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

Jean-Pierre Lindenmayer[1,3]; Anzalee Khan[1,2]; Vikram Ramanarayanan[5]; Hardik Kothare[5]; David Pautler[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

INTRODUCTION

The automatic analysis of facial and acoustic expressions has been applied to the study of facial and speech productions in individuals with schizophrenia. This new method has particular relevance for the assessment of negative symptoms. Features of negative symptoms are the reduction of facial movements, emotional facial expressions and changes in speech patterns.

Current methods of assessing negative symptoms depend on verbal report from patients and/or caregivers and a clinical rating scales, which can be insensitive to change in treatment, subjective, require extensive training and are subject to cultural disparities.

With passive digital 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 disease and conversion to psychosis.

Similarly, digital therapeutics platforms can be used to supplement clinical interviews for more objective and precise measurements of negative symptoms.

Would a novel artificial intelligence (AI) system analyzing facial and acoustic features be equivalent to traditional measurement of expressive negative symptoms in schizophrenia?

AIMS

- 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 of negative symptoms.

METHOD

Experimental Approach: Subjects undergo two **computer-based NEMSI sessions. In addition, a**t the first visit, the following instruments are administered: sociodemographic and clinical questionnaire, PANSS, BNSS, CDSS, CGI-S, AIMS, SAS, BARS. The second visit occurs within a one-week period and is done by the same clinician to assess for test-retest reliability and inter-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 (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

BASELINE DEMOGRAPHICS AND EPS

Characteristics	Schizophrenia		Healthy Controls	
	Mean	SD	Mean	SD
Age (in years)	39.98	11.70	42.15	12.20
	n	%	n	%
Gender				
Male	59	78.67	49	59.76
Female	16	21.33	33	40.24
Race				
Black	51	68.00	50	60.98
White	22	29.33	21	25.61
Asian	2	2.67	9	10.98
Other	0	0	2	2.44
Ethnicity				
Hispanic	17	40.00	37	45.12
Non-Hispanic	45	60.00	43	52.44
Not reported	0	0	3	3.66

PANSS	n	Mean	SD
Positive Subscale	75	20.23	4.44
Negative Subscale	75	26.18	3.32
General Psychopathology	75	39.23	7.00
PANSS Total	75	82.33	11.10
Marder Positive Symptom	75	26.31	3.97
Marder Negative Symptom	75	22.47	2.69
Marder Disorganized Symptom	75	18.32	3.98
Marder Hostility Symptom	75	7.99	2.38
Marder Anxiety Symptom	75	6.73	2.53

Schizophrenia compared to HC

A significant difference ($p < 0.05$) was observed between patients and HCs for all NEMSI metrics.

A selection of NEMSI metrics compared to Extrapyramidal Symptoms Scale (EPS)

NEMSI Markers	SAS	AIMS	BARS
Phonation	-0.101	-0.201	-0.101
Cepstral Peak Prominence (level of noise in vocal signal, measures dysphonia)	-0.009	-0.222	-0.201
Speech Intelligibility (SIT), Duration, and Rate (with and without pauses)	-0.203	-0.203	-0.120
Articulation Rate and Loudness	-0.301	-0.110	-0.121
DDK - also known as syllable alternating motion rate (AMR), assesses repetitive movements of oral articulators	-0.220	-0.212	-0.122
Internal Silence (pauses)	-0.210	-0.203	-0.212
Syllable Rate and Count	-0.214	-0.301	-0.201
Lip Aperture	**-0.411**	-0.333	**-0.499**
Mouth Surface Area	-0.223	-0.222	-0.201
Jaw Velocity and acceleration	-0.231	-0.32	-0.301
Lower Lip Velocity and Acceleration	-0.304	-0.301	-0.333
Eye Opening and Eyebrow vertical position	-0.303	-0.321	-0.301
Head Tilt	**-0.481**	-0.301	-0.423

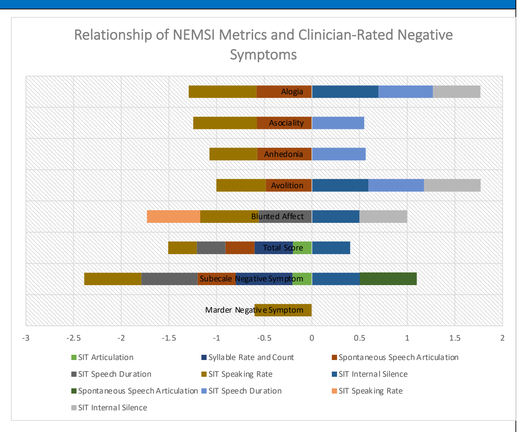
Correlations were performed using Pearson r correlations. Bolded r values are significant at < 0.05, no correlations were significant at < 0.01.

SPEECH AND VOCAL METRICS

Speech Metrics from NEMSI compared to PANSS subscale and BNSS Item Scores

	PANSS		BNSS					
	Marder Negative Symptom	Subscale Negative Symptom	Total Score	Blunted Affect	Avolition	Anhedonia	Asociality	Alogia
Speech Metrics								
Articulation Loudness	**-0.530**	-0.403	-0.105	**-0.488**	**-0.489**	**-0.550**	**-0.501**	**-0.512**
Speaking Rate	**-0.600**	**-0.532**	0.201	**-0.544**	**-0.546**	**-0.510**	**-0.511**	**-0.601**
Speech Duration	**-0.500**	**-0.430**	-0.201	**-0.588**	**-0.500**	-0.420	**-0.512**	**-0.610**
DDK AMR	**-0.582**	-0.402	-0.101	-0.421	-0.420	**-0.440**	**-0.503**	**-0.501**
SIT Articulation	-0.394	-0.304	-0.100	**-0.497**	-0.303	-0.203	**-0.602**	**-0.904**
Syllable Rate and Count	-0.203	-0.203	-0.200	-0.309	-0.304	**-0.450**	**-0.500**	**-0.517**
Spontaneous Speech Articulation	**-0.599**	**-0.596**	-0.402	**-0.559**	**-0.500**	**-0.542**	**-0.589**	**-0.591**
SIT Speech Duration	-0.403	-0.400	-0.304	**-0.450**	**-0.483**	**-0.574**	**-0.578**	**-0.581**
SIT Speaking Rate	**-0.598**	**-0.590**	-0.300	**-0.559**	**-0.500**	**-0.577**	**-0.600**	**-0.608**
SIT Internal Silence	**0.599**	**0.503**	0.401	**0.499**	**0.593**	**0.580**	**0.612**	**0.698**

Correlations were performed using Pearson r correlations. Bolded r values are significant at < 0.01.



Relationships of 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 articulation loudness of speech was negatively related to BNSS Items and the PANSS Negative Marder Factor
  - The syllable rate and count was negatively related to Alogia (poverty of speech), Anhedonia and Asociality
  - Spontaneous Speech was negatively related to the BNSS Items, PANSS Negative Symptom Subscale and PANSS Negative Marder Factor
- Speech Intelligibility (SIT) refers to how well someone can be understood when they're speaking
  - Speaking Rate was negatively related to the BNSS Items, PANSS Negative Symptom Subscale and PANSS Negative Marder Factor
  - Speech Duration was negatively correlated with BNSS Items of Blunted Affect, Avolition, Anhedonia, Asociality and Alogia
  - Internal Silence was positively correlated BNSS Items, PANSS Negative Symptom Subscale and PANSS Negative Marder Factor

FACIAL EXPRESSION AND GESTURES

	PANSS		BNSS					
	Marder Negative Symptom	Subscale Negative Symptom	Total Score	Blunted Affect	Avolition	Anhedonia	Asociality	Alogia
Facial Metrics								
Jaw Velocity	**-0.588**	**-0.561**	-0.301	**-0.549**	-0.403	-0.400	**-0.501**	**-0.599**
Articulatory: Total Mouth Surface	**-0.598**	**-0.520**	-0.409	**-0.599**	-0.400	-0.450	**-0.503**	**-0.601**
Eyebrow Movement	**-0.596**	-0.400	**-0.505**	**-0.598**	-0.411	-0.401	**-0.510**	**-0.604**
Lip Aperature	**-0.550**	**-0.599**	**-0.484**	**-0.601**	-0.349	-0.400	**-0.527**	**-0.630**

Correlations were performed using Pearson r correlations. Bolded r values are significant at < 0.01.

Relationships with AI Facial Metrics and Clinician-Rated Assessments

- Facial Features
  - Jaw velocity (mean and maximum speed of the jaw) was negatively correlated with PANSS Negative Subscale, Marder Negative Factor, and BNSS Items
  - Total Mouth Surface area and eyebrow movement was also negatively correlated with PANSS Negative Symptoms, Marder Negative Factor, and BNSS Items
  - Lip Aperature was negatively correlated with PANSS Negative Symptoms, Marder Negative Factor and BNSS Items

RELIABILITY AND VALIDITY

- Reliability NEMSI AI (Time 1 and Time 2): ICC = 0.980
- Reliability BNSS Total Score (Time 1 and Time 2): ICC = 0.957
- Validity of NEMSI with 1. BNSS Total Score = 0.805, 2. PANSS Marder Negative Symptom = 0.810, BNSS Alogia = 0.810, BNSS Avolition = 0.851
- Internal Consistency of NEMSI: 0.878 Test-Retest Reliability NEMSI: $p < 0.01$
- Overall low correlations of NEMSI variables with EPS measures

CONCLUSIONS

- Speech and facial AI technology could aid in negative symptoms assessments
- The NEMSI showed adequate reliability, validity, and internal consistency
- The NEMSI metrics showed significant separation between schizophrenia and HCs
- The NEMSI variables showed good correlations with clinical measures of negative symptoms