

Multimodal conversational technology for remote assessment of symptom severity in people with schizophrenia

Hardik Kothare, Michael Neumann, Oliver Roesler, Jackson Liscombe, David Pautler, David Suendermann-Oeft, Christian Yavorsky, Anzalee Khan, Jean-Pierre Lindenmayer and Vikram Ramanarayanan

Abstract:

The causes of schizophrenia are complex and largely unknown. Current treatments focus on the management of positive or psychotic symptoms. However, negative symptoms like lack of facial and verbal expression, monotone speech, social withdrawal and isolation affect people with schizophrenia on a day-to-day basis and are harder to recognize and interpret. A shortage of mental health services in parts of the world makes it difficult to provide care and support to people with schizophrenia. Additionally, the COVID-19 pandemic exposed the pressing need for clinicians to be able to monitor treatment and symptoms in psychiatric disorders remotely. We present an exploratory study highlighting the feasibility of a novel multimodal conversational platform with real-time automated extraction of speech and facial metrics to assess symptom severity in people with schizophrenia.

Eight participants diagnosed with schizophrenia took part in this study. Their symptom severity was evaluated in clinic using standard clinical psychometric instruments like the Positive and Negative Syndrome Scale (PANSS) and the Brief Negative Symptom Sale (BNSS) along with cognitive assessments like the Wide Range Achievement Test -4 (WRAT-4) Reading. They were then guided by a virtual conversational agent through an interactive session on a cloud-based multimodal dialogue platform that included a battery of tasks to elicit speech and facial behaviours like syllable repetitions, read speech and spontaneous speech. An audio-visual analytics module automatically extracted a variety of measures capturing speech acoustic properties, facial landmark positions and articulatory kinematics during these tasks.

We found that articulation rate ($r^2=0.64$ $p=0.02$), loudness($r^2=0.67$ $p=0.01$), internal silence ($r^2=0.63$ $p=0.02$) and lip aperture ($r^2=0.61$ $p=0.02$) correlated highly with WRAT-4 Reading scores. Vertical position of the eyebrows ($r^2=0.89$ $p<0.01$) correlated *positively* with the PANSS positive subscale whereas measures like cepstral peak prominence ($r^2=0.57$ $p=0.03$) and mouth surface area ($r^2=0.63$ $p=0.02$) correlated *negatively* with the PANSS general psychopathology subscale. Metrics like cepstral

peak prominence ($r^2=0.66$ $p=0.01$) and lower lip velocity ($r^2=0.56$ $p=0.03$) also correlated with the BNSS total score.

The small sample size notwithstanding, these observed correlations between standard clinical assessments and speech and visual analytics suggest that audio-visual measures extracted through multimodal conversational interfaces could be an effective tool for remote monitoring of symptoms and the efficacy of treatment strategies in people with schizophrenia.