Title	Assessment of patients with negative symptoms of schizophrenia from movement, and prosodic and conversational speech signals
Author(s)	Chakraborty, Debsubhra; Yang, Zixu; Yasir, Tahir; Dauwels, Justin; Magnenat-Thalmann, Nadia; Tan, Bhing -Leet; Lee, Jimmy
Citation	Chakraborty, D., Yang, Z., Yasir, T., Dauwels, J., Magnenat-Thalmann, N., Tan, B.-L., & Lee, J. (2018). Assessment of patients with negative symptoms of schizophrenia from movement, and prosodic and conversational speech signals. Conference of the IEEE Engineering in Medicine and Biology Society.
Date	2018
URL	http://hdl.handle.net/10220/49700
Rights	

Assessment of Patients with Negative Symptoms of Schizophrenia from Movement, and Prosodic and Conversational Speech Signals

Debsubhra Chakraborty, Zixu Yang, Yasir Tahir, Justin Dauwels, Nadia Magnenat-Thalmann, Bhing-Leet Tan and Jimmy Lee

***Abstract*— Negative symptoms of schizophrenia are often characterized by speech and motor impairments. Therefore, in this paper we combine audio (prosody and conversation) and video (body movement) signals to distinguish patients (43) from healthy control (23) subjects. First, we used these different modalities individually as features in machine learning algorithms yielding an accuracy of 67-79%. Next, we combined the different modalities as features, improving the accuracy to 85%.**

I. Introduction

Negative symptoms of schizophrenia (characterized by apathy, blunting of affect, alogia) are difficult to detect with no effective drug treatments, and hence are considered significant unmet needs by clinicians. Speech and psychomotor impairments have been identified as two of the key indicators of negative symptoms. All previous studies exploring the relationship of these two factors with schizophrenia are non-automated and require the expertise of a clinician. We wish to move towards more objective and automated methods of differentiating between individuals with schizophrenia and healthy individuals.

We recorded (audio and video) 66 individuals (43 *Patients*, 23 *Healthy Controls*) while they were interviewed by psychometricians of the Institute of Mental Health (IMH), our collaborators in this study. Ethics approval for this study was provided by the National Healthcare Group's Domain-Specific Review Board in Singapore. All individuals were consenting adults and were matched for their age, gender, ethnicity and educational qualifications. Each interview lasted for about 30 minutes, and we have analysed about 33 hours of audio and video data.

*Research supported by NMRC, NITHM, RRIS and BTC 2.0.

D. Chakraborty is a PhD student with the Interdisciplinary Graduate School, Nanyang Technological University, Singapore (Phone: +65 8346 4863; e-mail: chak0008@e.ntu.edu.sg).

Z. Yang is a Senior Research Psychologist with the Institute of Mental Health, Singapore (e-mail: zixu_yang@imh.com.sg).

Y. Tahir is a Senior Research Engineer with the Institute for Media Innovation, Nanyang Technological University, Singapore (e-mail: yasirtahir@ntu.edu.sg).

J. Dauwels is an Associate Professor at the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore (e-mail: jdauwels@ntu.edu.sg).

N. Thalmann is the Director of the Institute for Media Innovation, Nanyang Technological University, Singapore (e-mail: nadiathalmann@ntu.edu.sg).

Dr. B-L. Tan is an Associate Professor at the Singapore Institute of Technology (e-mail: bhingleet.tan@singaporetech.edu.sg).

Dr. J. Lee is a Psychiatrist and Senior Consultant at the Institute of Mental Health, Singapore (e-mail: jimmy_lee@imh.com.sg).

II. System Overview

We recorded the audio using portable H4N recorder (2-channel) and the video using Microsoft Kinect camera. Our feature-sets include 988 low-level prosodic features extracted using the open-Source Media Interpretation by Large feature-space Extraction (openSMILE) toolkit, [1], 14 features associated with the dynamics of conversation between the participant and psychometrician [2], and 14 features related to body movement of the participant [3].

III. Results

Table I gives the results of binary classification of participants into *Patient* and *Control* (target labels) groups using body movement, prosodic speech and conversational speech features as attributes to machine learning algorithms. The baseline accuracy is for a dummy classifier which always predicts the most frequent class as its output.

TABLE I. Classification of Patients v/s Controls.

Feature-set	***Accuracy***	***Mean F-score***	***Algorithm***	***Baseline Accuracy***
Movement	66.67 %	0.60	Stochastic Gradient Descent (SGD) from Weka	65.15%
Prosodic	78.79%	0.76		
Conversation	66.67%	0.65		
Movement + Prosodic + Conversation	84.85%	0.83		

IV. Discussion and Conclusion

The feature-sets can individually differentiate between Patient and Control groups but fusing all the feature-sets together yields the best accuracy. This alludes to a multi-modal approach of considering all manifested symptoms together while designing an automated system for the reliable detection of schizophrenia.

References

[1] Chakraborty, Debsubhra et al. "Prediction of negative symptoms of schizophrenia from emotion related low-level speech signals." *Acoustics, Speech and Signal Processing (ICASSP), 2018 IEEE International Conference on*. IEEE, In press.

[2] Tahir, Yasir, et al. "Non-verbal speech analysis of interviews with schizophrenic patients." *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on*. IEEE, 2016.

[3] Chakraborty, Debsubhra, et al. "Assessment and prediction of negative symptoms of schizophrenia from RGB+ D movement signals." *Multimedia Signal Processing (MMSP), 2017 IEEE 19th International Workshop on*. IEEE, 2017.