•Original research article•

Clinical investigation of speech signal features among patients with schizophrenia

Jing ZHANG[1,2], Zhongde PAN[1,3,4], Chao GUI[5], Jie ZHU[5,*], Donghong CUI[1,3,4,*]

Background: A new area of interest in the search for biomarkers for schizophrenia is the study of the acoustic parameters of speech called 'speech signal features'. Several of these features have been shown to be related to emotional responsiveness, a characteristic that is notably restricted in patients with schizophrenia, particularly those with prominent negative symptoms.

Aim: Assess the relationship of selected acoustic parameters of speech to the severity of clinical symptoms in patients with chronic schizophrenia and compare these characteristics between patients and matched healthy controls.

Methods: Ten speech signal features – six prosody features, formant bandwidth and amplitude, and two spectral features – were assessed using 15-minute speech samples obtained by smartphone from 26 inpatients with chronic schizophrenia (at enrollment and 1 week later) and from 30 healthy controls (at enrollment only). Clinical symptoms of the patients were also assessed at baseline and 1 week later using the Positive and Negative Syndrome Scale, the Scale for the Assessment of Negative Symptoms, and the Clinical Global Impression-Schizophrenia scale.

Results: In the patient group the symptoms were stable over the 1-week interval and the 1-week test-retest reliability of the 10 speech features was good (intraclass correlation coefficients [ICC] ranging from 0.55 to 0.88). Comparison of the speech features between patients and controls found no significant differences in the six prosody features or in the formant bandwidth and amplitude features, but the two spectral features were different: the Mel-frequency cepstral coefficient (MFCC) scores were significantly lower in the patient group than in the control group, and the linear prediction coding (LPC) scores were significantly higher in the patient group than in the control group. Within the patient group, 10 of the 170 associations between the 10 speech features considered and the 17 clinical parameters considered were statistically significant at the $p<0.05$ level.

Conclusions: This study provides some support for the potential value of speech signal features as indicators (i.e., biomarkers) of the severity of negative symptoms in schizophrenia, but more detailed studies using larger samples of more diverse patients that are followed over time will be needed before the potential utility of such acoustic parameters of speech can be fully assessed.

Keywords: schizophrenia; speech; speech signal features; biomarkers; negative symptoms; China

[*Shanghai Arch Psychiatry*. 2016, **28**(2): 95-102. doi: http://dx.doi.org/10.11919/j.issn.1002-0829.216025]

[1] Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, Shanghai, China
[2] Jiading District Mental Health Center, Shanghai, China
[3] Shanghai Key Laboratory of Psychotic Disorders, Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, Shanghai, China
[4] Key Laboratory of Translational Psychiatry, Shanghai Jiao Tong University School of Medicine, Shanghai, China
[5] School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China

*co-corresponding authors: Dr. Donghong Cui, Shanghai Key Laboratory of Psychotic Disorders, Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, 3210 Humin Road, Shanghai 201108, China. E-mail: manyucc@126.com; Jie Zhu, School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai, China. E-mail: zhujie@sjtu.edu.cn

A full-text Chinese translation of this article will be available at http://dx.doi.org/10.11919/j.issn.1002-0829.216025 on August 25, 2016.

1. Introduction

Schizophrenia is a complex mental disorder caused by multiple factors including heredity, development, and environment.[1] The fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5)[2] lists the following five prominent psychopathological characteristics of the disorder: delusions, hallucinations, disorganized speech, grossly disorganized or catatonic behavior, and negative symptoms. Negative symptoms include emotional blunting, poverty of speech, avolition, an inability to experience pleasure, and the lack of desire to form relationships. Present methods for determining the diagnosis and for assessing the effectiveness of treatment primarily rely on the subjective judgment of the clinician who uses information provided by family members, the mental status examination, and various clinical symptom scales. In the absence of objective measures and the frequent uncooperativeness of patients – particularly those with prominent negative symptoms – assessing the severity and course of the illness is often challenging for clinicians. To address this fundamental problem, psychiatric researchers are actively searching for objective biomarkers that can be used both in the diagnosis of the condition and in the monitoring of the clinical progress of the disorder.

Fluctuations in speech that parallel patients' physio-psychological state might be suitable candidates as biomarkers for schizophrenia. Studies of signal processing and artificial intelligence find that the features of speech signals can contain substantial emotional information.[3,4] Changes of emotions and the range and variability of emotions can be quantified by changes in speech parameters, particularly by changes in prosody – that is, the vocal pitch (fundamental frequency), loudness (acoustic intensity), and rhythm (phoneme and syllable duration) of speech. For example, when a person is in an angry state, changes in physiological characteristics (e.g., increased heart rate, elevated skin voltage, and elevated blood pressure) are often associated with changes in the rate, volume, and tone of speech.

There is considerable interest in developing methods for extracting the acoustic parameters which reflect emotions from speech samples and in assessing the relationship of these parameters to emotionally restrictive states, such as the negative symptoms of schizophrenia. Identification of the emotional content of speech signals is primarily accomplished by two processes: first, the features of the speech signals are extracted from speech samples and then judgments are made about the emotional content of the identified features based on pre-existing models. The quality of the extraction process largely determines the functional quality of the speech identification system.[5,6] Studies about speech identification generally start by investigating the prosodic features and acoustic characteristics of speech content, focusing on the features which are directly relevant to the emotional characteristics of speech.[7]

Patients with schizophrenia who have prominent negative symptoms such as emotional blunting and poverty of speech may be particularly prone to having restricted emotional content in their speech content. This can directly limit their social functioning and make it difficult for clinicians to detect changes in their clinical status over the course of their illness. Identification of the specific speech abnormalities in such patients could both help in monitoring the course of the illness and potentially be used to develop targeted interventions for patients with prominent negative symptoms. Several researchers[8-11] have reported relationships between specific phonetic parameters and the negative symptoms and impaired emotional perception of schizophrenia, but the results to date are far from robust.

The current study assessed the characteristics of the speech signals of patients with schizophrenia with prominent negative symptoms, considered the association between these features and the severity of different types of negative symptoms, and compared the speech signal features in these patients with those in healthy control subjects.

2. Methods

2.1 Participants

The patient group consisted of patients with schizophrenia who were inpatients at the Shanghai Mental Health Center from September 2013 to December 2015. The inclusion criteria were as follows: (a) aged from 18 to 65 years; (b) met the diagnostic criteria of schizophrenia specified in DSM-5[2] as assessed by a psychiatrist using the Mini-International Neuropsychiatric Interview (M.I.N.I. 6.0).[12] (c) had prominent negative symptoms of schizophrenia; (d) a minimum of duration of illness of two years; (e) no co-morbid psychiatric or substance abuse disorder; (f) no evidence of severe impulsivity; (g) not using antipsychotic medication that could impair speech; and (h) both the patient and the patient's family member provided written informed consent to participate in the study, including the use of smartphones to record speech.

We recruited volunteers from the community by advertisement as healthy controls. Volunteers were similar for patients in age and duration of education, underwent a through psychiatric exam (using the M.I.N.I. 6.0) and physical exam. Inclusion criteria for controls were as follows (a) 18 to 65 years of age; (b) Han Chinese ethnicity; (c) no current or past physio-psychological, substance abuse, or serious neurological disorder; (d) no serious physical illness; (e) no history of severe impulsivity; (f) no history of suicide attempt; (g) not using antipsychotic medication that could affect speech; (h) no family history of psychiatric disorder or serious neurological disorder; and (i) provided written informed consent to participate in the study.

2.2 Measures

We explored a smartphone APP which could record the participant's outgoing speech (i.e., no incoming speech is captured or recorded). Each participant was provided with a preloaded Samsung GALAXY Mega 6.3 (a sampling frequency of 44 kHz and a resolution of 32 bit). Participants (both patients and controls) sat comfortably in a noise-controlled room (background sound below 30 dB) and were asked to use the specially designed smartphone to call a psychiatrist from the study and speak naturally for 15 minutes about any topic of interest.

All call samples were saved in the Advanced Audio Coding (aac) formant. After pre-processing of the data, speech features of interest were extracted and analyzed at the School of Electronic Information and Electrical Engineering of Shanghai Jiao Tong University. We extracted speech features with high emotional identification rates that were identified in previous reports of speech models of bipolar disorder[13,14] and from our own speech model for schizophrenia generated from data in the current study. These parameters include prosodic features (formant 1 to 6 [F1 to F6, unit: Hz], formant bandwidth [unit: Hz], formant amplitude [unit: dB]), and two spectral features (the linear prediction coding [LPC], and the Mel-frequency cepstral coefficient [MFCC]). To assess the stability of the speech data extracted by the software, participants in the patient group repeated the phone call 1 week after the baseline assessment.

In addition, trained attending psychiatrists administered the Positive and Negative Syndrome Scale (PANSS)[15], the Scale for the Assessment of Negative Symptoms (SANS),[16] and the Clinical Global Impression-Schizophrenia Scale (CGI-S)[17] to the participating patients at baseline and 1 week after baseline.

2.3 Statistical analysis

All data were processed and analyzed using SPSS 17.0 software. The in-group comparisons from baseline to 1-week post-baseline were analyzed by paired t-tests. The test-retest reliability of the acoustic parameters was assessed by comparing the baseline and 1-week results using intraclass correlation coefficients (ICCs). For between-group comparisons, continuous data with normal distributions were analyzed using independent t-tests; non-normal continuous variables were analyzed using Mann-Whitney U tests; and nominal data were analyzed using Chi-square tests. In the patient group, we use Pearson correlation analyses to assess the strength of the relationship between the acoustic parameters and the severity of clinical symptoms. All statistical analyses used two-tailed tests and statistical significance was set at $p<0.05$.

3. Results

As shown in Figure 1, 26 patients completed the two assessments. These patients included 16 males (61.5%); their mean (sd) age was 43.3 (10.9) years; their mean years of education was 9.5 (3.0) years; the mean course of their illness was 21.7 (8.5) years; the mean number of psychiatric admissions was 3.4 (2.4) admissions; and the mean total length of hospitalization was 7.0 (5.4) years. A total of 30 control subjects completed the phonetic assessment; they included 16 males (53.3%), had a mean (sd) age of 37.0 (14.3) years and had a mean duration of education of 11.6 (2.5) years. Comparison between the 26 patients who completed the assessment and the 30 controls who completed the assessment found no statistically significant differences by gender (X^2=0.38, p=0.536), by age (t=1.70, p=0.098), or by duration of education (t=1.95, p=0.058).

As shown in Table 1, in the patient group there were no statistically significant differences between the baseline and 1-week assessment of CGI-S, PANSS total and subscale scores, and SANS total and subscale scores. Thus the patients' clinical status was stable over the 1-week interval.

Table 2 shows the baseline and 1-week results for the acoustic parameters in the patient group and the baseline acoustic parameters in the control group. The test-retest reliability of these measures (only assessed in the patient group) was good, with ICC values ranging from 0.55 to 0.88. The prosody features and formant amplitude and bandwidth were not significantly different between patients and controls at baseline, but the two spectral features were different between the groups: MFCC was significantly lower in the patient group than in the control group and the LPC was significantly higher in the patient group than in the control group.

Table 3 shows the correlation of 17 clinical and demographic measures with the 10 acoustic parameters in the 26 patients. Among the 170 associations considered, ten coefficients were $\geq$0.40 and, thus, statistically significant at the $p<0.05$ level: Formant 1 was negatively correlated with the SANS alogia subscale score, Format 2 was negatively correlated with the PANSS negative symptoms subscale score and the SANS alogia subscale score; Formant 6 was significantly more prominent in male than female respondents; bandwidth was negatively correlated with the SANS affective blunting subscale score and stronger in female respondents than in male respondents; and MFCC was positively correlated with the PANSS general psychopathology subscale score and with the patients' total number of hospitalizations, and it was more prominent in male respondents than female respondents.

4. Discussion

4.1 Main findings

We found that when the severity of psychiatric symptoms remains stable, the speech features selected to assess the emotional content of the voice samples of patients with schizophrenia with prominent negative

Figure 1. Flowchart of the study

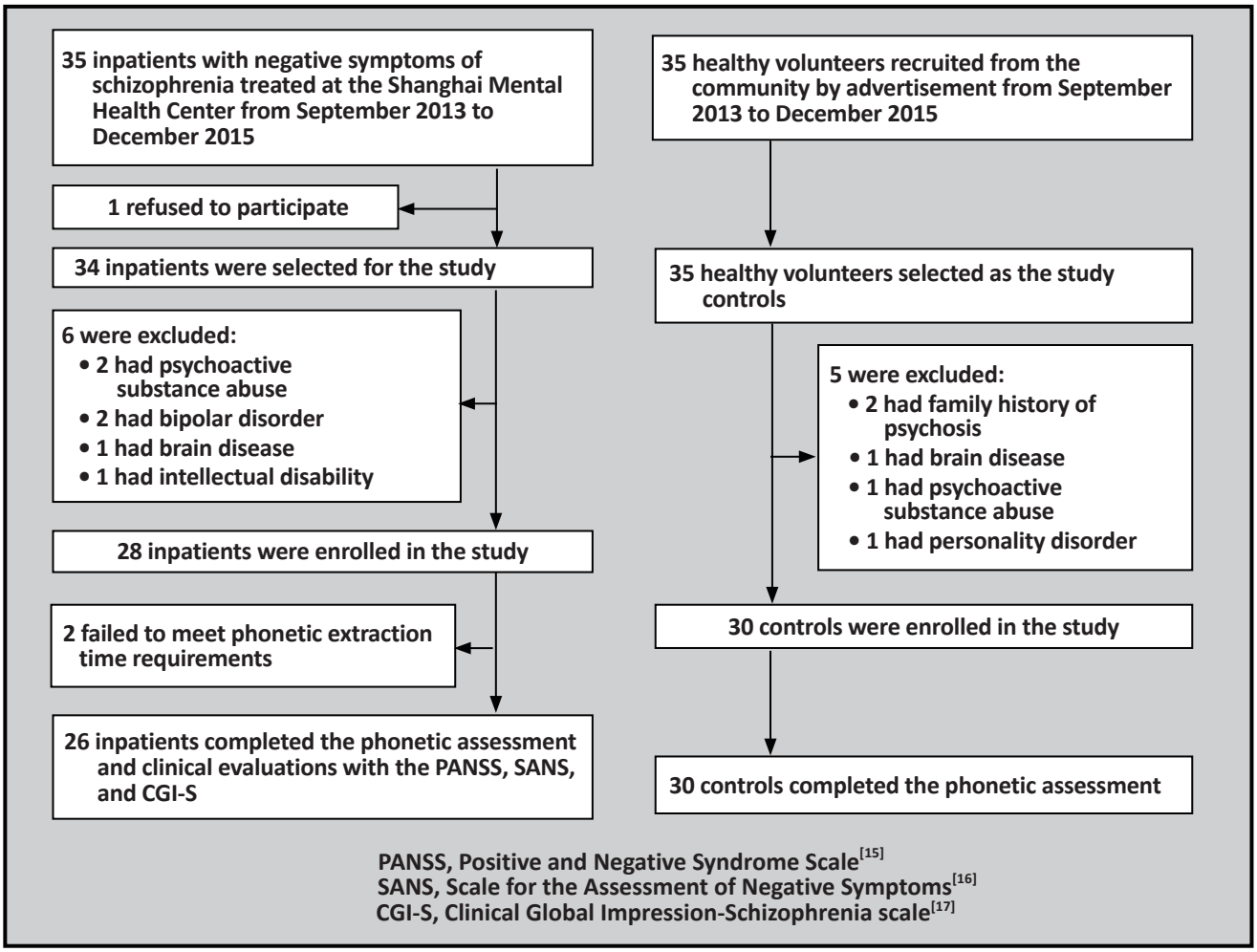

Table 1. Comparisons of clinical symptoms in 26 patients with schizophrenia at baseline and after 1 week

	baseline mean (sd)	after 1 week mean (sd)	paired t-test	p-value
CGI-S	4.69 (0.74)	4.73 (0.78)	1.00	0.327
PANSS	---	---	---	---
total score	69.23 (9.62)	68.96 (9.64)	1.16	0.258
positive symptoms score	8.15 (1.54)	8.23 (1.53)	1.44	0.161
negative symptoms score	28.23 (4.03)	27.81 (4.24)	2.03	0.054
general psychopathology score	32.88 (5.83)	32.96 (5.86)	1.00	0.327
SANS	---	---	---	---
total score	76.00 (7.43)	74.92 (9.83)	0.87	0.391
affective blunting score	26.27 (3.77)	24.38 (4.04)	1.94	0.064
alogia score	14.15 (1.83)	14.04 (2.36)	0.37	0.713
avolition score	15.92 (2.11)	16.42 (2.63)	1.64	0.114
anhedonia score	19.27 (3.62)	19.73 (3.53)	1.95	0.063
attention score	0.42 (1.03)	0.38 (0.85)	1.00	0.327

PANSS, Positive and Negative Syndrome Scale[15]
SANS, Scale for the Assessment of Negative Symptoms[16]
CGI-S, Clinical Global Impression-Schizophrenia scale[17]

Table 2. Comparisons of speech features at baseline and after 1 week in the patient group and between patient and control groups at baseline

phonetic parameter	baseline patient group result (n=26) mean (sd)	patient group result after 1 week (n=26) mean (sd)	test-retest reliability of patient results ICC (p-value)	baseline control group result (n=30) mean (sd)	comparison of baseline patient v. control group results *t* (p-value)
F1 (Hz/dB)	0.036 (0.007)	0.237 (0.028)	**0.76 (0.002)**	0.040 (0.005)	1.78 (0.081)
F2 (Hz/dB)	0.040 (0.052)	0.084 (0.009)	**0.68 (0.009)**	0.083 (0.009)	0.95 (0.353)
F3 (Hz/dB)	0.174 (0.015)	0.179 (0.014)	**0.84 (<0.001)**	0.182 (0.012)	1.43 (0.153)
F4 (Hz/dB)	0.265 (0.023)	0.264 (0.019)	**0.67 (0.011)**	0.257 (0.019)	0.35 (0.725)
F5 (Hz/dB)	0.359 (0.020)	0.359 (0.014)	**0.88 (<0.001**)	0.357 (0.013)	0.21 (0.836)
F6 (Hz/dB)	0.426 (0.011)	0.426 (0.012)	**0.55 (0.045)**	0.426 (0.015)	0.06 (0.951)
formant bandwidth (Hz)	18.83 (11.05)	21.26 (12.41)	**0.63 (0.019)**	18.10 (9.81)	1.16[a] (0.247)
formant amplitude (dB)	0.043 (0.005)	0.042 (0.004)	**0.61 (0.024)**	0.041 (0.009)	1.01 (0.317)
MFCC	-0.085 (0.500)	-0.027 (0.193)	**0.87 (<0.001)**	0.236 (0.043)	**4.97 (<0.001)**
LPC	0.249 (0.067)	0.035 (0.006)	**0.72 (0.004)**	-0.203 (0.367)	**5.69 (<0.001)**

ICC, Intraclass Correlation Coefficient MFCC, Mel-frequency cepstral coefficient LPC, linear prediction coding

[a] The homogeneity of variances tests showed that the data were heterogeneous, so comparison of the results of the two groups used the Mann Whitney U test; this is the Z-value of the Mann-Whitney U

Table 3. Correlation of demographic characteristics and the severity of clinical symptoms (at baseline) with the speech features in 26 patients with schizophrenia (Pearson's *r*)

	F1	F2	F3	F4	F5	F6	bandwidth	amplitude	MFCC	LPC
CGI-S	-0.28	-0.20	0.04	-0.10	0.03	-0.13	-0.13	0.19	0.04	-0.03
PANSS										
total score	-0.30	-0.38	-0.16	-0.26	-0.003	0.15	-0.14	0.08	0.31	0.09
positive symptoms score	0.11	0.04	0.09	0.18	0.23	-0.12	-0.23	-0.25	-0.03	-0.36
negative symptoms score	-0.23	**-0.40**[a]	-0.21	-0.30	0.05	0.10	-0.01	0.17	0.13	0.26
general psychopathology score	-0.36	-0.36	-0.15	-0.28	-0.11	0.21	-0.16	0.08	**0.43**[a]	0.06
SANS										
total score	-0.09	0.07	-0.04	-0.18	0.04	0.02	-0.20	0.002	0.06	0.02
affective blunting score	-0.04	0.28	-0.30	**-0.50**[a]	0.07	0.31	**-0.41**[a]	-0.001	0.15	-0.12
alogia score	**-0.42**[a]	**-0.42**[a]	0.17	-0.01	0.10	0.23	0.06	0.26	0.20	-0.11
avolition score	-0.04	0.12	0.05	0.08	-0.02	-0.17	-0.04	0.08	0.02	0.09
anhedonia score	0.09	0.10	0.09	0.11	0.02	-0.24	-0.06	-0.18	-0.03	0.12
attention score	-0.04	-0.34	0.22	0.08	-0.22	-0.23	-0.23	0.06	-0.24	0.18
Age	-0.05	0.22	0.28	0.05	0.17	-0.27	-0.09	-0.18	-0.04	-0.05
Gender (1=female, 2=male)	-0.18	0.18	-0.08	0.03	-0.19	**0.45**[a]	**-0.50**[b]	0.20	**0.69**[b]	-0.08
Years of education	-0.18	-0.22	-0.10	-0.003	0.179	-0.25	0.17	0.11	0.07	0.04
Duration of illness	0.07	0.30	0.16	-0.01	-0.04	-0.25	-0.25	-0.21	-0.01	0.01
Number of hospitalizations	-0.23	-0.13	0.10	-0.16	-0.01	-0.06	-0.26	-0.27	**0.40**[a]	-0.06
Total time hospitalized	-0.04	0.10	0.10	0.18	-0.11	-0.08	-0.13	0.08	0.22	0.09

PANSS, Positive and Negative Syndrome Scale[15]
SANS, Scale for the Assessment of Negative Symptoms[16]
CGI-S, Clinical Global Impression-Schizophrenia scale[17]
MFCC, Mel-frequency cepstral coefficient
LPC, linear prediction coding

[a] $0.01<p<0.05$
[b] $0.001<p<0.01$

symptoms were also stable over a 1-week period. Correlation analyses of these measures with clinical and demographic characteristics of the patients identified several potentially important relationships, a finding that has been reported in previous studies.[8-10] Comparison of these speech features between patients and matched healthy controls found no statistically significant differences in the prosody features or formant bandwidth and amplitude, but there were significant differences in the two spectral features considered: the MFCC was significantly lower in patients than controls, while the LPC was significantly higher in patients than controls.

Other studies have reported that these two spectral features play an important role in everyday communications.[18] Spectral features have also been found to be useful for discriminating emotions in artificial intelligence studies. The LPC is a relatively efficient and accurate measure of the waveform and spectrum of speech that is used in speech coding, speech synthesis, speech identification, and other applications.[19,20] The MFCC, which modifies external signals in a manner similar to the human ear, is a reliable parameter for discriminating different emotional states.[21,22] Similar to our results, a study by Sun and colleagues[23] found that (when using a sorter based on a Gaussian mixture model [GMM] algorithm) the MFCC was better at discriminating different emotional states than the prosody features. Further work is needed to determine whether or not these spectral features can be used as biomarkers for the identification and monitoring of schizophrenia or of the prominent negative symptom subtype of schizophrenia.

The correlation analysis identified some intriguing associations between 6 of the 10 speech signal features considered (F1, F2, F4, F6, bandwidth, and MFCC) and 6 of the 17 clinical and demographic parameters considered (gender, the PANSS negative symptoms and general psychopathology subscale scores, the SANS affective blunting and alogia subscale scores, and the number of hospitalizations). Other studies have also identified significant correlations between different speech features and the negative symptoms of schizophrenia.[11] This raises the possibility that a subset of acoustic parameters of a standardized speech sample – potentially transmitted over a smart phone to clinicians – could be used to either monitor the severity of negative symptoms or predict the subsequent course of the illness. However, given the small sample size and the large number of potential associations considered in the current study, these results need to be replicated before they can be meaningfully interpreted.

4.2 Limitations

The present study has several limitations that need to be considered when interpreting the results. The speech features selected may not be the most sensitive measures of changes in chronic schizophrenia; further research using a wider range of measures will be needed to find other, potentially more sensitive, measures. The 1-week interval we used to assess the test-retest reliability of the speech features indicated short-term stability in the phonetic parameters, but we are uncertain how stable such measures are over a longer time period. A total of 170 correlations between the 10 speech features assessed and 17 clinical and demographic characteristics are considered, so the statistically significant relationships identified may be chance findings; repeat studies are needed to confirm their importance. The comparison between patients and controls was cross-sectional so we cannot report on the sensitivity of the speech features to changes in clinical symptoms; longitudinal studies that compare changes in the speech features to changes in the clinical measures will be needed to determine their potential utility as biomarkers of clinical changes. All the patients included in the study had a prolonged course of illness and had been on antipsychotic medication for many years, so it is possible that this may have had an effect on the assessed speech features.[15,16,18] Finally, the sample was quite small – only 26 patients – so some of the negative findings (e.g., failure to identify differences between patients and controls) may have been due to Type II errors.

4.3 Importance

This study focused on the negative symptoms of schizophrenia, symptoms that are often not improved with standard antipsychotic medications and that often predict a poor prognosis and progressive deterioration in social functioning.[15] The study is a preliminary assessment of the feasibility of using speech features that assess the emotional characteristics of speech as biomarkers for the severity of negative symptoms in schizophrenia and, thus, as potential predictors of the prognosis of the disorder. The selected speech features included both the prosodic variables used in prior studies (i.e., rate, volume, rhythm, etc.) and two spectral features (MFCC and LPC) that have previously been shown to be useful in the emotional characterization of speech samples. These speech features proved to be stable (over a short period), some of them – the spectral features rather than the prosody features reported in some previous studies – were significantly different between patients and controls, and some of them were significantly correlated with clinical measures of negative symptoms. However, this was a cross-sectional study in a small group of chronic patients, so much more detailed studies using larger samples of more diverse patients that are followed over time will be needed before the potential utility of such speech features can be fully assessed.

Funding

Shanghai public health outstanding academic leader training program (GWDTR201230); Shanghai Jiao Tong

University Key Program for the Medical Engineering Cross Project (YG2012ZD04);and Shanghai Key Laboratory of Psychotic Disorders (13dz2260500).

Conflict of interest statement

The authors report no conflict of interest.

Informed consent

Every participant who participated in this study signed a consent form at the beginning of the study.

Ethical review approval

The study has been approved by the Shanghai Mental Health Center Institutional Review Board (approval number: 2011-15).

Authors' contributions

DHC designed the study and revised the manuscript; J Zhang and ZP recruited patients and healthy controls, gathered speech data, and conducted clinical evaluations; J Zhu and CG conducted the phonetic analyses; J Zhang did the statistical analyses and wrote the first draft of the manuscript.

精神分裂症患者语音信号的临床分析

张静，潘忠德，桂超，朱杰，崔东红

背景：语音参数是精神分裂症生物学指标研究的一个全新领域，其中一些已被证明与情感反应相关，情感反应是精神分裂症患者显著受限制的一个特点，特别是对那些具有突出阴性症状的患者。

目标：评估慢性精神分裂症患者的选择性语音参数与临床症状严重程度之间的关系，并比较患者与所匹配的健康对照者的这些特征。

方法：对 26 例住院慢性精神分裂症患者（入组时和一周后）和 30 名健康对照者（仅在入组时）通过电话采集的 15 分钟语音样本，对该样本进行 10 项语音测量参数的评估，包括 6 个语音韵律参数、共振峰带宽和振幅、以及 2 个频谱特征。采用阳性与阴性症状量表 (Positive and Negative Syndrome Scale)、阴性症状评估量表 (Scale for the Assessment of Negative Symptoms)、临床总体印象量表 - 精神分裂症分量表 (the Clinical Global Impression-Schizophrenia scale) 分别在基线和 1 周后进行患者临床特征的评估。

结果：患者组症状在 1 周的时间间隔中保持稳定，并且 10 项语音参数的前后一周重测信度良好（内部相关系数 [intraclass correlation coefficient, ICC] 介于 0.55 到 0.88 之间）。语音参数中 6 项韵律参数、共振峰带宽和振幅参数在患者组和对照组之间没有显著差异，但 2 项光谱参数在组间有差异：患者组美尔频率倒谱系数 (the Mel-frequency cepstral coefficient, MFCC) 评分显著低于对照组，并且患者组的线性预测系数 (linear prediction coding, LPC) 评分显著高于对照组。在患者组中，在 10 个本研究所考虑的语音参数和 17 个所考虑的临床参数之间构成的 170 个相关性中，有 10 个达到了 $p<0.05$ 的统计学显着性水平（相关系数 >0.40）。

结论：这项研究支持了语音参数具有作为精神分裂症阴性症状严重程度指标（即，生物指标）的潜在价值，但在这些语音参数的潜在效用获充分评估前，我们需要对更多样化的患者进行更大样本量、更详细的随访研究。

关键词：精神分裂症；语音；生物标志；阴性症状；中国

本文全文中文版从 2016 年 8 月 25 日起在 http://dx.doi.org/10.11919/j.issn.1002-0829.216025 可供免费阅览下载

References

1. Green MF, Bearden CE, Cannon TD, Alan PF, Hellemann GS, Horan WP, et al. Social cognition in schizophrenia, part 1: performance across phase of illness. *Schizophr Bull*. 2012; **38**(4): 854-864. doi: http://dx.doi.org/10.1093/schbul/sbq171
2. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5).* Arlington, VA: American Psychiatric Association; 2013
3. Zhao L, Jiang CH, Zou CR, Wu ZY. [A study on emotional feature analysis and recognition in speech signal]. *Dian Zi Xue Bao*. 2004; **32**(4): 606-608. Chinese. doi: http://dx.chinadoi.cn/10.3321/j.issn:1000-436X.2000.10.004
4. Nwe TL, Foo SW, Silva LCD. Speech emotion recognition using hidden Markov models. *Speech Communication*. 2003; **41**(3): 603-623. doi: http://dx.doi.org/10.1016/S0167-6393(03)00099-2
5. Lin YL, Wei G, Yang KC. [A survey of emotion recognition in speech]. *Guangzhou Dian Lu Yu Xi Tong Xue Bao*. 2007; **12**(1): 90-91. Chinese. doi: http://dx.chinadoi.cn/10.3969/j.issn.1007-0249.2007.01.019
6. Yuan J, Xu HH, He X. [Research progress of speech emotion recognition]. *Ji Suan Ji Guang Pan Ruan Jian Yu Ying Yong*. 2010; **1**: 36-38. Chinese

7. Bhatti MW, Wang Y, Guan L. A neural network approach for human emotion recognition in speech (conference paper). Sydney: Circuits and Systems Conference, ISCAS '04; 2004. p.181-184. doi: http://dx.doi.org/10.1109/ISCAS.2004.1329238
8. Stassen HH, Albers M, Püschel J, Scharfetter C, Tewesmeier M, Woggon B. Speaking behavior and voice sound characteristics associated with negative schizophrenia. *J Psychiatr Res*. 1995; **29**(4): 277-296. doi: http://dx.doi.org/10.1016/0022-3956(95)00004-O
9. Püschel J, Stassen HH, Bomben G, Scharfetter C, Hell D. Speaking behavior and speech sound characteristics in acute schizophrenia. *J Psychiatr Res*. 1998; **32**(2): 89-97. doi: http://dx.doi.org/10.1016/S0022-3956(98)00046-6
10. Leitman DI, Laukka P, Juslin PN, Saccente E, Butler P, Javitt DC. Getting the cue: sensory contributions to auditory emotion recognition impairments in schizophrenia. *Schizophr Bull*. 2008; **36**(3): 545-556. doi: http://dx.doi.org/10.1093/schbul/sbn115
11. Gold R, Butler P, Revheim N, Leitman DI, Hansen JA, Gur RC, et al. Auditory emotion recognition impairments in schizophrenia: relationship to acoustic features and cognition. *Am J Psychiatry*. 2012; **169**(4): 424-432. doi: http://dx.doi.org/10.1176/appi.ajp.2011.11081230
12. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry*. 1998; **59** (Suppl 20): 22-33
13. Xu D, Zhang J, Zhu J, Cui D. *Acoustic Analysis and Identification of Manic Psychosis Patients*. London: Intelligent Signal Processing Conference; 2013
14. Gui C, Li W, Pan Z, Zhang J, Zhu J, Cui D. *A classifier for diagnosis of manic psychosis state based on SVM-GMM.* Sydney: The 10th International Conference on Information Technology and Applications (ICITA2015); 2015
15. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr Bull.* 1987; **13**(2): 261-276. PMID 3616518. doi: http://dx.doi.org/10.1093/schbul/13.2.261.
16. Andreasen NC. Negative symptoms in schizophrenia: definition and reliability. *Arch Gen Psychiatry*. 1982; **39**(7): 784-788
17. Haro J, Kamath S, Ochoa S, Novick D, Rele K, Fargas A, et al. The Clinical Global Impression-Schizophrenia scale: a simple instrument to measure the diversity of symptoms present in schizophrenia. *Acta Psychiatrica Scandinavica.* 2003; **107**(Suppl 416): 16-23. doi: http://dx.doi.org/10.1034/j.1600-0447.107.s416.5.x
18. Sun Y, Jiang ZC, Wang DF. [Analysis and application of voice spectrum]. *Ji Suan Ji Yu Xian Dai Hua*. 2010; **4**: 200-202. Chinese. doi: http://dx.chinadoi.cn/10.3969/j.issn.1006-2475.2010.04.054
19. Yu BK, Yu M. [Analysis of the resonance peak of speech signal extracted with LPC method]. *Dian Sheng Ji Shu*. 2000; **3**: 3-8. Chinese. doi: http://dx.chinadoi.cn/10.3969/j.issn.1002-8684.2000.03.001
20. Nica A, Caruntu A, Toderean G, Buza O. *Analysis and synthesis of vowels using Matlab*. IEEE Conference on Automation, Quality and Testing, Robotics; 2006. p. 371-374. doi: http://dx.doi.org/10.1109/AQTR.2006.254662
21. Lin W, Yang LL, Xu BL. [Speaker identification in Chinese whispered speech based on modified-MFCC]. *Nanjing Da Xue Xue Bao (Zi Ran Ke Xue)*. 2006; **42**(1): 54-61. Chinese. doi: http://dx.chinadoi.cn/10.3321/j.issn:0469-5097.2006.01.008
22. Dave N. Feature extraction methods LPC, PLP and MFCC in speech recognition. *Ijaret Org*. 2013; **1**(6): 1-5
23. Sun MH, Jiang BC. [*Analysis and Research on the Emotional Information of Mandarin Speech*]. Shandong: Shandong University. 2011; p. 22-24. Chinese

(received, 2016-02-19; accepted, 2016-03-22)

Dr. Jing Zhang graduated from Wannan Medical College with a bachelor's degree in medicine in 2007. She is now a master's student at the Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine. She has been working in the Department of Psychiatry as an attending physician in the Jiading District Mental Health Center in Shanghai since 2009. Her main research interest is the diagnostic relevance of speech signals in patients with schizophrenia.