NON-VERBAL SPEECH ANALYSIS OF INTERVIEWS WITH SCHIZOPHRENIC PATIENTS

Yasir Tahir[†], *Debsubhra Chakraborty*[†], *Justin Dauwels*[*], *Nadia Thalmann*[†], *Daniel Thalmann*[†], *and Jimmy Lee*[‡]

[†]Institute for Media Innovation, Nanyang Technological University, Singapore
[*]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore
[‡]Institute of Mental Health, Singapore

ABSTRACT

Negative symptoms in schizophrenia are associated with significant burden and functional impairment, especially speech production. In clinical practice today, there are no robust treatments for negative symptoms and one obstacle surrounding its research is the lack of an objective measure. To this end, we explore non-verbal speech cues as objective measures. Specifically, we extract these cues while schizophrenic patients are interviewed by psychologists. We have analyzed interviews of 15 patients who were enrolled in an observational study on the effectiveness of Cognitive Remediation Therapy (CRT). The subject (undergoing CRT) and control group (not undergoing CRT) contains 8 and 7 individuals respectively. The patients were recorded during three sessions while being evaluated for negative symptoms over a 12-week follow-up period. In order to validate the non-verbal speech cues, we computed their correlation with the Negative Symptom Assessment (NSA-16). Our results suggest a strong correlation between certain measures of the two rating sets. Supervised prediction of the subjective ratings from the non-verbal speech features with leave-one-person-out cross-validation has reasonable accuracy of 53-80%. Furthermore, the non-verbal cues can be used to reliably distinguish between the subjects and controls, as supervised learning methods can classify the two groups with 80-93% accuracy.

Index Terms— Schizophrenia, Negative Symptoms Assessment, non-verbal cues, correlation, supervised learning

1. INTRODUCTION

Schizophrenia is a chronic and disabling mental disorder that often develops in adolescence and has a heterogeneous presentation characterized broadly by positive (hallucinations and delusions), negative (apathy, blunting of affect and alogia) and cognitive (attention, memory and executive functioning) symptoms [1]. Although the pathogenesis of schizophrenia remains unclear but research so far has pointed to a strong genetic basis with estimates of heritability of risk at about 80% [2].

Current pharmacological treatments are effective in treating positive symptoms, but have at most doubtful efficacy on negative and cognitive symptoms, sometimes with harmful side-effects [3], [4]. Negative and cognitive symptoms in schizophrenia do contribute significantly to the disability seen in clinical practice. Cognitive function has also been consistently shown to be highly correlated to functioning in daily living, and this has spurred a flurry of activity to develop treatments for cognitive impairments in schizophrenia [5]. One of the most promising strategies in the literature today for the treatment of cognitive deficits in schizophrenia is Cognitive Remediation Therapy (CRT) [6]. CRT is designed to enhance cognition, and hence is likely to improve functioning outcomes and "life skills such as work and social functioning" [7]. Recent meta-analysis has shown that CRT not only led to a moderate effect size improvement in cognition, it additionally appeared to improve negative symptoms by a similar margin [6].

However, it was also found that CRT did not seem to suit all patients with schizophrenia. One of the possible explanations for this observation could be that cognitive deficits in schizophrenia have a wide range of impairment and could be as heterogeneous as the diagnostic phenotype. This makes it difficult to identify suitable patients for CRT, or to even tailor the CRT to the specific deficits. Identifying cognitive and sociological biomarkers will greatly aid the stratification and tailoring of CRT for patients with schizophrenia and can have potential to be used as an objective way to gauge an individual's response to CRT.

One of such biomarkers can be language usage among patients as speech impairments is one of the key negative symptoms in schizophrenic patients [8]. However, previous non-automated attempts to utilize the different aspects of speech and language as differentiators between schizophrenic and healthy individuals have had limited success. Although there existed some distinction in verbal fluency tasks between patients and healthy controls [9], other studies involving semantic boundary [10] or metaphor interpretation [11] indicated no significant differences between the two groups. However, automated efforts based on speech deficiencies to distinguish patients and healthy controls have had greater success with the recent advancements in computer science

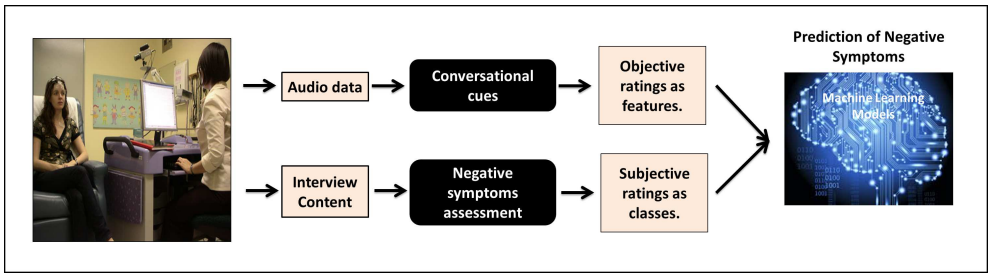

Fig. 1. An overview of data acquisition and the analysis of subjective and objective features.

and signal processing techniques. Subtle differences in communication discourses were detected among patients, their first-degree relatives and healthy controls employing Latent Semantic Analysis (LSA) in [12]. LSA was again utilized to identify lack of semantic and phonological fluency, disconnected speech, and thought disorder in [13], and LSA and machine learning were used to analyse free-speech and predict onset of psychosis respectively of high-risk youths in [14]. However, all the above methods are based on semantic analysis and natural language processing, with little attention to the restricted non-verbal cues display of patients suffering from negative symptom schizophrenia [15]. Non-verbal speech cues such as voice tone, volume, and interjections play crucial role in human interaction and communication [16], and the muted display of such signals in patients can be made use for both distinguishing them with healthy controls and developing specific and objective treatments.

In this paper, we present the preliminary results on the relationship of such non-verbal speech cues extracted from audio recordings of patient interviews with the Negative Symptoms Assessment (NSA-16) tool [17]. We also report how the non-verbal cues were applied as features in machine learning algorithms to differentiate between subjects and controls, the former group suffering from a greater degree of cognitive impairments of schizophrenia than the latter.

This paper is organized as follows: in Section 2, we describe the hardware system used to collect data and the extraction of specific non-verbal cues from the collected data. In Section 3, we elucidate the experimental design with the subjects and controls and their interview cycles with the psychologists. In Section 4, we discuss the correlation between the non-verbal speech cues and subjective ratings of the NSA-16 from the patient interviews. We also assess the accuracy of machine learning algorithms with leave-one-person-out cross-validation technique in identifying the subjective ratings from objective cues (accuracy of 53-80%) and attributing the features to subject cases and controls (accuracy of 80-93%). In Section 5, we provide concluding remarks.

2. SYSTEM OVERVIEW

In this section we briefly describe the data acquisition system and feature extraction. First, we explain the hardware setup for audio recording of conversations. Next, we briefly describe the extraction of non-verbal speech cues from the recorded audio. The overall system is illustrated in Fig. 1. The hardware system utilised for audio recording is identical to the one used in our earlier work [18], where similar non-verbal audio features were extracted.

2.1. Sensing and Recording

We employed easy-to-use portable equipment for recording conversations; it consisted of lapel microphones for each of the two speakers and an audio H4N recorder that allowed multiple microphones to be interfaced with a laptop. The audio data was recorded in brief consecutive segments as a 2-channel audio .wav file.

2.2. Extraction of Non-Verbal Cues

The conversational cues account for *who* is speaking, *when* and *how much*, while the prosodic cues quantify *how* people talk during their conversations. We computed the following conversational cues: *the number of Natural Turns*, *Speaking Percentage*, *Mutual Silence Percentage*, *Turn Duration*, *Natural Interjections*, *Speaking Interjections*, *Interruptions*, *Failed Interruptions*, *Speaking Rate* and *Response Time* [18]. Fig. 2 shows an illustration of conversational cues [18].

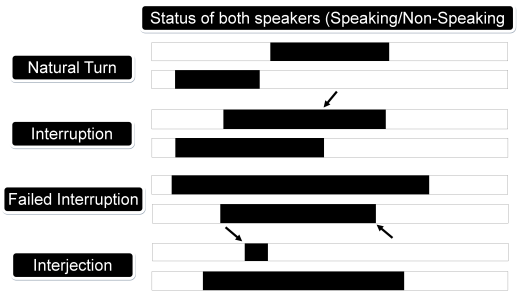

Fig. 2. Illustration of turn-taking, interruption, failed interruption, and interjection. Those conversational cues are derived from the binary speaking status (speaking vs. non-speaking) [18].

Table 1. Demographics of participants in the study.

	Subjects (N=8)		Controls (N=7)	
	Mean	**Range**	**Mean**	**Range**
Age	29.25	23-40	30.43	25-39
	N	**%**	**N**	**%**
Female	4	50.0	5	71.43
Ethnicity (Chinese)	5	62.5	7	100.0
Highest Level of Education (University)	2	25.0	1	14.29

3. EXPERIMENTAL DESIGN

We are conducting this study in collaboration with the Institute of Mental Health (IMH) in Singapore. It is an ongoing study, and the results in this paper are for the cases who have completed the 12-week study. There are two groups of participants: *Subjects* who are patients with schizophrenia undergoing CRT [6] in sessions at IMH, and *Controls* who are patients with schizophrenia at IMH, matched for age, gender, ethnicity, and education, but not undergoing CRT treatment. The control patients suffer from less severe cognitive impairments of schizophrenia compared to the subjects. The subject and the control groups were assessed for cognitive impairments related to schizophrenia at the start of the study period using the Brief Assessment of Cognition in Schizophrenia (BACS) tool [19]. The subject group had a mean BACS Composite score of 27.52, whereas the control group had a mean score of 42.49 on the same metric, a higher score indicating lesser cognitive impairment. The participants are recruited by IMH based on the recommendations of clinicians. The participants are provided with monetary compensation for participation in the study. The participants are all adults above 21 years of age, and have provided written informed consent. All experiments are performed in accordance to the relevant guidelines and regulations, and the study protocol has been approved by the National Healthcare Group's domain-specific Review Board in Singapore. So far we have collected data for 15 completed participants, including 8 subjects and 7 controls. Table 1 gives the demographics data of the participants.

The experiment has been designed such that each participant is assessed at three timepoints: the first at week 0 (before the start of the CRT sessions), the second at week 2 and the third at week 12, at the completion of CRT. Each session consists of cognitive tasks, clinical interview, and functioning tasks. In this paper we will discuss the analysis of audio data acquired during the structured clinical interview. A trained psychologist from IMH conducts these interviews in English and rates each participant on a scale of 1-6, where 1 indicating no symptoms and 6 indicating severe negative symptoms, on the Negative Symptom Assessment (NSA-16) tool. There is no pre-determined duration for the interview, instead it depends on participant's response to the questions asked by the psychologist. On average, the interviews lasted about 30 minutes. We have analysed each interview in its entirety. Therefore, in this study we have analysed about 22.5 hours of audio recordings (0.5 hour/interview $\times$ 3 sessions $\times$ 15 patients).

4. ANALYSIS AND RESULTS

In this section we will present our analysis of the data collected so far. First we show the correlation between the objective audio features and the subjective negative symptoms ratings, then we present our results for automated prediction of negative symptoms from audio features. At the end we present the the results for the classification of the control and subject groups. The classification performance was computed by leave-one-person-out cross-validation, i.e., for each participant the classifier was tested on the instances of that participant and trained on all the remaining instances.

We extracted conversational and prosodic features from the patient interviews conducted by the psychologists at IMH. As there was no role playing involved in the interviews, most of the prosodic features remained similar. Therefore, we only focus on conversational features in this section. Table 2 shows the linear correlation of conversational speech features with the patients' negative symptoms assessed by the IMH psychologists. We calculate the linear correlation value $\rho_{X,Y}$ between a non-verbal cue x_i of the ith recording of one person, say *Turn Duration*, with the rating y_i of the same recording, say *Restricted Speech Quantity*, as follows:

$$\rho_{X,Y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}, \quad (1)$$

where $\bar{x}$ and $\bar{y}$ denote the corresponding mean value for all recordings. Table 2 shows the correlation of 9 NSA-16 criteria with conversational audio features. Each cell of the table contains two values separated by a comma; the first value being the linear correlation value itself, whereas the second value is the corresponding *p*-value. When the *p*-value is less than 0.05, the corresponding correlation value is treated as statistically significant, and hence the cell is colored for easier understanding, green for positive correlation and red for negative correlation. It can be seen from Table 2 that features like *Failed Interrupt*, *Overlap*, and *Response Time* are directly correlated to the negative symptoms; on the other hand *Interjection*, *Interrupt*, *Speaking Percentage*, *Speech Rate*, and *Turn Duration* are anti-correlated to the negative symptoms. Only those objective non-verbal conversational features have been listed in the table which have at least one significant correlation with any one of the subjective features.

Although the set of subjective and objective measures were selected independently, the significant correlations between some of them indeed make sense. For example, the subjective feature *Prolonged time to respond* is negatively related to *Interjection*, and positively to *Failed Interrupt* and *Response Time*. A patient displays *Poor rapport with interviewer* when she takes longer time to respond (increased *Response Time*), speaks simultaneously with the interviewer

Table 2. Correlation values between subjective and objective measures

	Interjection	Failed Interrupt	Overlap	Turn Duration	Speaking %	Speech Rate	Response Time
Prolonged time to respond	-0.62, 0.014	0.53, 0.042	0.41, 0.130	-0.23, 0.400	-0.33, 0.236	-0.49, 0.066	0.55, 0.032
Restricted speech quantity	-0.56, 0.030	0.22, 0.429	0.16, 0.565	-0.51, 0.051	-0.53, 0.040	0.13, 0.647	0.59, 0.022
Impoverished speech content	-0.58, 0.023	0.38, 0.158	0.29, 0.294	-0.02, 0.935	-0.07, 0.801	-0.31, 0.254	0.50, 0.058
Inarticulate speech	-0.52, 0.045	0.59, 0.020	0.46, 0.083	-0.11, 0.686	-0.16, 0.570	-0.55, 0.033	0.53, 0.043
Emotion: Reduced range	-0.46, 0.082	0.01, 0.968	-0.11, 0.696	-0.15, 0.601	-0.13, 0.644	-0.32, 0.245	0.37, 0.176
Affect: Reduced modulation of intensity	-0.40, 0.144	0.09, 0.755	0.01, 0.966	-0.41, 0.133	-0.49, 0.062	-0.09, 0.744	0.35, 0.195
Affect: Reduced display on demand	-0.31, 0.258	0.14, 0.618	0.11, 0.690	-0.57, 0.026	-0.60, 0.017	-0.04, 0.898	0.51, 0.051
Reduced social drive	-0.59, 0.021	0.14, 0.607	0.08, 0.765	-0.30, 0.270	-0.41, 0.129	-0.06, 0.828	0.48, 0.072
Poor rapport with interviewer	-0.71, 0.002	0.71, 0.003	0.66, 0.007	-0.42, 0.120	-0.50, 0.059	-0.12, 0.673	0.71, 0.003

Table 3. Accuracies of predicting Negative Symptoms using conversational speech features.

Negative Symptoms	SVM	SVR
Prolonged time of response	67%	60%
Restricted speech quantity	73%	53%
Affect reduced display on demand	73%	60%
Reduced social drive	67%	53%
Poor rapport with interviewer	60%	80%

(increased *Failed Interrupts* and *Overlaps*) and a lack of *Interjections*. The relation between other subjective features and objective can be explained in a similar manner. As a next step in our analysis we determined the prediction accuracy for each NSA-16 tool from two multi-class pattern recognition classifiers, viz., Support Vector Machine (SVM) [20] and Support Vector Regression (SVR) [20], trained in a supervised manner. We used subjective ratings as class labels (1-6) and conversational features as feature-set, then we performed leave-one-person-out cross-validation to calculate the prediction accuracy of each criterion. In Table 3 we display only those criteria which could be predicted with more than 60% accuracy. The negative symptoms dimensions of *Prolonged Time to Respond*, *Restricted Speech Quantity*, *Affect Reduced Display on Demand*, *Reduced Social Drive*, and *Poor Rapport with Interviewer* fall under this category.

To determine whether the subjects and the control cases can be distinguished based on the objective features of non-verbal speech cues, we utilized binary SVM and SVR classifiers in a supervised manner with subjects and controls as training target labels, and performed leave-one-person-out cross-validation to calculate the accuracy of classification. Table 4 contains the classification accuracies. We have presented session-wise accuracies as well as combined accuracy for all sessions. The results clearly indicate that audio features for control and subjects contain major differences and thus could be differentiated with a high accuracy. This difference can be attributed to the difference in schizophrenia severity in the subject and control cases. As mentioned earlier, both the subjects and controls were suffering from schizophrenia, but the subjects were undergoing CRT, whereas the controls were not recommended for CRT by doctors. The controls have relatively superior mental health compared to subjects, and this contrast is also translated into the differences in the non-verbal speech cues.

Table 4. Classification of conversational speech features into controls and subjects.

Session	SVM	SVR
Session 1	86%	86%
Session 2	87%	87%
Session 3	87%	80%
Sessions Combined	87%	93%

5. CONCLUSION

The results in Table 3 are promising, albeit based only on data from a relatively small number of patients. We plan to increase the number of participants (both subjects and controls) to obtain more reliable results. These results can be the stepping stone towards building an automated tool which could predict negative symptoms by analysing the speech of a patient in an automated manner. Such tool may serve as an aid to psychologists and could potentially help them in providing better monitoring of schizophrenia patients.

6. ACKNOWLEDGEMENTS

This study was funded by the NMRC Center Grant awarded to the Institute of Mental Health Singapore (NMRC/CG/004/2013) and by NITHM grant M4081187.E30.

7. REFERENCES

[1] Caroline Demily and Nicolas Franck, "Cognitive remediation: a promising tool for the treatment of schizophrenia," 2008.

[2] Alastair G Cardno, E Jane Marshall, Bina Coid, Alison M Macdonald, Tracy R Ribchester, Nadia J Davies, Piero Venturi, Lisa A Jones, Shon W Lewis, Pak C Sham, et al., "Heritability estimates for psychotic disorders: the maudsley twin psychosis series," *Archives of general psychiatry*, vol. 56, no. 2, pp. 162–168, 1999.

[3] 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.

[4] Uriel Heresco-Levy, Daniel C Javitt, Richard Ebstein, Agnes Vass, Pesach Lichtenberg, Gali Bar, Sara Catinari, and Marina Ermilov, "D-serine efficacy as add-on pharmacotherapy to risperidone and olanzapine for treatment-refractory schizophrenia," *Biological psychiatry*, vol. 57, no. 6, pp. 577–585, 2005.

[5] Maria Semkovska, Marc-André Bédard, Lucie Godbout, Frédérique Limoge, and Emmanuel Stip, "Assessment of executive dysfunction during activities of daily living in schizophrenia," *Schizophrenia research*, vol. 69, no. 2, pp. 289–300, 2004.

[6] Susan R McGurk, Elizabeth W Twamley, David I Sitzer, Gregory J McHugo, and Kim T Mueser, "A meta-analysis of cognitive remediation in schizophrenia," *The American journal of psychiatry*, vol. 164, no. 12, pp. 1791–1802, 2007.

[7] Til Wykes, Clare Reeder, Sabine Landau, Brian Everitt, Martin Knapp, Anita Patel, and Renee Romeo, "Cognitive remediation therapy in schizophrenia," *The British journal of psychiatry*, vol. 190, no. 5, pp. 421–427, 2007.

[8] Lynn E DeLisi, "Speech disorder in schizophrenia: Review of the literature and exploration of its relation to the uniquely human capacity for language.," *Schizophrenia Bulletin*, vol. 27, no. 3, pp. 481, 2001.

[9] B Elvevåg, DM Weinstock, M Akil, JE Kleinman, and TE Goldberg, "A comparison of verbal fluency tasks in schizophrenic patients and normal controls," *Schizophrenia Research*, vol. 51, no. 2, pp. 119–126, 2001.

[10] B Elvevåg, T Weickert, M Wechsler, R Coppola, DR Weinberger, and TE Goldberg, "An investigation of the integrity of semantic boundaries in schizophrenia," *Schizophrenia Research*, vol. 53, no. 3, pp. 187–198, 2002.

[11] Brita Elvevåg, Kim Helsen, Marc De Hert, Kim Sweers, and Gert Storms, "Metaphor interpretation and use: a window into semantics in schizophrenia," *Schizophrenia research*, vol. 133, no. 1, pp. 205–211, 2011.

[12] Brita Elvevåg, Peter W Foltz, Mark Rosenstein, and Lynn E DeLisi, "An automated method to analyze language use in patients with schizophrenia and their first-degree relatives," *Journal of neurolinguistics*, vol. 23, no. 3, pp. 270–284, 2010.

[13] Katherine Holshausen, Philip D Harvey, Brita Elvevåg, Peter W Foltz, and Christopher R Bowie, "Latent semantic variables are associated with formal thought disorder and adaptive behavior in older inpatients with schizophrenia," *Cortex*, vol. 55, pp. 88–96, 2014.

[14] Gillinder Bedi, Facundo Carrillo, Guillermo A Cecchi, Diego Fernández Slezak, Mariano Sigman, Natália B Mota, Sidarta Ribeiro, Daniel C Javitt, Mauro Copelli, and Cheryl M Corcoran, "Automated analysis of free speech predicts psychosis onset in high-risk youths," *npj Schizophrenia*, vol. 1, 2015.

[15] Alfonso Troisi, G Spalletta, and A Pasini, "Non-verbal behaviour deficits in schizophrenia: an ethological study of drug-free patients," *Acta Psychiatrica Scandinavica*, vol. 97, no. 2, pp. 109–115, 1998.

[16] Mark Knapp, Judith Hall, and Terrence Horgan, *Nonverbal communication in human interaction*, Cengage Learning, 2013.

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

[18] 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.

[19] Richard SE Keefe, Terry E Goldberg, Philip D Harvey, James M Gold, Margaret P Poe, and Leigh Coughenour, "The brief assessment of cognition in schizophrenia: reliability, sensitivity, and comparison with a standard neurocognitive battery," *Schizophrenia research*, vol. 68, no. 2, pp. 283–297, 2004.

[20] Chih-Chung Chang and Chih-Jen Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*.