Assessment and Prediction of Negative Symptoms of Schizophrenia from RGB+D Movement Signals

Debsubhra Chakraborty*, Yasir Tahir†, Zixu Yang‡, Tomasz Maszczyk§, Justin Dauwels§, Daniel Thalmann†,
Nadia Magnenat Thalmann†, Bhing-Leet Tan‡, and Jimmy Lee‡
*Institute for Media Innovation, Interdisciplinary Graduate School, Nanyang Technological University, Singapore
†Institute for Media Innovation, Nanyang Technological University, Singapore
‡Institute of Mental Health, Singapore
§School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore

***Abstract*—Negative symptoms of schizophrenia significantly affect the daily functioning of patients, especially movement and expressive gestures. The diagnosis of such symptoms is often difficult and require the expertise of a trained clinician. Apart from these subjective methods, there is little research on developing objective methods to quantify the symptoms. Therefore, we explore body movement signals as objective measures of negative symptoms. Specifically, we extract the signals from video recordings of patients being interviewed. We analysed the interviews of 69 paid participants (46 patients and 23 healthy controls) in this study. Correlation between movement signals (linear and angular speeds of upper limbs and head, acceleration and gesture angles) and subjective ratings (assigned during same interview) from the NSA-16 scale were calculated. As hypothesized, the movement signals correlated strongly with the movement impairment aspect of the NSA-16 questionnaire. Also, not quite surprisingly, strong correlations were obtained between the movement signals and speech items of NSA-16, indicating lack of associated gestures in patients during speech. These subjective ratings could also be reasonably predicted from the objective signals with an accuracy of 61-78% using machine-learning algorithms with leave-one-out cross-validation technique. Furthermore, these objective measures can be reliably utilized to distinguish between the patient and healthy groups, as supervised learning methods can classify the two groups with 74-87% accuracy.**

***Index Terms*—Schizophrenia, Negative symptoms, Kinect video, Body movement, Correlation, Prediction**

I. Introduction

Schizophrenia is a chronic and disabling serious mental disorder that often develops in adolescence and runs a lifelong course. Schizophrenia has a heterogeneous presentation characterized broadly by positive (hallucinations and delusions), negative (apathy, blunting of affect, alogia) and cognitive (attention, memory and executive function) symptoms [1]. While positive symptoms tend to be overt and easily identifiable with effective pharmacological treatments, negative and cognitive symptoms are often neglected with ineffective treatments [2]. Negative and cognitive symptoms have been consistently reported to contribute to the observed disability in schizophrenia and have been highlighted as significant unmet needs by practitioners and researchers in the field [3]. Despite the failure of pharmacological treatments of cognitive and negative symptoms, recent randomised trials of non-pharmacological treatments such as Cognitive Remediation Therapy (CRT) have reported improvement on both cognitive and negative symptoms [4]–[6].

978-1-5090-3649-3/17/$31.00 ©2017 IEEE

Apart from speech, impaired psychomotor speed has been identified as one of the strongest predictors of the severity of negative symptoms [7], [8]. Although there have been numerous rating scale-based studies exploring the relation between motor function impairment and schizophrenia, there have been few attempts to quantify the negative symptoms of schizophrenia with some rater-independent, objective method. Kupper et al. [9] employed Motion Energy Analysis (MEA) to measure the head and body movement of patients and found correlations of such movements with the ratings from the PANSS scale. These movements were calculated from rather short video clips (median duration 52 s) of patients during role-play tasks. However, role-playing may heighten the emotions and body-movements of patients, and thus, may not be an accurate reflection of the negative symptoms exhibited during daily activities. MEA was also employed [10] to analyse the coordination of movements between individuals with schizophrenia and the psychologist during dyadic therapy sessions. In another study [11], the authors conducted an objective assessment of motor behaviours in people with schizophrenia using partial auto-correlation analysis of actigraphy time-series data. Researchers have been looking at other similar objective methods to evaluate the symptoms of schizophrenia. Automated analysis of facial expressions was carried out to gauge the blunting of emotional expressiveness as a feature of schizophrenia and other neuropsychotic disorders [12], [13]. Lavelle et al. [14] used 3-D motion capture technology to determine the difference in non-verbal communication during social interactions in the presence of a patient with schizophrenia. More recently, electromyography signals were explored to evaluate the consequences of reduced expressive gestures in schizophrenia, specifically smiling, in social interactions [15]. Other researchers have worked on a more non-invasive approach to examine objectively the effects of other mental diseases. In [16], the authors studied the viability of Microsoft Kinect and other motion-tracking systems to capture the social motor coordination in children suffering from autism.

Fig. 1. The video recording hardware and joint position extraction system.

Similarly, automatic non-verbal behaviour descriptors such as eye gaze, smile duration, and others were extracted from Kinect recordings “to identify indicators of psychological disorders such as depression, anxiety, and post-traumatic stress disorder” [17]. However, we did not find any study which employed objective measures to differentiate individuals with schizophrenia from healthy individuals. This is a potential limitation in the application of automated identification and quantification of schizophrenia symptoms. As a result, we wanted to answer the following research questions:

1) What is the association between objective and subjective evaluations of motor movement? Can the objective measurements be used to predict the same subjective ratings?
2) Can the objective measures differentiate individuals with schizophrenia from healthy individuals?

In this paper, we focus on objective body movement signals. These signals are extracted from video recordings of participants during a semi-structured clinical interview. To establish the validity of our movement signals, we correlate them with the subjective ratings assigned by the psychometricians on the Negative Symptoms Assessment (NSA-16) scale — a standard, well-established rating scale for negative symptoms [18]. We also report how the movement signals were applied as features in machine learning algorithms to predict the previously obtained subjective ratings, and again to distinguish the patients from healthy individuals.

This paper is organized as follows: in Section II, we describe the experimental design, and give the demographics data of the participants. In Section III, we mention the hardware for video recording and elucidate the body movement signals in detail. The results of the correlation between the objective and subjective measures are given in Section IV; it also lists the prediction accuracies of the subjective ratings (61-78%) and the classification accuracies of patient v/s healthy individuals (74-87%). Section V provides a brief discussion of the results from the previous section. Finally, in Section VI, we present our concluding remarks.

II. Design of Experiment

This study was conducted in collaboration with the Institute of Mental Health (IMH) in Singapore. There are two groups of participants (Total participants = 69): *Patients* (N = 46), individuals with schizophrenia, and *Controls* (N = 23), individuals who are not diagnosed with any disorders. All participants are matched for age, gender, ethnicity and educational qualifications, and are recruited by IMH. Table I gives the demographics of the participants. The participants receive monetary remuneration for taking part in the study. All the participants are above 21 years of age and have given their written informed consent. Ethics approval for this study was provided by the National Healthcare Group’s Domain-Specific Review Board in Singapore.

TABLE I
Demographics data of participants.

		Patients (N = 46)	**Controls** (N = 23)
Age	Mean (years)	31.2	28.4
	Range (years)	20-51	19-47
Gender	Male	23	11
	Female	23	12
Ethnicity	Chinese	38	19
	Malay	5	3
	Indian	3	1
Education	University	6	3
	Diploma/ Vocational	25	14
	High School	15	6

Following the experiment design, each participant is assessed on a cognitive battery - the Brief Assessment of Cognition (BAC), and a semi-structured clinical interview. The interview of the participant is conducted in English by a trained psychometrician from IMH, and this interview is audio and video recorded. The psychologist rates the behaviour displayed by the participant during the interview on a scale of 1-6 (1 indicating no symptoms, and 6 indicating severe symptoms) on the NSA-16 [18]. There is no role-playing involved during the interview. As mentioned before, the interview is semi-structured, and there is no fixed time-limit for the patient responses. The videos of these interviews have been analysed from start to finish, instead of selecting only a part of the interview. This provides a more comprehensive understanding of the participants’ behaviour in a conversation setting. On average, the interviews lasted for about 30 minutes, and we have analysed about 34 hours of recorded video data.

III. System Overview

In this section we briefly describe the hardware utilized to record the video data, and the body movement signals extracted from such captured recordings. Fig.1 illustrates the video recording and joint position extraction system.

A. Video recording hardware

Microsoft Kinect® v.1.0 camera is used to capture video and depth data of participant during the interview. The camera is placed about 2 metres in front of the seated participant, and its RGB+D sensor can track the body skeleton and the positions (x, y, and z co-ordinates in terms of pixels) of 20 joints in real-time. These joints are *Head*, *Shoulder* (*Right*, *Left*, and *Center*), *Elbows* (*Right* and *Left*), *Wrists* (*Right* and *Left*), *Hands* (*Right* and *Left*), *Spine*, *Hip* (*Right*, *Left*, and *Center*), *Knee* (*Right* and *Left*), *Ankles* (*Right* and *Left*), and *Feet* (*Right* and *Left*). The joint positions are extracted from the recorded video frames and stored sequentially in a .csv file. The frame rate of Kincet® is experimentally found to be 25 fps. The joint positions from one such frame, plotted on a 2-D plane, along with the proper limbs connecting the joints, form a skeleton of the participant, as demonstrated in Fig.2.

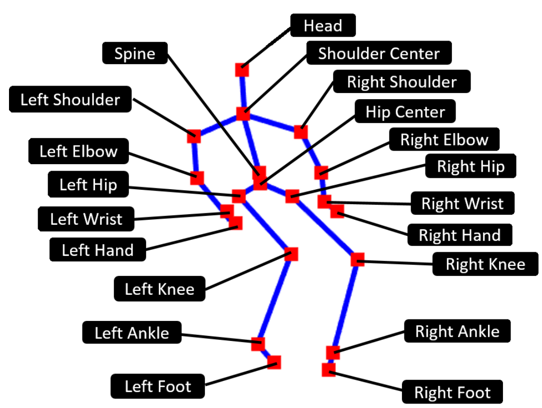

Fig. 2. The joints of the participant as captured by Kinect, and associated skeleton.

B. Body movement signals

We focus our attention only to the joints of the upper body limbs, i.e., *Elbows*, *Wrists*, *Hands*, and *Head* as we observed that most of the movement (more than 70%) of the participants are restricted to these 7 joints. Moreover, for a few files, the lower body joints are not available, due to the inability of the Kinect® to track those joints for the participant in a seated position. The joint positions are first filtered using a median filter with a window length of 2 seconds (50 samples - 25 fps $\times$ 2 seconds) to remove any spike noises.

Table II gives a summary of the extracted movement signals. The difference in position of these upper body joints between consecutive frames is the distance moved by the joints during the frame interval time. Therefore, this distance is proportional to the linear speed of the particular joint, since this distance when divided by the time interval between frames (a constant value of 1/25 seconds) yields the linear speed. Similarly, the difference between the linear speeds of successive frames is proportional to the joint's linear acceleration. To get a broader overall picture of the movement of the limbs of a participant during the interview, the linear speeds of the 6 joints are added together (named as *Gross upper limb linear speed*, and their mean and standard deviation are calculated. Similar action is performed for the linear accelerations of the 6 joints, yielding the *Gross upper limb linear acceleration*. Furthermore, to have a sense of the angular movement of the limbs, the angle at the elbows and at the wrists are calculated for every frame. The difference of the angles between consecutive frames generates the angular speeds, and one further level of difference of the angular speeds produces the angular accelerations. As before, the angular speeds and accelerations of the *Elbows* and *Wrists* are added together to respectively yield *Gross upper limb angular speed* and *Gross upper limb angular acceleration*, and their mean and standard deviations are calculated. The linear and angular speeds for the *Head* joint are also computed, and their means and standard deviations noted. Also, the top 0.1 percent i.e., the 99.9^{th} percentile and above values of the angular difference in *Elbow* and *Wrist* joints are considered as significant gestures. The mean and standard deviation of all such gesture angles are taken into account.

TABLE II
SUMMARY OF THE BODY MOVEMENT SIGNALS.

Name	Explanation	Signals	Movement Signal No.
Gross upper limb linear speed	Summation of linear speeds of *Elbow*, *Wrist* and *Hand* joints	Mean	1
		Standard deviation	2
Gross upper limb linear acceleration	Summation of linear accelerations of *Elbow*, *Wrist* and *Hand* joints	Mean	3
		Standard deviation	4
Gross upper limb angular speed	Summation of the angular speeds of *Elbow* and *Wrist* joints	Mean	5
		Standard deviation	6
Gross upper limb angular acceleration	Summation of angular accelerations of *Elbow* and *Wrist* joints	Mean	7
		Standard deviation	8
Head	Linear speed	Mean	9
		Standard deviation	10
	Angular speed	Mean	11
		Standard deviation	12
Gestures	Top 0.1 percentile values of the angular difference of *Elbow* and *Wrist* joints	Mean of the angles	13
		Standard deviation of the angles	14

IV. RESULTS

In this section we present the results of the analysis of the analysis conducted on the video recordings. First, we furnish the correlation between the objective body movement signals and the subjective ratings on the NSA-16 scale, as assigned by the psychometrician during the interview. Next, we put forward the prediction results of few of the NSA-16 ratings from the objective movement signals. Finally, we present the results of binary classification of participants into *Patient* and *Control* groups, using the movement signals as features to supervised pattern recognition classifiers.

A. Correlation

Table III shows the linear correlation between the objective body movement signals extracted from the video recordings and the subjective NSA-16 ratings provided by the psychometricians. The significant correlations ($p<0.05$) are

TABLE III
CORRELATIONS OF BODY MOVEMENT SIGNALS AND NSA-16 (N = 46).

NSA-16 Items	Movement signal no.													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Prolonged Time to Respond	-0.31*	-0.28	-0.30*	-0.21	-0.27	-0.27	-0.26	-0.27	-0.21	-0.21	-0.31*	-0.17	-0.28	-0.33*
Restricted Speech Quantity	-0.44†	-0.33*	-0.40†	-0.23	-0.42†	-0.47‡	-0.40†	-0.44†	-0.34*	-0.16	-0.28	-0.06	-0.50‡	-0.51‡
Impoverished Speech Content	-0.15	-0.15	-0.12	-0.09	-0.14	-0.12	-0.12	-0.11	-0.31*	-0.31*	-0.33*	-0.17	-0.14	-0.19
Inarticulate Speech	0.13	-0.01	0.08	-0.06	0.15	0.05	0.11	0.02	0.14	0.18	0.28	0.24	0.05	0.00
Emotion Reduced Range	-0.27	-0.15	-0.23	-0.12	-0.20	-0.26	-0.18	-0.25	-0.27	-0.31*	-0.22	-0.03	-0.29*	-0.27
Affect Reduced Modulation of Intensity	-0.25	-0.30*	-0.26	-0.25	-0.25	-0.28	-0.25	-0.28	-0.31*	-0.10	-0.24	0.01	-0.30*	-0.30*
Affect Reduced Display on Demand	-0.04	-0.01	-0.04	0.00	-0.13	-0.18	-0.14	-0.17	-0.11	0.04	-0.06	0.02	-0.21	-0.19
Reduced Social Drive	0.02	-0.04	0.04	-0.02	0.11	0.03	0.13	0.06	-0.16	-0.21	-0.20	-0.06	-0.02	-0.06
Poor Rapport with Interviewer	-0.11	-0.13	-0.13	-0.10	-0.03	-0.11	-0.05	-0.12	-0.08	-0.04	-0.16	-0.18	-0.15	-0.20
Interest in Emotional and Physical Intimacy	-0.04	-0.24	-0.11	-0.26	-0.01	-0.08	-0.04	-0.10	0.00	0.01	-0.05	-0.08	-0.06	-0.09
Poor Grooming and Hygiene	-0.10	-0.09	-0.14	-0.08	-0.03	0.02	-0.01	0.03	0.27	0.32*	0.19	0.15	0.04	0.15
Reduced Sense of Purpose	0.13	0.17	0.11	0.16	0.03	-0.05	0.02	-0.05	-0.11	-0.02	-0.03	0.11	-0.06	-0.17
Reduced Interests	-0.08	-0.16	-0.10	-0.19	0.00	-0.05	0.01	-0.04	-0.30*	-0.23	-0.17	-0.07	-0.06	-0.12
Reduced Daily Activity	-0.18	-0.06	-0.19	-0.02	-0.05	-0.09	-0.04	-0.10	-0.29*	-0.06	-0.18	0.02	-0.10	-0.21
Reduced Expressive Gestures	-0.33*	-0.25	-0.30*	-0.17	-0.29	-0.32*	-0.28	-0.31*	-0.43†	-0.29	-0.30*	0.04	-0.34*	-0.40†
Slowed Movements	-0.08	-0.03	-0.07	-0.03	-0.07	-0.10	-0.06	-0.08	-0.12	0.04	0.17	0.36*	-0.09	-0.23
Global Negative Symptoms Rating	-0.28	-0.30*	-0.29	-0.25	-0.17	-0.23	-0.17	-0.22	-0.37*	-0.28	-0.29	-0.03	-0.25	-0.31*

*$p<0.05$, †$p<0.01$, ‡$p<0.001$

TABLE IV
PREDICTION RESULTS FOR NSA-16 ITEMS.

NSA-16 Item	Confusion matrix		Predicted class 0	Predicted class 1	Precision	Recall	F-score	ROC Area	Accuracy	Algorithm	Feature selection
Restricted Speech Quantity	**True class**	0	25	2	0.758	0.926	0.833	0.752	78.26 %	Linear SVM with Stochastic Gradient Descent	Subset with minimum redundancy
		1	8	11	0.846	0.579	0.688	0.752			
Reduced Expressive Gestures	**True class**	0	24	4	0.750	0.857	0.800	0.706	73.91 %	Linear SVM with Stochastic Gradient Descent	No
		1	8	10	0.714	0.556	0.625	0.706			
Impoverished Speech Content	**True class**	0	14	10	0.737	0.583	0.651	0.678	67.39 %	Linear SVM with Stochastic Gradient Descent	Subset with the significantly correlated features
		1	5	17	0.630	0.773	0.694	0.678			
Affect Reduced Modulation of Intensity	**True class**	0	16	7	0.615	0.696	0.653	0.630	63.04 %	k-Nearest Neighbours	No
		1	10	13	0.650	0.565	0.605	0.630			
Prolonged Time to Respond	**True class**	0	21	9	0.700	0.700	0.700	0.569	60.87 %	Linear SVM with Stochastic Gradient Descent	Subset with the significantly correlated features
		1	9	7	0.438	0.438	0.438	0.569			

highlighted. Only the *Patients* are considered in the correlation calculation, since the NSA-16 ratings for *Controls* have no clinical significance as the questionnaire was not designed for evaluating healthy individuals. As can be seen from Table III, there were significant correlations between the subjective ratings and the objective body movement signals. Overall, the correlations were mostly negative in value, indicating that *Patients* with higher ratings on the NSA-16 scale, i.e., having greater severity of negative symptoms, generally exhibit less movement, indicated by lower mean and standard deviation values of linear and angular speeds, acceleration, and angular gestures. *Reduced Expressive Gestures* correlated inversely with the body movement signals, thus objectively confirming and corroborating the subjective ratings. Another interesting finding was the negative correlation of the body movement signals with the speech items of NSA-16, especially *Restricted Speech Quantity* and *Prolonged Time to Respond*.

B. NSA-16 items prediction

Since several NSA-16 items were correlated with the body movement signals, we evaluated the reliability of movement signals to predict subjective ratings. As ratings of NSA-16 were skewed, the items were categorized into two classes: Unobservable (Class 0: ratings of 1 or 2 on the items, implying no observable symptom), and Observable (Class 1: ratings of 3 and above, implying observable symptom). We trained machine learning classifiers with the movement signals as features and the class labels as targets, and performed leave-one-out cross-validation. Below, in Table IV, we present the results for a few of the correlated NSA-16 items (which have accuracy over 60% and ROC area over 0.500), giving the confusion matrix and associated statistics for the best-performing classifier.

TABLE V
PATIENTS V/S CONTROLS CLASSIFICATION WITH BODY MOVEMENT SIGNALS.

Confusion matrix		Predicted class 0	Predicted class 1	Precision	Recall	F-score	ROC Area	Accuracy	Algorithm	Feature selection
True class	0	11	12	0.647	0.478	0.550	0.674	73.91%	Linear SVM with Stochastic Gradient Descent	Body movement signals except *Head* signals
	1	6	40	0.769	0.870	0.816	0.674			

C. Classification of participants

Finally, we utilized the objective body movement signals as features to binary classifiers in order to distinguish between the Controls (Class 0, N = 23) and Patients (Class 1, N = 46) groups. The participant groups were given as target labels, and leave-one-out cross-validation was performed to calculate the accuracy of prediction. Table V lists the classification accuracy, confusion matrix and associated metrics for the best classifier.

TABLE VI
PATIENTS V/S CONTROLS CLASSIFICATION WITH SPEECH AND BODY MOVEMENT SIGNALS.

Confusion matrix				Precision	Recall	F-score	ROC Area	Accuracy	Algorithm	Feature selection
		Predicted class								
		0	1							
True class	0	21	2	0.750	0.913	0.824	0.913	86.76%	Multilayer Perceptron	Least redundant features from the subset of top 75% of the most correlated features
	1	7	38	0.950	0.844	0.894	0.913			

Previously, we had described the relation of objective, non-verbal speech cues with the subjective NSA-16 ratings in [19], where we could distinguish between the *Patient* and *Control* groups based on their speech cues. In this study, we extracted 41 conversational and prosodic speech cues similar to those extracted in our previous works [19] and [20]. We appended these speech signals (41 in number) to the movement signals, and utilized this extended set to perform classification of participants into *Controls* (Class 0, N = 23) and *Patients* (Class 1, N = 45) groups with leave-one-out cross-validation. The results of the classification and associated metrics for the best-performing classifier are given in Table **??**. A number of algorithms and feature combinations were tested, and the results for the best classifier and feature-set is listed. The best performance was obtained with a subset of the feature-set; first, a smaller set was created with top 41 (i.e., 75%) of the features which are most correlated with the class label, and from that smaller set, 5 of the features were selected which had the least redundancy among them, i.e., they are correlated with the class label, but not among each other. As can be seen from the Table **??**, the *Control* and *Patient* groups can be distinguished with high accuracy. It can be noted that our dataset is unbalanced, with the number of *Patients* being almost twice as many as *Controls*, which biases the classifiers towards the majority class. To deal with this problem, the Asymmetric Partial Least Squares Regression algorithm has been proposed in literature [21], [22]. We applied this algorithm to the speech and body movement signals, and although it slightly improves the *Recall* value of the minority class, the *Precision* and overall *F-score* values of both the majority and minority classes degrade considerably.

V. DISCUSSION

In the previous section, results from classification and correlation analyses were presented. In this section, we discuss a few observations regarding them. As mentioned before, the correlation coefficient values were mostly negative, indicating an inverse relationship between the subjective and objective measures as hypothesized. This points to the fact that patients suffering from higher severity of negative symptoms of schizophrenia indeed have more muted body movements and gestures, which is a well-known symptom of schizophrenia. Another observation is the strong correlation of the speech items of NSA-16 scale with the body movement signals. Another key symptom of schizophrenia is speech impairment, and natural speech is often accompanied with associated gestures (non-verbal cues). Hence, speech deficiency in patients can be coupled with blunted display of gestures or body movements as captured in their video recordings. An additional observation is the apparent lack of strong correlation of the body movement signals with the NSA-16 item *Slowed Movements*, for which it seems obvious that there should be strong correlations among them. This could be explained by the fact that the participants were rated on this item during a different task, separate from the interview, and thus were not video recorded.

VI. CONCLUSION

Schizophrenia is a debilitating mental disorder with significant burden and affecting of millions of people globally. The negative symptoms of schizophrenia are often manifested through amotivation, alogia, and avolition, and are difficult to identify without the expertise of a trained clinician. In this paper, we established the moderate to strong correlations between objective body movement signals and negative symptoms in schizophrenia. We also demonstrated that these movement signals alone, or in combination with speech signals, could also be utilized to distinguish between individuals with schizophrenia from healthy individuals. Therefore, these objective signals have the potential to supplement the subjective system of negative symptoms assessment. It can be a helpful aid in clinical practice to screen for presence and severity of negative symptoms.

ACKNOWLEDGEMENT

This study was funded by the Singapore Ministry of Health's National Medical Research Council Center Grant awarded to the Institute of Mental Health Singapore (NMRC/CG/004/2013) and by NITHM grant M4081187.E30. This research is also supported in part by the Being Together Centre, a collaboration between Nanyang Technological University (NTU) Singapore and University of North Carolina (UNC) at Chapel Hill. The Being Together Centre is supported by the National Research Foundation, Prime Ministers Office, Singapore under its International Research Centres in Singapore Funding Initiative. Moreover, this project is funded in part by the RRIS Rehabilitation Research Grant RRG2/16009. The authors would also like to acknowledge the Interdisciplinary Graduate School and Nanyang Technological University for their support of this research.

REFERENCES

[1] Caroline Demily and Nicolas Franck, "Cognitive remediation: a promising tool for the treatment of schizophrenia," *Expert Review of Neurotherapeutics*, vol. 8, no. 7, pp. 1029–1036, 2008.

[2] Brendan P Murphy, Young-Chul Chung, Tae-Won Park, and Patrick D McGorry, "Pharmacological treatment of primary negative symptoms in schizophrenia: a systematic review," *Schizophrenia research*, vol. 88, no. 1, pp. 5–25, 2006.

[3] Brian Kirkpatrick, Wayne S Fenton, William T Carpenter, and Stephen R Marder, "The nimh-matrics consensus statement on negative symptoms," *Schizophrenia bulletin*, vol. 32, no. 2, pp. 214–219, 2006.

[4] Manouchehr Gharaeipour and BJ Scott, "Effects of cognitive remediation on neurocognitive functions and psychiatric symptoms in schizophrenia inpatients," *Schizophrenia research*, vol. 142, no. 1, pp. 165–170, 2012.

[5] Pedro Sánchez, Javier Peña, Eneritz Bengoetxea, Natalia Ojeda, Edorta Elizagárate, Jesus Ezcurra, and Miguel Gutiérrez, "Improvements in negative symptoms and functional outcome after a new generation cognitive remediation program: a randomized controlled trial," *Schizophrenia bulletin*, vol. 40, no. 3, pp. 707–715, 2014.

[6] Bhing-Leet Tan and Robert King, "The effects of cognitive remediation on functional outcomes among people with schizophrenia: A randomised controlled study," *Australian & New Zealand Journal of Psychiatry*, vol. 47, no. 11, pp. 1068–1080, 2013.

[7] Dwight Dickinson and Robert D Coursey, "Independence and overlap among neurocognitive correlates of community functioning in schizophrenia," *Schizophrenia research*, vol. 56, no. 1, pp. 161–170, 2002.

[8] Micah A Hammer, Joanna Katsanis, and William G Iacono, "The relationship between negative symptoms and neuropsychological performance," *Biological Psychiatry*, vol. 37, no. 11, pp. 828–830, 1995.

[9] Zeno Kupper, Fabian Ramseyer, Holger Hoffmann, Samuel Kalbermatten, and Wolfgang Tschacher, "Video-based quantification of body movement during social interaction indicates the severity of negative symptoms in patients with schizophrenia," *Schizophrenia research*, vol. 121, no. 1, pp. 90–100, 2010.

[10] Fabian Ramseyer and Wolfgang Tschacher, "Nonverbal synchrony in psychotherapy: coordinated body movement reflects relationship quality and outcome.," *Journal of consulting and clinical psychology*, vol. 79, no. 3, pp. 284, 2011.

[11] Sebastian Walther, Fabian Ramseyer, Helge Horn, Werner Strik, and Wolfgang Tschacher, "Less structured movement patterns predict severity of positive syndrome, excitement, and disorganization," *Schizophrenia bulletin*, vol. 40, no. 3, pp. 585–591, 2014.

[12] Alex S Cohen, Sean C Morrison, and Dallas A Callaway, "Computerized facial analysis for understanding constricted/blunted affect: Initial feasibility, reliability, and validity data," *Schizophrenia research*, vol. 148, no. 1, pp. 111–116, 2013.

[13] Jihun Hamm, Christian G Kohler, Ruben C Gur, and Ragini Verma, "Automated facial action coding system for dynamic analysis of facial expressions in neuropsychiatric disorders," *Journal of neuroscience methods*, vol. 200, no. 2, pp. 237–256, 2011.

[14] Mary Lavelle, Patrick GT Healey, and Rosemarie McCabe, "Is nonverbal communication disrupted in interactions involving patients with schizophrenia?," *Schizophrenia bulletin*, vol. 39, no. 5, pp. 1150–1158, 2013.

[15] Marcel Riehle and Tania M Lincoln, "Social consequences of subclinical negative symptoms: An emg study of facial expressions within a social interaction," *Journal of Behavior Therapy and Experimental Psychiatry*, 2017.

[16] Veronica Romero, Joseph Amaral, Paula A Fitzpatrick, Richard C Schmidt, and Mike Richardson, "Capturing social motor coordination: A comparison of the microsoft kinect, video-motion analysis and the polhemus latus motion tracking system.," in *CogSci*, 2015.

[17] Stefan Scherer, Giota Stratou, Gale Lucas, Marwa Mahmoud, Jill Boberg, Jonathan Gratch, Louis-Philippe Morency, et al., "Automatic audiovisual behavior descriptors for psychological disorder analysis," *Image and Vision Computing*, vol. 32, no. 10, pp. 648–658, 2014.

[18] Nancy C Andreasen, "Negative symptoms in schizophrenia: definition and reliability," *Archives of general psychiatry*, vol. 39, no. 7, pp. 784–788, 1982.

[19] Yasir Tahir, Debsubhra Chakraborty, Justin Dauwels, Nadia Thalmann, Daniel Thalmann, and Jimmy Lee, "Non-verbal speech analysis of interviews with schizophrenic patients," in *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on*. IEEE, 2016, pp. 5810–5814.

[20] Yasir Tahir, Debsubhra Chakraborty, Tomasz Maszczyk, Shoko Dauwels, Justin Dauwels, Nadia Thalmann, and Daniel Thalmann, "Real-time sociometrics from audio-visual features for two-person dialogs," in *Digital Signal Processing (DSP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 823–827.

[21] Hai-Ni Qu, Guo-Zheng Li, and Wei-Sheng Xu, "An asymmetric classifier based on partial least squares," *Pattern Recognition*, vol. 43, no. 10, pp. 3448–3457, 2010.

[22] Dong-Yan Huang, Shuzhi Sam Ge, and Zhengchen Zhang, "Speaker state classification based on fusion of asymmetric simpls and support vector machines," in *Twelfth Annual Conference of the International Speech Communication Association*, 2011.