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

Anzalee Khan1,2; Jean-Pierre Lindenmayer1,2,3; Sebastian Prokop1,2,4; Saqib Bashir1,2; Beverly Insel1,2; Mohan Parak1,2; Benedicto Parker1,2; Christian Vavorysk5; David Pautler6; David Suendermann-Oeft6; Vikram Ramanarayanan6


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 cost-effective and sensitive assessment of negative symptoms.
- Many individuals with schizophrenia present with negative symptoms including abnormalities in vocal expression, such as altered vocal production (e.g., alogia, pressured speech) and intonation/emphasis (e.g., blunted affect; affective lability). This is reflected in communication via coupled mechanisms: vocal articulation, facial gestures, and dialogue content.
- One barrier to understanding and measuring vocal abnormalities in negative symptoms is a reliance on clinician-rated 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 Neurocognitive and Mental Health Screening Instrument (NEMSI) by comparing speech metrics (e.g., prosody, rate, intelligibility, pauses, 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, BARR 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 (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 ≥ 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, PANS, Marder Negative Factor and the CDSS.

RESULTS: SPEECH AND VOCAL METRICS

**TABLE:**

<table>
<thead>
<tr>
<th>Metric</th>
<th>Mean</th>
<th>SD</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BNSS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avolition</td>
<td>52.88</td>
<td>2.22</td>
<td>21.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Anhedonia</td>
<td>43.70</td>
<td>2.00</td>
<td>21.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Asociality</td>
<td>43.70</td>
<td>2.00</td>
<td>21.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Alogia</td>
<td>43.70</td>
<td>2.00</td>
<td>21.00</td>
<td>0.00</td>
</tr>
<tr>
<td><strong>Marder</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive Subscale</td>
<td>66.67</td>
<td>0.50</td>
<td>21.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Negative Subscale</td>
<td>33.33</td>
<td>2.00</td>
<td>21.00</td>
<td>0.00</td>
</tr>
<tr>
<td><strong>PANS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive Subscale</td>
<td>66.67</td>
<td>0.50</td>
<td>21.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Negative Subscale</td>
<td>33.33</td>
<td>2.00</td>
<td>21.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

**Table:** 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 (MAR), 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: RELIABILITY AND VALIDITY

- **Reliability NEMSI AI (Time 1 and Time 2): ICC = 0.982**
- **Reliability BNSS Marder Negative Symptoms (Time 1 and Time 2): ICC = 0.995**
- **Reliability BNSS Total Score (Time 1 and Time 2): ICC = 0.956**
- **Validity of NEMSI with HC:** A significant difference (p < 0.05) was observed between patients and HC for most NEMSI metrics; NS = not significant

RESULTS: FACIAL EXPRESSION AND GESTURES

- **Facial Symmetry**
- **Facial Movements**

CONCLUSIONS

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