Phonation "Ah" Articulation Loudness								0.486
Spontaneous Speech Articulation			0.506					
SIT Speech Duration					0.530	0.540	0.548	0.431
SIT Speaking Rate				-0.459				
SIT Internal Silence				0.440	0.530			0.468



Relationships with AI Speech and Vocal Metrics and Clinician-Rated Assessments

- 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

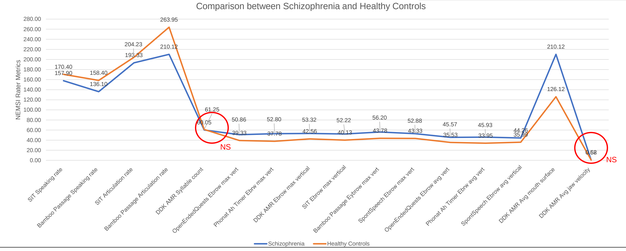
Characteristics	Schizophrenia		Healthy Controls	
	Mean	SD	Mean	SD
Age (in years)	39.95	11.59	42.11	12.18
	n	%	n	%
Gender				
Male	16	76.19	3	33.33
Female	5	23.81	6	67.67
Race				
Black	14	66.67	6	67.67
White	7	33.33	1	11.11
Asian	0	0	0	0
Other	0	0	2	22.22
Ethnicity				
Hispanic	3	14.29	2	22.22
Non-Hispanic	18	85.71	6	67.67
Not reported	0	0	1	11.11

PANSS	n	Mean	SD
Positive Subscale	21	18.62	4.08
Negative Subscale	21	23.38	3.02
General Psychopathology	21	39.67	6.26
PANSS Total	21	81.67	11.32
Marder Positive Symptom	21	25.76	4.3
Marder Negative Symptom	21	22.67	2.67
Marder Disorganized Symptom	21	19.19	3.71
Marder Hostility Symptom	21	7.71	2.37
Marder Anxiety Symptom	21	6.62	2.42

Schizophrenia compared to HC

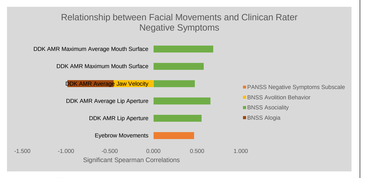
A significant difference ($p < 0.05$) was observed between patients and HCs for most NEMSI metrics; *NS = not significant*

Comparison between Schizophrenia and Healthy Controls



RESULTS: FACIAL EXPRESSION AND GESTURES

	PANSS Negative Symptoms Subscale	BNSS Avolition Behavior	BNSS Asociality	BNSS Alogia
Eyebrow Movements	0.466			
DDK AMR Lip Aperture			0.552	
DDK AMR Average Lip Aperture			0.654	
DDK AMR Average Jaw Velocity		-0.456	0.474	-0.529
DDK AMR Maximum Mouth Surface			0.576	
DDK AMR Maximum Average Mouth Surface			0.686	



Relationships with AI Facial Metrics and Clinician-Rated Assessments

- Asociality or lack of motivation to engage in social interaction or activities was positively correlated to multiple facial metrics: lip aperture, jaw velocity and mouth surface
- Negative Symptom Subscale (PANSS) was positively correlated to Eyebrow movements

RESULTS: RELIABILITY AND VALIDITY

- Reliability NEMSI AI (Time 1 and Time 2): ICC = 0.982
- Reliability PANSS Marder Negative Symptoms (Time 1 and Time 2): ICC = 0.953
- Reliability BNSS Total Score (Time 1 and Time 2): ICC = 0.956
- Validity of NEMSI with 1. BNSS Total Score = 0.801, 2. PANSS Marder Negative Symptom = 0.800, BNSS Alogia = 0.812, BNSS Avolition = 0.844
- Internal Consistency of NEMSI: 0.867 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.