

RESEARCH ARTICLE

Non-verbal speech cues as objective measures for negative symptoms in patients with schizophrenia

Yasir Tahir¹, Zixu Yang², Debsubhra Chakraborty¹, Nadia Thalmann¹, Daniel Thalmann¹, Yogeswary Maniam², Nur Amirah binte Abdul Rashid², Bhing-Leet Tan^{2,4}, Jimmy Lee Chee Keong^{2,3}, Justin Dauwels^{5*}

1 Institute for Media Innovation, Nanyang Technological University, Singapore, Singapore, **2** Institute of Mental Health, Singapore, Singapore, **3** Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore, **4** Singapore Institute of Technology, Singapore, Singapore, **5** School of Electrical and Electronics Engineering, Nanyang Technological University, Singapore, Singapore

☯ These authors contributed equally to this work.

* jdauwels@ntu.edu.sg



OPEN ACCESS

Citation: Tahir Y, Yang Z, Chakraborty D, Thalmann N, Thalmann D, Maniam Y, et al. (2019) Non-verbal speech cues as objective measures for negative symptoms in patients with schizophrenia. PLoS ONE 14(4): e0214314. <https://doi.org/10.1371/journal.pone.0214314>

Editor: Sinan Guloksuz, Department of Psychiatry and Neuropsychology, Maastricht University Medical Center, NETHERLANDS

Received: February 13, 2019

Accepted: March 8, 2019

Published: April 9, 2019

Copyright: © 2019 Tahir et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Due to the anonymity guaranteed in the informed consent paperwork, data cannot be publicly shared, and are controlled by the Institute of Mental Health, Singapore. Researchers who wish to request access to these data may contact imhresearch@imh.com.sg.

Funding: This research is supported by the BeingTogether Centre, a collaboration between Nanyang Technological University (NTU) Singapore and University of North Carolina (UNC)

Abstract

Negative symptoms in schizophrenia are associated with significant burden and possess little to no robust treatments in clinical practice today. One key obstacle impeding the development of better treatment methods is the lack of an objective measure. Since negative symptoms almost always adversely affect speech production in patients, speech dysfunction have been considered as a viable objective measure. However, researchers have mostly focused on the verbal aspects of speech, with scant attention to the non-verbal cues in speech. In this paper, we have explored non-verbal speech cues as objective measures of negative symptoms of schizophrenia. We collected an interview corpus of 54 subjects with schizophrenia and 26 healthy controls. In order to validate the non-verbal speech cues, we computed the correlation between these cues and the NSA-16 ratings assigned by expert clinicians. Significant correlations were obtained between these non-verbal speech cues and certain NSA indicators. For instance, the correlation between Turn Duration and Restricted Speech is -0.5, Response time and NSA Communication is 0.4, therefore indicating that poor communication is reflected in the objective measures, thus validating our claims. Moreover, certain NSA indices can be classified into observable and non-observable classes from the non-verbal speech cues by means of supervised classification methods. In particular the accuracy for Restricted speech quantity and Prolonged response time are 80% and 70% respectively. We were also able to classify healthy and patients using non-verbal speech features with 81.3% accuracy.

1 Introduction

Schizophrenia is a chronic and disabling mental disorder with heterogeneous presentations. They are characterized broadly by positive (hallucinations and delusions), negative (avolition,

at Chapel Hill. The BeingTogether Centre is supported by the National Research Foundation, Prime Minister's Office, Singapore under its International Research Centres in Singapore Funding Initiative. This study was also funded by the NMRC Center Grant awarded to the Institute of Mental Health Singapore (NMRC/CG/004/2013) and by NITHM grant M4081187.E30. It is also supported in part by the Rehabilitation Research Institute of Singapore's (RRIS) research grant RRG2/16009. The funders had no roles in study design, data collection, experimental design, or decision to publish.

Competing interests: The authors have declared that no competing interests exist.

anhedonia, asociality, blunted affect and alogia) and cognitive (deficits in attention, memory and executive functioning) symptoms [1–4]. Unlike positive symptoms which are easily identified and effectively treated with pharmacological treatments, negative symptoms tend to be neglected with ineffective treatments [5, 6]. Moreover, negative symptoms have been consistently reported to contribute to poor function and quality of life in patients with schizophrenia [7–9] and highlighted as a significant unmet needs in a large percentage of patients [4].

Negative symptoms are assessed on clinical assessment scales (see [10] for a review) via interviews which relies on interviewer's subjective judgement and normally takes about 15–30 minutes. The implementation of routine negative symptoms assessment in clinical practice is hampered by both training of interviewers and the time. Therefore there is a need to develop a convenient objective measurement of negative symptoms for detection of impairments and monitoring of treatment effectiveness to be used in fast-paced clinical setting.

In recent years, the advancements in speech processing research such as INTERSPEECH competitions [11, 12] and the Audio-Visual Emotion Challenge (AVEC) [13, 14] have paved the way for researchers to investigate the assessment or diagnosis of mental disorders utilizing speech analysis. Non-verbal speech or voice based cues such as prosodic features, formant features, source features and spectral analysis features were extracted from speech and machine learning models would be applied on these features to predict the existence or the severity of various mental disorders such as depression, PTSD, and autism [15–17]. The speech features could be used along with traditional clinical ratings to improve the detection of the mental illness [18].

Previous non-automated attempts to utilize the different aspects of speech and language as differentiators between individuals with schizophrenia and healthy individuals have had limited success. Although there existed some distinction in verbal fluency tasks between patients and healthy controls [19], other studies involving semantic boundary [20] or metaphor interpretation [21] reported 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 and signal processing techniques. The ability of individuals with schizophrenia to express emotions was compared to that of a healthy group [22]. Subtle differences in communication discourses were detected among patients, their first-degree relatives and healthy controls employing Latent Semantic Analysis (LSA) in [23]. LSA was again utilized to identify lack of semantic and phonological fluency, disconnected speech, and thought disorder in [24], and LSA and machine learning were used to analyze free-speech and predict the onset of psychosis respectively of high-risk youths in [25].

However, all the above methods are based on semantic analysis and natural language processing. Restricted non-verbal cues display of patients suffering from negative symptom schizophrenia [26] have not been explored much. In [27] computerized measures of flat affect, alogia and anhedonia were examined in their relationships to clinically-rated negative symptoms and social functioning for 14 patients with flat affect, 46 patients without flat affect and 19 healthy controls. The results suggest that speech rate measure significantly discriminated patients with flat affect from controls. The computer measures of alogia and negative emotions significantly discriminated flat affect and non-flat affect patients. In [28] a system was designed to assess the speech for patient's negative symptoms. Results show a relationship between dyadic interaction and flat affect, pause production and alogia, and rate and asociality. In [29] acoustic analysis of speech in response of visual stimuli was conducted for 48 stable outpatients for schizophrenia and mood disorders. The results show that computerized acoustic analysis appears to be a promising method for understanding the pathological manifestation of these disorders. In [30] the correlation between the variability of tongue front and back movement,

and formant frequencies to the negative symptoms was studied for 25 first episode psychosis patients. In [31] interviews of 20 subjects with flat affect, 26 with non-flat affect, and 20 healthy controls were analysed to determine the motor expressive deficiency in schizophrenic patients. In [18] clinical ratings of flat affect and alogia were compared to the patient's speech prosody and productivity. The results suggest that acoustic analysis can provide objective measures that may help in clinical assessment. In [32] a semi-automatic method was employed to quantify the degree of expressive prosody deficits in schizophrenia for 45 patients and 35 healthy controls. The results suggest that using non-verbal speech analysis the researchers were capable of classifying patient and controls with 93.8% accuracy. Non-verbal speech cues such as voice tone, volume, and interjections play a crucial role in human interaction and communication [33], and the display of such signals in patients can be used for both distinguishing them with healthy controls and developing specific and objective treatments. In existing work speech analysis has mostly been used to determine the presence and/or the severity of symptoms.

In this study we built upon our earlier work [34] and attempted to explore the utility of non-verbal speech cues of determine the severity of negative symptoms in schizophrenia. Specifically, we study the correlations between subjective ratings of negative symptoms on a clinical scale during interviews and the objective non-verbal speech features extracted from audio recording of the same interview.

2 Methods

2.1 Subjects

Fifty-four outpatients diagnosed with schizophrenia from the Institute of Mental Health, Singapore and twenty-six healthy individuals from general population were recruited in this study. The inclusion criteria of the study included diagnosis of schizophrenia for patient group or no mental illness for control group, aged 16–65, English speaking and fit to provide informed consent. The capacity of consent was assessed by asking participants to describe the purpose and procedures of the study to interviewers. All the participants finally selected for the study were consenting adults, above 18 years of age. The exclusion criteria included history of strokes, traumatic brain injuries and neurological disorders such as epilepsy. The Structured Clinical Interview for DSM-IV (SCID) was conducted for all participants to ascertain the diagnoses of schizophrenia for patients and no mental illness for healthy individuals by trained research psychologists. Ethics approval for the study was provided by the National Healthcare Group Domain Specific Review Board. Written informed consent was obtained from all participants. The sample characteristics were presented in [Table 1](#).

2.2 Clinically-rated symptom measures

The 16-item Negative Symptom Assessment (NSA-16) [35] is a reliable and validated scale used to measure the severity of negative symptoms through semi-structured interview. It contains 16 items, each rated on a 6-point Likert scale where higher ratings indicate more severe impairments. In addition to the individual item scores, the scale also provides a global negative symptom rating (based on the overall clinical impression of negative symptoms in the patient), a total score (sum of the scores on the 16 items), and five negative symptoms domains scores including Communication (sum of the scores of item 1-4), Affect/Emotion (sum of the scores of item 5-7), Social Involvement (sum of the scores of item 8-10), Motivation (sum of the scores of item 11-14), and Retardation (sum of the scores of item 15 and 16). The NSA-16 demonstrated high internal consistency (Cronbach's alpha = 0.92) and inter-rater reliability (Kappa value = 0.89) [36]. NSA was rated by trained research psychologists and inter-rater reliability was 0.92. The meanings of NSA items and domains are shown in [Table 2](#).

Table 1. Sample characteristics.

	Schizophrenia sample	Healthy Control Sample	p-values
Gender (male:female)	25:29	12:14	0.990
Age (years)	31.06 ± 7.52	29.58 ± 8.09	0.424
Total years of education	13.67 ± 2.76	13.53 ± 2.23	0.825
Duration of illness	9.06 ± 7.36	-	
Age of illness onset	22.56 ± 5.30	-	
Medication (%)			
Antipsychotics	94.44		
Typical antipsychotics	7.41	-	
Atypical antipsychotics	81.48	-	
Anticholinergics	25.93	-	
Antidepressants	31.48	-	
Mood Stabilizers	25.93	-	
Benzodiazepine	14.81	-	
CPZ equivalence (mg/day)	412.19 ± 352.01	-	
BPRS Total Score	32.81 ± 8.86	19.81 ± 1.86	< 0.001
NSA Total Score	41.28 ± 9.39	26.77 ± 3.77	< 0.001
NSA—Communication Domain Score	7.96 ± 3.38	4.46 ± 0.71	< 0.001
NSA—Emotion Affect Domain Score	8.65 ± 1.99	6.04 ± 1.73	< 0.001
NSA—Social Involvement Domain Score	9.02 ± 2.67	7.04 ± 2.01	< 0.001
NSA—Motivation Domain Score	11.85 ± 2.68	6.92 ± 1.94	< 0.001
NSA—Retardation Domain Score	3.80 ± 1.81	2.31 ± 0.47	< 0.001

CPZ = Chlorpromazine;

BPRS = The Brief Psychiatric Rating Scale;

NSA = Negative Symptom Assessment

<https://doi.org/10.1371/journal.pone.0214314.t001>

2.3 Acquisition of non-verbal speech features

2.3.1 Equipment and procedure. The system deployed in this paper is based on our earlier work [37, 38], where we developed a machine learning based system that is able to infer the levels of interest, dominance, and agreement with 85%, 86% and 82% accuracy respectively from dyadic conversations. 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 as a 2-channel audio.wav file (one channel for each speaker). This file allows us to detect which speaker is speaking at any given time.

The patient and the psychologist wore their respective microphones during the whole interview of NSA and the whole conversation was recorded. There was no pre-determined duration for the interview; instead it depended on participant's response to the questions asked by the psychologist. On average, the interviews lasted about 30 minutes. Before the actual recording, we ensured that all the devices are connected. During the interview the software applications were monitored from another room via remote desktop to ensure the recording device functioned properly, and simultaneously maintain confidentiality of the participants' speech.

2.3.2 Extraction of non-verbal cues. We considered two types of low-level speech metrics: conversational and prosody related cues. The conversational cues accounted for who was speaking, when and by how much, while the prosodic cues quantified how people talked during their conversations. A detailed list of conversational cues is shown in Table 3. We used

Table 2. NSA items and their explanations.

Label	Criteria	Explanation
NSA 1	Prolonged time to respond	After asking the subject a question, he/she pauses for inappropriately long periods before answering
NSA 2	Restricted speech quantity	Ratings on this item suggest that the subject gives brief answers to questions and/or provides elaborating details only after the interviewer prods him
NSA 3	Impoverished speech content	The subject may talk a lot or a little but the information conveyed is very limited
NSA 4	Inarticulate speech	The subject's speech cannot be understood because enunciation is poor
NSA 5	Emotion: Reduced range	Emotion is the feeling content of a person's inner life. This item assesses the range of emotion experienced by the subject during the last week (or other specified time period)
NSA 6	Affect: Reduced modulation of intensity	This item assesses the subject's modulations of intensity of affect shown during the interview while discussing matters that would be expected to elicit significantly different affective intensities in a normal person
NSA 7	Affect: Reduced display on demand	This items assesses the subject's ability to display a range of affect as expressed by changes in his/her facial expression and gestures when asked by the interviewer to show how his/her face appears when he/she feels happy, sad, proud, scared, surprised, and angry
NSA 8	Reduced social drive	This item assesses how much the subject desires to initiate social interactions. Desire may be measured in part by the number of actual or attempted social contacts with others
NSA 9	Poor rapport with interviewer	This item assesses the interviewer's subjective sense that he/she and the subject are actively engaged in communication with one another
NSA 10	Interest in emotional and physical intimacy	This item assesses how much the subject retains interest in emotional and physical intimacy or sexual activity
NSA 11	Poor grooming and hygiene	The subject presents with poorly groomed hair, dishevelled clothing, etc.
NSA 12	Reduced sense of purpose	This item assesses whether the subject possesses integrated goals for his/her life
NSA 13	Reduced interests	This item assesses the range and intensity of the subject's interests
NSA 14	Reduced daily activity	This item assesses the level of the subject's daily activity and his/her failure to take advantage of the opportunities his/her environment offers
NSA 15	Reduced expressive gestures	Gestures and body movements that normally facilitate communication during speech are less than normal, or are not observed at all
NSA 16	Slowed movements	This item assesses how much the subject's voluntary movements are slowed. At a minimum, one should rate movements as gait and those of rising from a chair
NSA 17	Global negative symptoms rating	This item assesses the overall impression of negative symptoms in the subject
NSA 18	NSA total	Sum of the ratings from questions 1-16
NSA 19	NSA communication	Sum of the ratings from questions 1-4
NSA 20	NSA emotion affect	Sum of the ratings from questions 5-7
NSA 21	NSA social involvement	Sum of the ratings from questions 8-10
NSA 22	NSA motivation	Sum of the ratings from questions 11-14
NSA 23	NSA retardation	Sum of the ratings from questions 15-16

<https://doi.org/10.1371/journal.pone.0214314.t002>

Table 3. List of conversational, and prosodic features.

Category	Features
Conversational	
Speaking duration	Speaking %, Mutual silence, Difference in Speaking %, Overlap, Response time
Speaking turns	Natural turns, Turn duration
Interruption	Interruptions, Failed interruptions
Interjection	Interjection, Speaking interjection
Prosodic	
Frequencies	Larynx frequency (F0), Formant (F1, F2, F3)
MFCC	Mel-frequency cepstral coefficients
Amplitude	Mean volume, Max volume, Min volume, Entropy

<https://doi.org/10.1371/journal.pone.0214314.t003>

Matlab to compute 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 [37]. Fig 1 explains extraction of some of the conversational cues.

We considered the following prosodic cues: amplitude, larynx frequency (F0), formants (F1, F2, F3), and mel-frequency cepstral coefficients (MFCCs). These cues were extracted from 30 ms segments at a fixed interval of 10 ms; they tended to fluctuate rapidly in time. Therefore, we computed various statistics of those cues over a time period of several seconds, including minimum, maximum, mean and entropy. The prosodic features are the standard audio features used in research, but the conversational features have been designed specifically for dyadic conversations. Table 4 provides the definition for these conversational features.

3 Statistical analyses

First, Matlab was used to test the Pearson’s correlation between the objective audio features and the subjective negative symptoms ratings. In the second step, we analyzed the automated prediction of negative symptoms from audio features. We determined the prediction accuracy

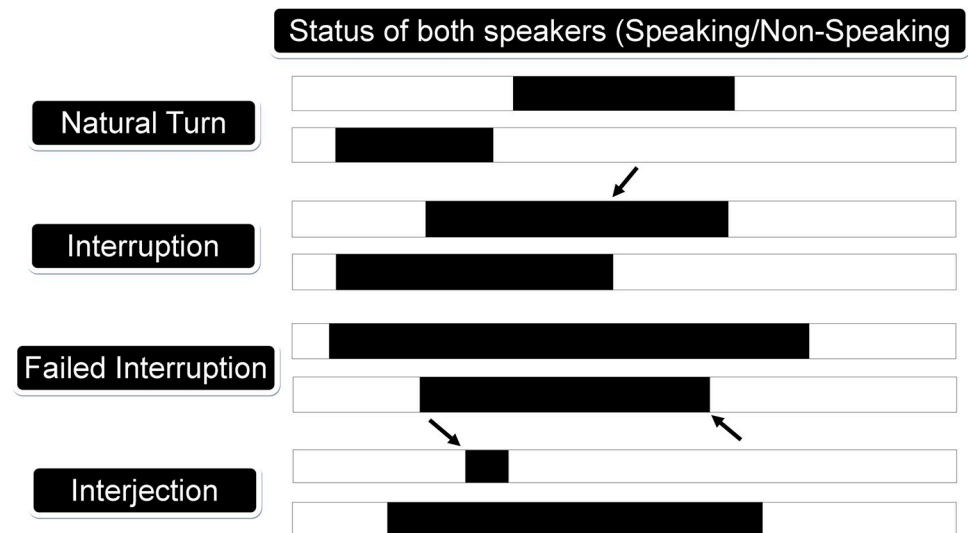


Fig 1. Illustration of conversational cues or features. Periods of speaking and non-speaking are indicated in black and white respectively.

<https://doi.org/10.1371/journal.pone.0214314.g001>

Table 4. Explanation of non-verbal conversational cues.

Non-Verbal Feature	Description
Natural Turn-Taking	The number of times person 'A' speaks in the conversation without interrupting person 'B' (see Fig 1). Normalized to per minute.
Turn Duration	The average duration of a speaker's turn.
Speaking %	The percentage of time a person speaks in the conversation.
Speaking % Difference	The difference between the speaking percentages of both speakers.
Mutual Silence %	The percentage of time when both participants are silent.
Interruption	Person 'A' interrupts person 'B' while speaking, and takes over. Person 'B' stops speaking before person 'A' does (see Fig 1).
Speaking Interjection	Short utterances such as 'okay', 'hmm' etc. when other speaker is speaking (see Fig 1).
Speech Gap	The gap that a person takes between his/her consecutive turns.
Response Time	If person 'A' finishes speaking, then the time taken for person 'B' to start speaking is called response time.

<https://doi.org/10.1371/journal.pone.0214314.t004>

for some NSA-16 criteria. The rating scale ranges from 1-6, where a value of 1 and 2 would be coded as non-observable group and a value between 3 and 6 would be coded as observable. We then used classification algorithms to determine the accuracy with which observable and non-observable classes can be differentiated. We performed leave-one-patient-out cross-validation to calculate the accuracy and AUC for these criteria. For feature selection we utilized CFSsubset attribute selection [39], and Correlation attribute selection [40]. CFSsubset attribute selection evaluates the worth of a subset of features by considering the individual predictive ability of each feature along with the degree of intercorrelation between the features. Subsets of features that are highly correlated with the class while having low intercorrelation are preferred. Correlation attribute selection [40] evaluates the worth of an attribute by measuring the Pearson correlation between it and the class label. At the end we present the results for the classification of the healthy controls and individuals with schizophrenia. The classification 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.

4 Results

4.1 The correlations between non-verbal speech features and NSA scores

The colormaps in Fig 2 showed the correlation of NSA-16 criteria with non-verbal conversational audio features, where Table 5 presents the values for significant correlations. It can be seen from the colormap that features like Failed Interrupt, Overlap, Mutual Silence, Speech Gap and Response Time are directly correlated to the negative symptoms; on the other hand Natural Turns, Interjections, Interrupts, Speaking Percentage and Turn Duration are inversely correlated to the negative symptoms.

4.2 The automated prediction of negative symptoms from audio features

Table 6 presents the results for Prolonged time to respond, Restricted speech quantity, Impoverished speech content, Emotion: Reduced range, Affect: Reduced modulation of intensity, Reduced social drive, and Reduced expressive gestures along with the best five features in individuals with schizophrenia.

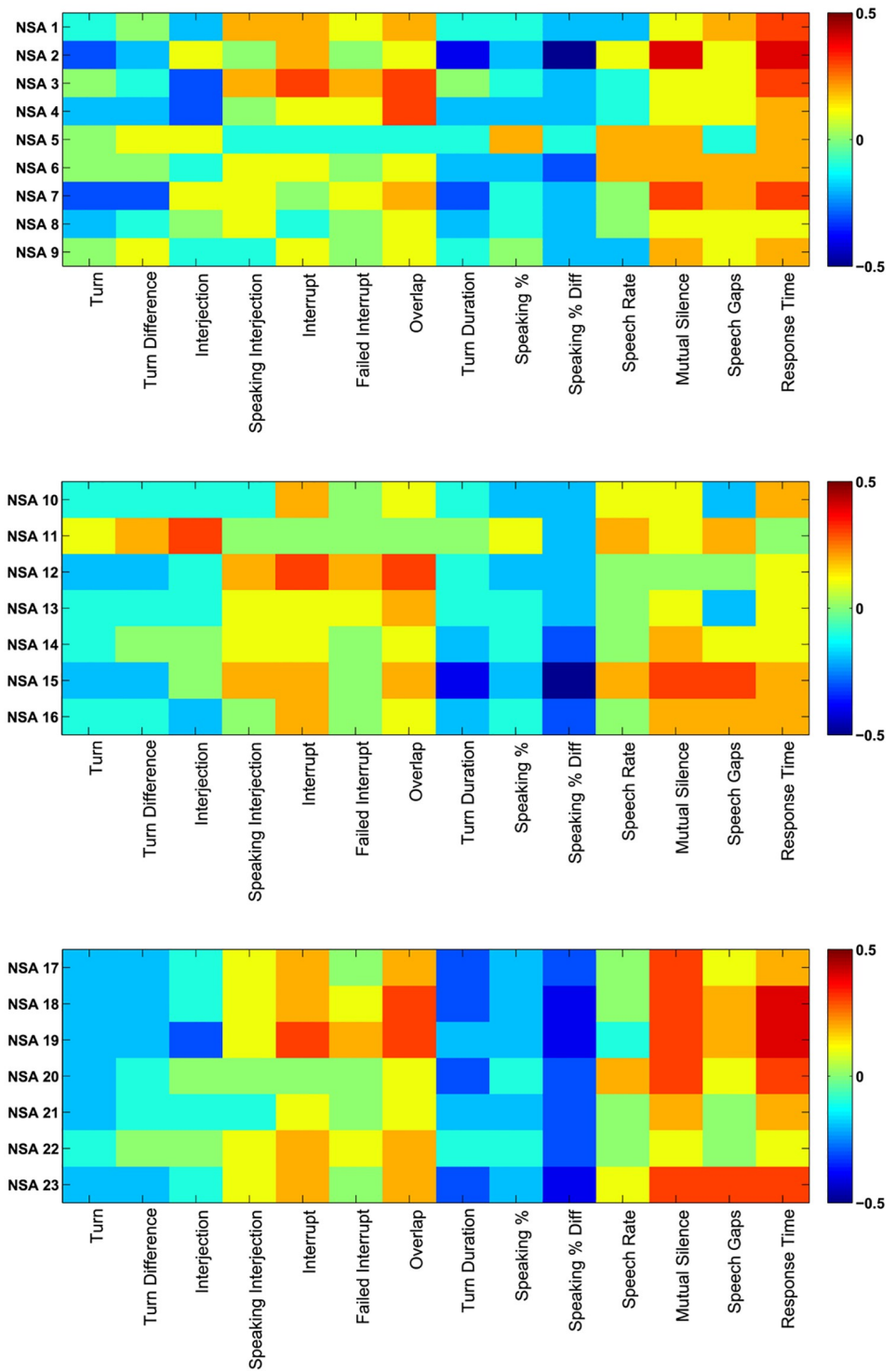


Fig 2. Colormap plots of NSA-16 ratings. Colormap plots between (a) NSA-16 features 1-9 and conversational features, (b) NSA-16 features 10-16 and conversational features, and (c) NSA-16 features 17-23 and conversational features.

<https://doi.org/10.1371/journal.pone.0214314.g002>

Table 5. Correlation values between NSA and speech features.

	Natural Turn	Difference Turn	Interject	Speaking Interject	Interrupt	Failed Interrupt	Overlap	Speaking	Difference Speaking	Turn Duration	Speaking Rate	Mutual Silence	Speech Gap	Response Time
NSA 1	-0.074 (0.593)	-0.039 (0.782)	-0.229 (0.096)	0.195 (0.158)	0.241 (0.079)	0.091 (0.515)	0.169 (0.223)	-0.081 (0.563)	-0.076 (0.584)	-0.174 (0.209)	-0.155 (0.265)	0.117 (0.399)	0.245 (0.075)	0.307 (0.024)
NSA 2	-0.276 (0.044)	-0.176 (0.204)	0.05 (0.718)	0.025 (0.859)	0.154 (0.267)	-0.035 (0.803)	0.098 (0.482)	-0.437 (0.001)	-0.178 (0.197)	-0.461 (0.001)	0.135 (0.331)	0.404 (0.002)	0.149 (0.283)	0.355 (0.008)
NSA 3	-0.044 (0.754)	-0.055 (0.691)	-0.309 (0.023)	0.177 (0.199)	0.277 (0.043)	0.246 (0.073)	0.296 (0.03)	-0.027 (0.844)	-0.102 (0.463)	-0.227 (0.099)	-0.125 (0.366)	0.096 (0.491)	0.129 (0.351)	0.29 (0.033)
NSA 4	-0.195 (0.158)	-0.205 (0.137)	-0.265 (0.053)	-0.016 (0.908)	0.088 (0.527)	0.137 (0.324)	0.255 (0.063)	-0.163 (0.238)	-0.24 (0.081)	-0.2 (0.147)	-0.146 (0.293)	0.119 (0.39)	0.054 (0.697)	0.216 (0.116)
NSA 5	-0.005 (0.973)	0.132 (0.34)	0.138 (0.318)	-0.14 (0.311)	-0.112 (0.42)	-0.105 (0.448)	-0.121 (0.383)	-0.081 (0.559)	0.16 (0.247)	-0.061 (0.66)	0.185 (0.18)	0.172 (0.214)	-0.051 (0.715)	0.192 (0.164)
NSA 6	-0.044 (0.752)	-0.048 (0.731)	-0.108 (0.439)	0.074 (0.596)	0.144 (0.3)	-0.012 (0.933)	0.126 (0.362)	-0.172 (0.213)	-0.165 (0.232)	-0.323 (0.017)	0.237 (0.085)	0.157 (0.258)	0.153 (0.268)	0.226 (0.1)
NSA 7	-0.349 (0.01)	-0.326 (0.016)	0.053 (0.703)	0.102 (0.462)	0.019 (0.889)	0.144 (0.299)	0.248 (0.071)	-0.326 (0.016)	-0.14 (0.314)	-0.183 (0.186)	0.036 (0.796)	0.342 (0.011)	0.187 (0.175)	0.285 (0.036)
NSA 8	-0.224 (0.103)	-0.076 (0.586)	-0.01 (0.944)	0.095 (0.493)	-0.056 (0.688)	-0.019 (0.892)	0.071 (0.609)	-0.224 (0.104)	-0.139 (0.316)	-0.162 (0.241)	-0.045 (0.747)	0.119 (0.393)	0.067 (0.631)	0.092 (0.509)
NSA 9	-0.027 (0.846)	0.059 (0.671)	-0.073 (0.599)	-0.092 (0.508)	0.147 (0.29)	0.027 (0.846)	0.081 (0.559)	-0.07 (0.613)	0.014 (0.918)	-0.17 (0.219)	-0.158 (0.255)	0.219 (0.112)	0.137 (0.323)	0.226 (0.1)
NSA 13	-0.068 (0.625)	-0.064 (0.648)	-0.105 (0.451)	0.062 (0.656)	0.091 (0.514)	0.072 (0.605)	0.171 (0.218)	-0.057 (0.684)	-0.105 (0.449)	-0.151 (0.275)	-0.019 (0.893)	0.054 (0.698)	-0.198 (0.15)	0.147 (0.288)
NSA 14	-0.069 (0.618)	-0.023 (0.87)	0.027 (0.848)	0.111 (0.426)	0.103 (0.459)	0.029 (0.834)	0.14 (0.314)	-0.155 (0.263)	-0.101 (0.465)	-0.28 (0.04)	-0.026 (0.853)	0.17 (0.22)	0.12 (0.387)	0.084 (0.548)
NSA 15	-0.222 (0.107)	-0.175 (0.205)	-0.047 (0.734)	0.188 (0.172)	0.165 (0.234)	0.035 (0.801)	0.172 (0.214)	-0.359 (0.008)	-0.232 (0.092)	-0.466 (0.001)	0.199 (0.149)	0.332 (0.014)	0.306 (0.025)	0.246 (0.073)
NSA 16	-0.08 (0.564)	-0.125 (0.367)	-0.179 (0.196)	0.046 (0.742)	0.232 (0.091)	0.012 (0.93)	0.116 (0.402)	-0.161 (0.244)	-0.142 (0.304)	-0.273 (0.046)	-0.006 (0.963)	0.206 (0.135)	0.24 (0.08)	0.179 (0.196)
NSA Global	-0.242 (0.077)	-0.168 (0.225)	-0.082 (0.557)	0.099 (0.476)	0.188 (0.173)	0.018 (0.898)	0.161 (0.244)	-0.273 (0.046)	-0.171 (0.215)	-0.311 (0.022)	0.02 (0.886)	0.271 (0.047)	0.118 (0.396)	0.161 (0.246)
NSA 18	-0.24 (0.08)	-0.157 (0.258)	-0.131 (0.345)	0.109 (0.435)	0.248 (0.071)	0.105 (0.452)	0.263 (0.055)	-0.301 (0.027)	-0.217 (0.115)	-0.432 (0.001)	0.027 (0.845)	0.307 (0.024)	0.168 (0.225)	0.369 (0.006)
NSA 19	-0.201 (0.144)	-0.157 (0.256)	-0.259 (0.058)	0.148 (0.287)	0.278 (0.041)	0.151 (0.277)	0.282 (0.039)	-0.248 (0.07)	-0.2 (0.147)	-0.377 (0.005)	-0.099 (0.477)	0.264 (0.054)	0.215 (0.118)	0.42 (0.002)
NSA 20	-0.181 (0.19)	-0.112 (0.419)	0.03 (0.828)	0.021 (0.879)	0.032 (0.816)	0.012 (0.933)	0.122 (0.379)	-0.271 (0.048)	-0.076 (0.585)	-0.274 (0.045)	0.22 (0.111)	0.31 (0.023)	0.14 (0.313)	0.329 (0.015)
NSA 21	-0.221 (0.108)	-0.112 (0.42)	-0.075 (0.592)	-0.051 (0.712)	0.145 (0.296)	0.03 (0.832)	0.12 (0.386)	-0.21 (0.128)	-0.173 (0.21)	-0.254 (0.063)	-0.043 (0.759)	0.188 (0.174)	-0.026 (0.851)	0.236 (0.086)
NSA 22	-0.107 (0.44)	-0.036 (0.796)	0.002 (0.99)	0.131 (0.345)	0.197 (0.154)	0.118 (0.394)	0.239 (0.081)	-0.118 (0.395)	-0.129 (0.352)	-0.281 (0.039)	0.014 (0.922)	0.11 (0.427)	0.022 (0.875)	0.115 (0.406)
NSA 23	-0.186 (0.179)	-0.178 (0.198)	-0.121 (0.385)	0.147 (0.29)	0.225 (0.103)	0.029 (0.834)	0.171 (0.216)	-0.317 (0.02)	-0.224 (0.104)	-0.443 (0.001)	0.127 (0.359)	0.321 (0.018)	0.321 (0.018)	0.251 (0.067)

<https://doi.org/10.1371/journal.pone.0214314.t005>

Table 6. Classification results for NSA criteria.

NSA Criteria	Algorithm	Feature Selection	Confusion Matrix		Accuracy	AUC	Best Features
			Non-Observable	Observable			
NSA 1	SVM	Correlation	33	4	79.6%	0.74	Ent_F2, Ent_F3, Ent_F1 Speech_Gap, Response_Time
			7	10			
NSA 2	SVM	Correlation	26	7	70.4%	0.68	Max_Vol, Mean_Vol, Ent_Freq MFCC, Mutual_Silence
			9	12			
NSA 3	SVM	Correlation	18	9	59.3%	0.59	MFCC2, Overlap, Response_Time Failed_Interrupt, MFCC1
			13	14			
NSA 5	SVM	Correlation	13	12	53.7%	0.54	Ent_Vol, MFCC8, Max_Vol MFCC, Mean_Vol
			13	16			
NSA 6	SVM	Correlation	13	13	59.3%	0.59	Turn_Duration, Speech_Gap, MFCC5 MFCC10, Response_Time
			9	19			
NSA 8	SVM	Correlation	6	7	74.1%	0.65	MFCC2, Speaking, MFCC4 MFCC3, MFCC8
			7	34			
NSA 15	SVM	CFSsubset	30	5	77.8%	0.74	Turn_Duration, Ent_Freq, Mutual_Silence Max_Vol, Mean_Vol
			7	12			

<https://doi.org/10.1371/journal.pone.0214314.t006>

4.3 Classification of patients v/s healthy subjects

Table 7 presents the patients vs healthy controls classification results for various machine learning algorithms along with the best five features. In the last row we present the statistics for a baseline classifier which classifies every participant as a patient. The highest accuracy is

Table 7. Patient vs healthy classification.

Algorithm	Feature Selection	Confusion Matrix		Accuracy	AUC	Precision (Patient)	Recall (Patient)	F-Score (Patient)	Best Features
		Patient	Healthy						
SVM	ReliefF	40	14	70%	0.68	0.8	0.74	0.77	speech_gap, ent_freq, mean_vol ent_f1, ent_f3
		10	16						
Random Forest	CFSsubset	42	12	72.5%	0.8	0.81	0.78	0.79	speaking, speech_gap, ent_freq ent_f3, mfcc
		10	16						
MLP	None	44	10	81.3%	0.9	0.9	0.82	0.85	
		5	21						
Ensemble (Bagging)	CFSsubset	46	8	77.5%	0.8	0.82	0.85	0.84	speaking, speech_gap, ent_freq ent_f3, mfcc
		10	16						

<https://doi.org/10.1371/journal.pone.0214314.t007>

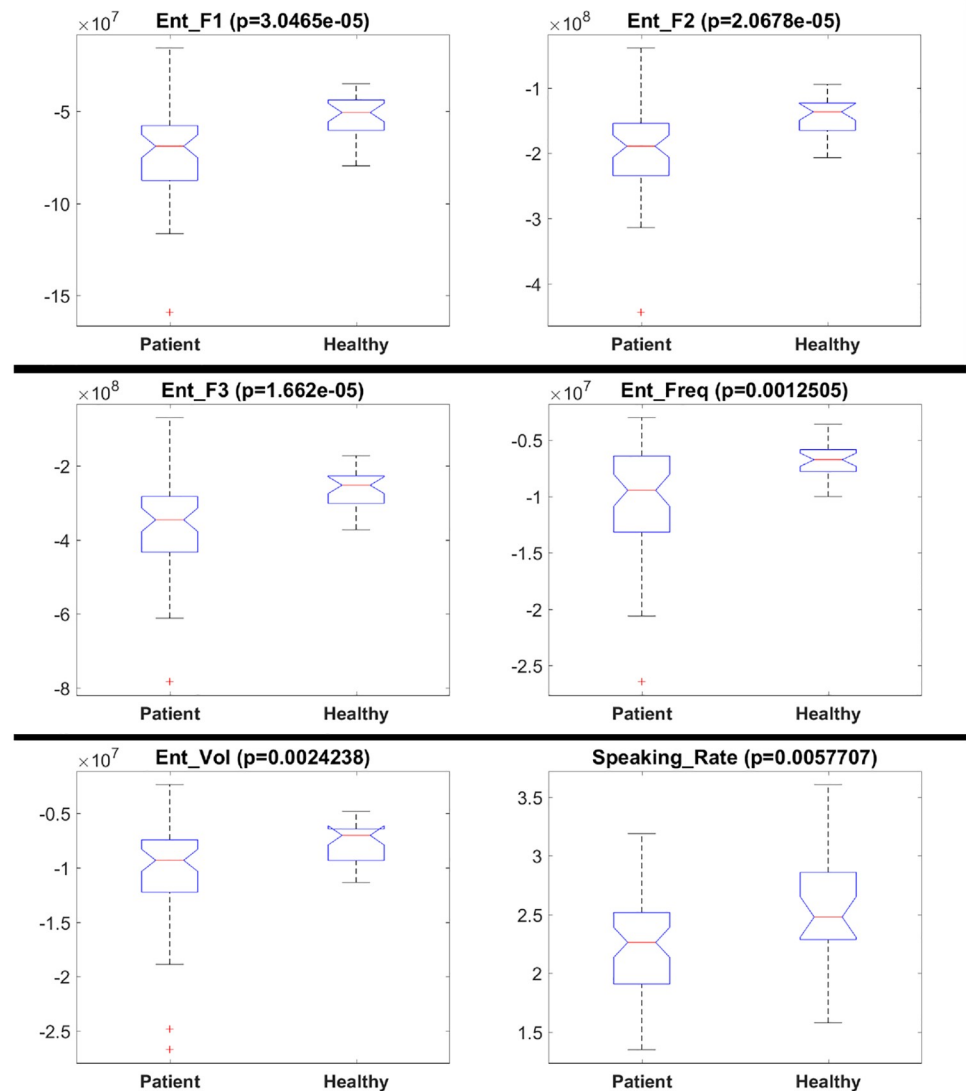


Fig 3. Boxplots for features that are significantly different for healthy subjects and patients. Boxplots for (a) F1 Entropy, (b) F2 Entropy, (c) F3 Entropy, (d) Frequency Entropy, (e) Volume Entropy, and (f) Speaking Rate.

<https://doi.org/10.1371/journal.pone.0214314.g003>

81.3% which is really promising, because it shows that using non-verbal speech features we can differentiate between healthy individuals and individuals with schizophrenia.

Fig 3 presents boxplots and results (p-values) for the Kruskal-Wallis test. We tested whether a feature is significantly different for the healthy and patient groups. The plots are only shown for features with p-values below 0.01 for the Kruskal-Wallis test. It can be seen from the figures that frequency and volume entropies show significant difference among the prosodic features, while speaking rate shows significant difference among the conversational features. The healthy group has higher frequency and volume entropy as compared to the patient group. This finding implies that the healthy subjects speak in a less monotonous manner compared to the patients, and have more variability in the volume of their speech. Similarly, the speaking rate seems to be higher for the healthy group. These findings are in line with the results presented in [27], where speech rate significantly discriminated patients and healthy controls, and [31], where patients were found to be less fluent.

5 Discussion

As can be observed from Fig 2, significant correlations exist between the subjective ratings (NSA-16 items) and the objective measures (conversational features). An interesting point to note here is that the absolute values of the correlations are higher in Fig 2(a) and 2(c) compared to those in Fig 2(b). This difference can be attributed to the fact that NSA-16 items 1-9 (see Table 2) are closely associated with speech impairments. Consequently, these specific NSA-16 items are in greater congruence with the objective speech-related measures compared to the rest of the NSA-16 items, yielding higher absolute values of correlations. Similarly, Fig 2(c) contains the cumulative NSA-16 items, which is reflected in the overall higher absolute correlations.

Table 5 presents the correlation values for the NSA-16 items and speech features, providing the complete picture by listing all the correlations and their corresponding p-values. It can be noted that features such as Failed Interrupts, Response Time and Overlap correlate directly with the negative symptoms, i.e., cases which saw an increased Response Time or Overlap during the interview also saw a higher rating on equivalent items of the subjective NSA-16 scale, indicating greater severity of negative symptoms. The reverse situation occurred with the features such as Natural Turns, Speaking %, or Turn Duration; the reduced values of such features, associated often with disjoint communication, saw increased ratings on the speech-related NSA-16 items, resulting in negative correlations.

It can be seen from the Table 5 that Response Time has significant positive correlations with the NSA criteria Prolonged Time to Respond, Restricted Speech Quantity, Impoverished Speech Content, NSA Total, NSA Communication, and NSA Emotion Affect. Mutual Silence has significant positive correlations with the NSA criteria Restricted Speech Quantity, Affect: Reduced Display on Demand, NSA Total, NSA Communication, NSA Emotion Affect, and NSA Retardation. Speaking % has significant negative correlations with the NSA criteria Restricted Speech Quantity, Affect: Reduced Display on Demand, NSA Total, NSA Emotion Affect, and NSA Retardation. Turn Duration has significant negative correlations with the NSA criteria Restricted Speech Quantity, Affect: Reduced Modulation of Intensity, NSA Total, NSA Communication, NSA Emotion Affect, and NSA Retardation.

An interesting observation is the significant correlations between Reduced Expressive Gestures with non-verbal speech features. It has significant positive correlation with Mutual Silence, and Speech Gap, showing a decrease in gesture usage if there is more silence. On the other hand Reduced Expressive Gestures has significant negative correlation with Turn Duration, and Speaking %, which shows that the gesture usage increases with increase in speech.

The results in Table 6 present the detection accuracies for Prolonged time to respond, Restricted speech quantity, Impoverished speech content, Emotion: Reduced range, Affect: Reduced modulation of intensity, Reduced social drive, and Reduced expressive gestures. We achieve higher accuracy of 79.6%, 77.8% and 70.4% for Prolonged time to respond, Reduced expressive gestures and Restricted speech quantity. For Impoverished speech content, Emotion: Reduced range, and Affect: Reduced modulation of intensity, we achieve rather moderate accuracies of 59.3%, 53.7%, and 59.3%. The NSA item Reduced social drive has a large imbalance between non-observable and observable classes. These results correspond to Table 5, where high correlation values lead to a better prediction in most of the cases.

The results in Table 7 present the accuracies for patient v/s healthy classification. The highest accuracy of 81.3% was achieved by Multi-layer perceptron. Other algorithms like SGD, Random Forest, and Random-subspace (ensemble) achieve 70.0%, 72.5%, and 77.5% respectively [41].

In this paper we have presented our study on the objective and automated analysis of the speech deficiencies associated with schizophrenia. Our study is unique in several aspects. First of all, our dataset contains 80 subjects, including 56 patients and 24 healthy controls, which is a larger group than in most related studies. Moreover, the recordings are about 30 minutes long, and are substantially longer than recordings in similar studies typically lasting only a few minutes. Also we have not edited the recordings in any way, and analyzed the entire recordings, instead of selecting particular segments. As a result, our approach could be applied in realistic settings such as face-to-face interviews or phone calls, as there is no need for manual editing.

Moreover, the semi-structured nature of our participant interviews attempts to recreate an environment that is close to real-life clinical settings. We are interested in the social and cognitive behavioral patterns of the participants in their everyday lives, hence we decided not to apply any external stimuli, in contrast to earlier studies [29].

Our approach is more comprehensive compared to earlier studies, as we combine both prosodic as well conversational speech cues; this allowed us to conduct a more in-depth analysis. These objective cues, once validated through their strong correlations with established subjective measures, were utilised to predict the aforementioned subjective measures and to distinguish patients from healthy controls. The conversation speech cues used in this study and the correlations between NSA items and these conversational features are unique to our research.

A few earlier studies describe their attempt to develop automated methods to analyze audio and speech features of individuals suffering from schizophrenia. However, these approaches have the following limitations. Either the studies only consider prosodic cues related to flattening of affect or alogia as in [29], [30], [31], and [18], or are limited in the duration of speech data as in [30] (duration—1 minute), [31] (duration—10 minutes) and [32] (first paragraph read from a medieval classic). Only Cohen et al. [27] explored a single conversational feature (speech rate) together with other prosody features (inflection). None of the above studies utilized these speech features to distinguish between the patients and healthy controls with the exception of the one conducted by Martinez et al. [32], who achieved a classification accuracy of 93.8%. However, as pointed out earlier, their data is of rather short duration, and hence, these results may be less reliable compared to our results, obtained from 30 minutes recordings. Such long-term recordings give us the opportunity to explore the correlations between non-verbal speech cues and negative symptoms in individuals with schizophrenia with greater depth. We believe that more studies of this nature are required, with more and longer recordings in realistic settings, to fully establish the effectiveness of non-verbal audio cues for assessing the negative symptoms of schizophrenia. These results can be the stepping stone towards building an automated tool which could predict negative symptoms by analyzing the speech of a patient in an automated manner. The results of our patient v/s healthy classification analysis are also very promising. It shows that participants can indeed be classified as individuals with schizophrenia or as healthy individuals on the basis of their non-verbal speech features.

Also, in the future, we plan to explore the temporal variation of the speech cues for the subjects and controls, specifically, how their speech features change over sessions, and, if at all, in different manner for subjects and controls. In this paper, we have only investigated the non-verbal cues related to speech. Visual non-verbal cues such as posture, gestures, etc., have not been investigated here, which we plan to address in the future.

The present study is not without limitations. As this is a cross-sectional study, no conclusion could be drawn about the stability of the relationship between the objective measure and subject measure over time. Some factors that might affect negative symptoms and speech production such as participants' smoking history and extrapyramidal symptoms were not assessed and controlled in this study. The correlations between the objective speech cues and

NSA ratings ranged between 0.5 and -0.5. Therefore there were a big percentage of variance could not be explained. In addition, this study only tested the relationship between objective measures and subjective measures of negative symptoms in schizophrenia. Future study might want to examine this relationship in other psychiatric disorders such as depression or bipolar disorder to explore whether the relationship is generalizable to other psychiatric conditions.

6 Conclusion

In this paper we presented our findings regarding the correlations between the non-verbal speech cues and negative symptom ratings. Existing schizophrenia related research mostly focuses on semantic analysis and natural language processing. Little attention has been given to non-verbal speech features. In this paper we have highlighted the significance of these non-verbal speech features by showing their correlations with the negative symptoms for schizophrenia.

The results of our analysis are promising as there are significant correlations between non-verbal speech features and NSA-16 ratings. We also predicted NSA-16 criteria using machine learning algorithms trained on subjective ratings. The results show that some of the NSA-16 criteria can be classified as observable and non-observable using non-verbal speech features with quite high accuracy.

The positive findings from our analysis pave the way for identifying speech characteristics as markers for negative symptoms, developing a technological application that detects clinically significant speech patterns may assist clinicians in postulating functional level of an individual. With the gathered data from this study, we will work towards the development of a prototype, possibly as a mobile application to facilitate implementation.

Acknowledgments

The authors acknowledge the efforts of Madeline Lim Wei Zhen, Nicholas Lin Weikang and Quek Yue Feng for their help with data collection. We also thank the participants for their participation.

Author Contributions

Conceptualization: Yogeswary Maniam, Bhing-Leet Tan, Jimmy Lee Chee Keong, Justin Dauwels.

Data curation: Yasir Tahir, Zixu Yang, Debsubhra Chakraborty.

Formal analysis: Yasir Tahir, Debsubhra Chakraborty.

Funding acquisition: Nadia Thalmann, Daniel Thalmann, Bhing-Leet Tan, Jimmy Lee Chee Keong, Justin Dauwels.

Investigation: Yasir Tahir, Zixu Yang, Nur Amirah binte Abdul Rashid.

Methodology: Yasir Tahir, Zixu Yang, Debsubhra Chakraborty, Nur Amirah binte Abdul Rashid, Jimmy Lee Chee Keong, Justin Dauwels.

Project administration: Jimmy Lee Chee Keong, Justin Dauwels.

Resources: Zixu Yang, Jimmy Lee Chee Keong, Justin Dauwels.

Supervision: Jimmy Lee Chee Keong, Justin Dauwels.

Validation: Yasir Tahir, Debsubhra Chakraborty.

Writing – original draft: Yasir Tahir, Zixu Yang, Debsubhra Chakraborty.

Writing – review & editing: Yasir Tahir, Zixu Yang, Debsubhra Chakraborty, Jimmy Lee Chee Keong, Justin Dauwels.

References

1. Andreasen NC, Olsen S. Negative v positive schizophrenia: definition and validation. *Archives of General Psychiatry*. 1982; 39(7):789–794. <https://doi.org/10.1001/archpsyc.1982.04290070020005> PMID: 7165478
2. Demily C, Franck N. Cognitive remediation: a promising tool for the treatment of schizophrenia. *Expert Review of Neurotherapeutics*. 2008; 8(7):1029–1036. <https://doi.org/10.1586/14737175.8.7.1029> PMID: 18590474
3. Brenner H, Kraemer S, Hermanutz M, Hodel B. Cognitive treatment in schizophrenia. In: *Schizophrenia*. Springer; 1990. p. 161–191.
4. Kirkpatrick B, Fenton WS, Carpenter WT, Marder SR. The NIMH-MATRICES consensus statement on negative symptoms. *Schizophrenia bulletin*. 2006; 32(2):214–219. <https://doi.org/10.1093/schbul/sbj053> PMID: 16481659
5. Buchanan RW, Javitt DC, Marder SR, Schooler NR, Gold JM, McMahon RP, et al. The Cognitive and Negative Symptoms in Schizophrenia Trial (CONSIST): the efficacy of glutamatergic agents for negative symptoms and cognitive impairments. *The American journal of psychiatry*. 2007; 164(10):1593–1602. <https://doi.org/10.1176/appi.ajp.2007.06081358> PMID: 17898352
6. Murphy BP, Chung YC, Park TW, McGorry PD. Pharmacological treatment of primary negative symptoms in schizophrenia: a systematic review. *Schizophrenia research*. 2006; 88(1):5–25. <https://doi.org/10.1016/j.schres.2006.07.002> PMID: 16930948
7. Savilla K, Kettler L, Galletly C. Relationships between cognitive deficits, symptoms and quality of life in schizophrenia. *Australian & New Zealand Journal of Psychiatry*. 2008; 42(6):496–504. <https://doi.org/10.1080/00048670802050512>
8. Lysaker PH, Lancaster RS, Nees MA, Davis LW. Attributional style and symptoms as predictors of social function in schizophrenia. *Journal of rehabilitation research and development*. 2004; 41(2):225. <https://doi.org/10.1682/JRRD.2004.02.0225> PMID: 15558376
9. Milev P, Ho BC, Arndt S, Andreasen NC. Predictive values of neurocognition and negative symptoms on functional outcome in schizophrenia: a longitudinal first-episode study with 7-year follow-up. *American Journal of Psychiatry*. 2005; 162(3):495–506. <https://doi.org/10.1176/appi.ajp.162.3.495> PMID: 15741466
10. Kane JM. Tools to assess negative symptoms in schizophrenia. *J Clin Psychiatry*. 2013; 74(6):e12. <https://doi.org/10.4088/JCP.12045tx2c> PMID: 23842020
11. Schuller B, Steidl S, Batliner A, Vinciarelli A, Scherer K, Ringeval F, et al. The INTERSPEECH 2013 computational paralinguistics challenge: social signals, conflict, emotion, autism. In: *Proceedings INTERSPEECH 2013, 14th Annual Conference of the International Speech Communication Association*, Lyon, France; 2013.
12. Schuller B, Steidl S, Batliner A, Hantke S, Höning F, Orozco-Arroyave JR, et al. The INTERSPEECH 2015 computational paralinguistics challenge: nativeness, parkinson's & eating condition. In: *Sixteenth Annual Conference of the International Speech Communication Association*; 2015.
13. Valstar M, Schuller B, Smith K, Eyben F, Jiang B, Bilakhia S, et al. AVEC 2013: the continuous audio/visual emotion and depression recognition challenge. In: *Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge*. ACM; 2013. p. 3–10.
14. Valstar M, Schuller B, Smith K, Almaev T, Eyben F, Krajewski J, et al. AVEC 2014: 3d dimensional affect and depression recognition challenge. In: *Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge*. ACM; 2014. p. 3–10.
15. Cummins N, Epps J, Sethu V, Breakspear M, Goecke R. Modeling spectral variability for the classification of depressed speech. In: *Interspeech*; 2013. p. 857–861.
16. Scherer S, Stratou G, Gratch J, Morency LP. Investigating voice quality as a speaker-independent indicator of depression and PTSD. In: *Interspeech*; 2013. p. 847–851.
17. Sturim D, Torres-Carrasquillo PA, Quatieri TF, Malyska N, McCree A. Automatic detection of depression in speech using gaussian mixture modeling with factor analysis. In: *Twelfth Annual Conference of the International Speech Communication Association*; 2011.
18. Alpert M, Shaw RJ, Pouget ER, Lim KO. A comparison of clinical ratings with vocal acoustic measures of flat affect and alogia. *Journal of Psychiatric Research*. 2002; 36(5):347–353. [https://doi.org/10.1016/S0022-3956\(02\)00016-X](https://doi.org/10.1016/S0022-3956(02)00016-X) PMID: 12127603

19. Elvevåg B, Weinstock D, Akil M, Kleinman J, Goldberg T. A comparison of verbal fluency tasks in schizophrenic patients and normal controls. *Schizophrenia Research*. 2001; 51(2):119–126. PMID: [11518632](#)
20. Elvevåg B, Weickert T, Wechsler M, Coppola R, Weinberger D, Goldberg T. An investigation of the integrity of semantic boundaries in schizophrenia. *Schizophrenia Research*. 2002; 53(3):187–198. [https://doi.org/10.1016/S0920-9964\(01\)00202-X](https://doi.org/10.1016/S0920-9964(01)00202-X) PMID: [11738532](#)
21. Elvevåg B, Helsen K, De Hert M, Sweers K, Storms G. Metaphor interpretation and use: a window into semantics in schizophrenia. *Schizophrenia research*. 2011; 133(1):205–211. <https://doi.org/10.1016/j.schres.2011.07.009> PMID: [21821395](#)
22. McGilloway S, Cooper SJ, Douglas-Cowie E. Can patients with chronic schizophrenia express emotion? A speech analysis. *Schizophrenia research*. 2003; 64(2):189–190. [https://doi.org/10.1016/S0920-9964\(03\)00012-4](https://doi.org/10.1016/S0920-9964(03)00012-4) PMID: [14613684](#)
23. Elvevåg B, Foltz PW, Rosenstein M, DeLisi LE. An automated method to analyze language use in patients with schizophrenia and their first-degree relatives. *Journal of neurolinguistics*. 2010; 23(3):270–284. <https://doi.org/10.1016/j.jneuroling.2009.05.002> PMID: [20383310](#)
24. Holshausen K, Harvey PD, Elvevåg B, Foltz PW, Bowie CR. Latent semantic variables are associated with formal thought disorder and adaptive behavior in older inpatients with schizophrenia. *Cortex*. 2014; 55:88–96. <https://doi.org/10.1016/j.cortex.2013.02.006> PMID: [23510635](#)
25. Bedi G, Carrillo F, Cecchi GA, Slezak DF, Sigman M, Mota NB, et al. Automated analysis of free speech predicts psychosis onset in high-risk youths. *npj Schizophrenia*. 2015; 1. <https://doi.org/10.1038/npjpsz.2015.30> PMID: [27336038](#)
26. Troisi A, Spalletta G, Pasini A. Non-verbal behaviour deficits in schizophrenia: an ethological study of drug-free patients. *Acta Psychiatrica Scandinavica*. 1998; 97(2):109–115. <https://doi.org/10.1111/j.1600-0447.1998.tb09971.x> PMID: [9517903](#)
27. Cohen AS, Alpert M, Nienow TM, Dinzeo TJ, Docherty NM. Computerized measurement of negative symptoms in schizophrenia. *Journal of psychiatric research*. 2008; 42(10):827–836. <https://doi.org/10.1016/j.jpsychires.2007.08.008> PMID: [17920078](#)
28. Alpert M, Pouget ER, Welkowitz J, Cohen J. Mapping schizophrenic negative symptoms onto measures of the patient's speech: Set correlational analysis. *Psychiatry Research*. 1993; 48(3):181–190. [https://doi.org/10.1016/0165-1781\(93\)90070-W](https://doi.org/10.1016/0165-1781(93)90070-W) PMID: [8272441](#)
29. Cohen AS, Najolia GM, Kim Y, Dinzeo TJ. On the boundaries of blunt affect/alogia across severe mental illness: implications for Research Domain Criteria. *Schizophrenia research*. 2012; 140(1-3):41–45. <https://doi.org/10.1016/j.schres.2012.07.001> PMID: [22831770](#)
30. Covington MA, Lunden SA, Cristofaro SL, Wan CR, Bailey CT, Broussard B, et al. Phonetic measures of reduced tongue movement correlate with negative symptom severity in hospitalized patients with first-episode schizophrenia-spectrum disorders. *Schizophrenia research*. 2012; 142(1-3):93–95. <https://doi.org/10.1016/j.schres.2012.10.005> PMID: [23102940](#)
31. Alpert M, Rosenberg SD, Pouget ER, Shaw RJ. Prosody and lexical accuracy in flat affect schizophrenia. *Psychiatry research*. 2000; 97(2-3):107–118. [https://doi.org/10.1016/S0165-1781\(00\)00231-6](https://doi.org/10.1016/S0165-1781(00)00231-6) PMID: [11166083](#)
32. Martínez-Sánchez F, Muela-Martínez JA, Cortés-Soto P, Meilán JGG, Ferrándiz JAV, Caparrós AE, et al. Can the acoustic analysis of expressive prosody discriminate schizophrenia? *The Spanish journal of psychology*. 2015; 18. <https://doi.org/10.1017/sjp.2015.85> PMID: [26522128](#)
33. Knapp ML, Hall JA, Horgan TG. *Nonverbal communication in human interaction*. Cengage Learning; 2013.
34. Tahir Y, Chakraborty D, Dauwels J, Thalmann N, Thalmann D, Lee J. Non-verbal speech analysis of interviews with schizophrenic patients. In: *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on*. IEEE; 2016. p. 5810–5814.
35. Andreasen NC. Negative symptoms in schizophrenia: definition and reliability. *Archives of general psychiatry*. 1982; 39(7):784–788. <https://doi.org/10.1001/archpsyc.1982.04290070020005> PMID: [7165477](#)
36. Daniel D, Alphs L, Cazorla P, Bartko J, Panagides J. Training for assessment of negative symptoms of schizophrenia across languages and cultures: comparison of the NSA-16 with the PANSS Negative Subscale and Negative Symptom factor. *Clinical schizophrenia & related psychoses*. 2011; 5(2):87–94.
37. Rasheed U, Tahir Y, Dauwels S, Dauwels J, Thalmann D, Magnenat-Thalmann N. Real-time comprehensive sociometrics for two-person dialogs. In: *International Workshop on Human Behavior Understanding*. Springer; 2013. p. 196–208.
38. Tahir Y, Chakraborty D, Maszczyk T, Dauwels S, Dauwels J, Thalmann N, et al. Real-time sociometrics from audio-visual features for two-person dialogs. In: *Digital Signal Processing (DSP), 2015 IEEE International Conference on*. IEEE; 2015. p. 823–827.

39. Hall MA. Correlation-based feature selection for machine learning. University of Waikato Hamilton; 1999.
40. Benesty J, Chen J, Huang Y, Cohen I. Pearson correlation coefficient. In: Noise reduction in speech processing. Springer; 2009. p. 1–4.
41. Witten IH, Frank E, Hall MA, Pal CJ. Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann; 2016.