ORIGINAL RESEARCH

Associations of Vocal Features, Psychiatric Symptoms, and Cognitive Functions in Schizophrenia

Tomoya Natsuyama[1], Enkhmurun Chibaatar[1], Yuko Shibata[2], Naomichi Okamoto [ID][1], John Noel Cruz Victorino [ID][2], Atsuko Ikenouchi [ID][1], Tomohiro Shibata[2], Reiji Yoshimura [ID][1]

[1]Department of Psychiatry, University of Occupational and Environmental Health, Kitakyushu, Fukuoka, 807-8555, Japan; [2]Department of Life Science and System Engineering, Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, Kitakyushu, Fukuoka, 808-0135, Japan

Correspondence: Reiji Yoshimura, Department of Psychiatry, University of Occupational and Environmental Health, 1-1, Iseigaoka, Yahatanishi, Kitakyushu, Fukuoka, 807-8555, Japan, Tel +81-93-691-7253, Email yoshi621@med.uoeh-u.ac.jp

Purpose: This study explored the use of advanced computational techniques in vocal analysis to improve the assessment of psychiatric symptoms and cognitive functions in schizophrenia. We hypothesized that digital signal processing techniques, such as mel spectrogram and mel-frequency cepstral coefficients (MFCC), could be used for objective evaluation of psychiatric symptoms and cognitive functions based on the analysis of alterations in the vocal characteristics.

Patients and Methods: Voice samples from 14 participants diagnosed with schizophrenia (92.9% female) were collected using a microphone array, and vocal features were extracted from the samples using mel spectrogram and MFCC techniques. Psychiatric symptoms and cognitive functions were assessed using the Positive and Negative Syndrome Scale (PANSS) and the computer-based tool Cognitrax.

Results: We found significant negative correlations between specific vocal features (mel spectrogram and MFCC) and cognitive functions, particularly working memory ($\beta = -0.645$, $p = 0.023$) and sustained attention ($\beta = -0.626$, $p = 0.029$). No direct correlations were found between vocal features and psychiatric symptoms, as measured by PANSS scores. However, the correlations between cognitive functions and PANSS total scores were significant ($\beta = -0.604$, $p = 0.037$), suggesting that cognitive functions may mediate the relationship between psychiatric symptoms and vocal characteristics.

Conclusion: This study underscores the potential of vocal analysis as a non-invasive tool for assessing cognitive impairment in schizophrenia. Future research should focus on expanding the sample size and including diverse populations to enhance the generalizability of these findings.

Plain Language Summary:

Why was the study done? Schizophrenia is a complex mental health disorder that affects how people think, feel, and behave, often leading to challenges in communication and cognitive function. This study aimed to explore how changes in vocal features can relate to the severity of psychiatric symptoms and cognitive abilities in individuals diagnosed with schizophrenia. We wanted to fill a gap in existing research about the connections between vocal patterns and cognitive functions in people with schizophrenia.

What did the researchers do and find? We analyzed voice data from 14 participants to see if there were links between their vocal features, psychiatric symptoms, and cognitive functions. Our findings showed that poorer cognitive abilities, like memory and attentional control, negatively correlated with certain vocal characteristics. Additionally, higher psychiatric symptom severity was associated with worse cognitive performance.

What do these results mean? This study suggests that analyzing vocal features could help us better understand cognitive impairments in schizophrenia, offering new ways to assess patients. By looking at how vocal features and cognitive functions relate, we can develop improved evaluation methods that may lead to better treatment outcomes. This research adds to our understanding of schizophrenia and highlights the importance of using vocal analysis in future studies to enhance clinical assessments and interventions.

943

Introduction

Schizophrenia is a complex and heterogeneous psychiatric disorder that affects millions of people worldwide, presenting significant challenges in both diagnosis and treatment.[1] Traditionally, schizophrenia has been viewed as a uniform condition; however, recent studies have uncovered significant variability in symptom presentation and individual responses to treatment.[2] This disorder is typically characterized by positive symptoms, such as delusions, hallucinations, and disorganized thinking, as well as negative symptoms, including reduced emotional expression, social withdrawal, and diminished speech output.[3]

Patients with schizophrenia are also known to exhibit widespread cognitive impairments.[4,5] Meta-analyses indicate that adult patients with schizophrenia experience a decline in IQ during the first episode, and this decline becomes more pronounced in chronic cases.[6] Cognitive impairments, including deficits in attention, working memory, executive function, verbal memory, and processing speed are also common and contribute to the complexity of the clinical picture.[7,8] These cognitive impairments substantially contribute to poor clinical outcomes, such as unemployment and difficulty with independent living.[9] In conjunction with positive and negative symptoms, these cognitive impairments significantly affect daily functioning and diminish quality of life.[10,11] Delayed treatment or prolonged course of the disorder can further exacerbate deterioration in executive function[12] and social functioning,[13] highlighting the importance of early diagnosis, comprehensive psychiatric assessment, and timely intervention.

Effective management of schizophrenia requires continuous, multidimensional assessment of symptom severity and progression. Clinical assessment tools, such as the Positive and Negative Syndrome Scale (PANSS), are routinely used to evaluate various dimensions of symptomatology.[14] Although PANSS and similar scales are useful, they may not fully capture the entire spectrum of schizophrenia's heterogeneity, underscoring the need for a comprehensive approach that integrates psychiatric symptoms, cognitive functions, social supports, and comorbid conditions.[10]

In recent years, interest in utilizing advanced computational techniques to analyze vocal features in patients with schizophrenia has grown.[15] Speech disturbances, including alterations in prosody, syntax, and semantics, are frequently observed in patients with schizophrenia and serve as valuable indicators for diagnosing and monitoring the progression of the disorder.[16,17] These disturbances are often associated with both positive and negative symptoms of schizophrenia, such as delusions, disorganized thinking, and emotional blunting.[18] Vocal analysis provides insights that enhance the understanding of psychiatric symptoms and cognitive functions, revealing patterns that traditional clinical assessments may overlook.[19] Vocal characteristics have been used to identify subtypes within psychotic disorders, with different vocal patterns linked to variations in symptom severity and cognitive functions.[20] A study evaluating the feasibility of collecting speech samples from patients with psychiatric disorders demonstrated that vocal features could serve as a useful tool for tracking clinical states over time, with individually trained models demonstrating strong correlations with clinician ratings.[21] Given the complexity and heterogeneity of this disorder, vocal analysis has gained considerable attention as a potential tool to enhance clinical assessments and provide a more objective measure of symptom severity.[22]

Among various computational techniques, vocal analysis using digital signal processing methods has emerged as one of the most promising approaches.[23] These methods allow for the quantification of acoustic features that may be associated with specific symptoms of schizophrenia.[24] Notably, techniques such as chromagram, mel spectrogram, and mel-frequency cepstral coefficients (MFCC) are being increasingly utilized for this purpose.[23,25] A chromagram provides a comprehensive representation of vocal signals in the time-frequency domain, capturing variations in pitch, intensity, and rhythm over time.[26] This approach has been employed in several studies to quantify prosodic changes in the vocal features of patients with schizophrenia.[16] Similarly, mel spectrogram represents the frequency spectrum of an audio signal over time with a frequency scale warped to match the human auditory system's response, highlighting subtle disruptions in vocal patterns, particularly regarding phonemic articulation.[16] MFCC values are derived from mel spectrogram using a discrete cosine transform. They are widely used in both speech recognition and classification tasks, as they capture the spectral and temporal properties of vocal features, providing a compact representation of these characteristics.[27] The process of generating MFCC involves several steps. First, the input audio signal is transformed into a spectrogram, and mel-frequency filters are applied. These filters resample the frequency axis to mimic the logarithmic pitch sensitivity of human hearing, offering finer resolution at lower frequencies and coarser resolution at higher frequencies. The resulting mel-filtered data are then log-transformed and

processed through a cepstral transform, which applies a series of cosine terms to the spectra. This transformation has proven valuable in reducing the dimensionality of the voice data while retaining essential acoustic features.[28]

Voice analysis has emerged as a valuable tool for assessing neurological and psychiatric disorders, providing insights into clinical symptoms through acoustic features. A recent systematic review highlighted the increasing role of voice analysis in neurological and psychiatric disorders, including schizophrenia, with MFCC, jitter, shimmer, and fundamental frequency (F0) being the most commonly used speech features.[29] Similarly, another systematic review on speech recognition emphasized the widespread use of deep neural networks and identified MFCC as the most commonly used speech feature.[30] Some studies have adopted the average values of vocal features, including MFCC, to capture broader speaker characteristics and mitigate the influence of short-term variations in speech.[31] Previous studies have demonstrated that alterations in MFCC values correlate with symptom severity and may offer a non-invasive, real-time method to track disease progression.[24]

Convergent progress in speech research and computer sciences opens avenues for implementing speech analysis to enhance the objectivity of assessment in clinical practice.[32] Machine learning approaches have been widely applied to classify speech patterns in schizophrenia and other psychiatric disorders, demonstrating their potential for improving the assessment of schizophrenia-related cognitive and psychiatric impairments.[29] Recent research has underscored the potential of these advanced vocal analysis techniques not only for assessing symptom severity but also for predicting treatment outcomes and relapse in schizophrenia.[33] This advancement would allow clinicians to incorporate more objective, quantitative tools to complement existing diagnostic methods such as clinical interviews and symptom rating scales.[34]

We hypothesized that vocal analysis could serve as a promising approach for evaluating psychiatric symptoms and cognitive functions in patients with schizophrenia. As an assessment tool, vocal analysis could enhance objective symptom evaluation, particularly given the findings that indicate distinctions in vocal characteristics between schizophrenia and other psychiatric disorders.[35] However, to date, no study has reported on the utility of relatively new machine learning-based vocal features, such as the chromagram, mel spectrogram, and MFCC, for evaluating the psychiatric symptoms of schizophrenia in real-world clinical settings.

Therefore, we conducted a comprehensive correlational analysis of vocal features, psychiatric symptoms, and cognitive functions in patients with schizophrenia to enhance both understanding and management of the disorder. Our study uses a range of acoustic features to examine their associations with psychiatric symptoms and cognitive functions. This approach offers a broader perspective compared to previous studies by exploring multiple features to gain a better understanding of the relationship between vocal characteristics and schizophrenia.

Methods

This study was conducted by the University of Occupational and Environmental Health and the Kyushu Institute of Technology in accordance with the tenets of the Declaration of Helsinki (1975, revised in 2008) and the Ethical Guidelines for Medical and Health Research Involving Human Participants in Japan. All procedures involving human participants were approved by the Ethics Committee of the University of Occupational and Environmental Health Sciences (approval number UOEHCRB21-090).

Participants

Patients were recruited from the outpatient clinic of UOEH Hospital. Written informed consent was obtained from all participants, who were assigned anonymous identification numbers to ensure privacy. An experienced psychiatrist (T.N.) diagnosed schizophrenia based on a clinical interview following the criteria outlined in the Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-5). Participants were selected based on the following inclusion criteria: (1) diagnosed with schizophrenia according to DSM-5, (2) aged 20 years or older but younger than 65 years at the time of participation, (3) receiving outpatient treatment, (4) able to provide informed consent for study participation. Exclusion criteria included: (1) individuals who are not native Japanese speakers or do not have sufficient proficiency in Japanese to complete the assessments, (2) individuals with severe neurological disorders (eg, Parkinson's disease, stroke, epilepsy), and (3) individuals with hearing impairments or speech disorders that could affect vocal analysis. The study initially included two male and 14 female participants; however, one female participant experienced an

exacerbation of psychiatric symptoms between the date of consent and the cognitive assessment, hindering her ability to comprehend the test. Another male participant could not undergo the cognitive assessment because of difficulties in understanding the contents of the test. Consequently, 14 participants (1 male and 13 females) were included in the analysis.

Audio Measurement and Processing

The development of an automated questioning system and voice samples acquisition were carried out by a team from the Kyushu Institute of Technology. An automated questioning system was designed to collect voice samples under standardized conditions, minimizing factors that could influence responses, such as the relationship between the questioner and the participant, or the distance between the participant and the microphone array. This system not only presented questions but also recorded the participants' spoken responses.

Participants were instructed to read 30 questions from the Japanese version of the Schizophrenia Quality of Life Scale to familiarize themselves with the task. The automated questioning system then presented these questions both audibly and visually, and participants responded with one of five possible answers: "Always", "Often", "Sometimes", "Rarely", or "Never". The system recorded the participants' spoken responses, captured by the microphone array positioned between the automated questioning system and the participant (see Figure 1), with a sampling frequency of 48 kHz. All voice samples were collected in a quiet, private room located in an area with no foot traffic, ensuring that external noise did not interfere with recordings. To further control environmental noise, the microphone and participant were kept at a fixed distance throughout the recordings, and noise reduction was applied during data processing to remove background interference and enhance recording clarity.

Voice samples for the five response categories (for example, "Never") were manually extracted using Audacity® 3.2.5 (https://support.audacityteam.org/additional-resources/changelog/older-versions/audacity-3.2/audacity-3.2.5), a voice editing software. This process facilitated the extraction of voice data for the aforementioned five types of responses. The vocal features, including 12-dimensional chromagram, 128-dimensional mel spectrogram, and 40-dimensional MFCC values, were then extracted using the Librosa audio library (https://librosa.org/doc/latest/index.html). The average values of these extracted features were then calculated for analysis.

Assessment of Psychiatric Symptoms and Cognitive Functions

On the same day as the audio measurements, an experienced psychiatrist (T.N.) assessed psychiatric symptoms using the Positive and Negative Syndrome Scale (PANSS). The positive subscale assesses symptoms such as delusions, hallucinations, and conceptual disorganization, reflecting the severity of positive symptoms. The negative subscale evaluates
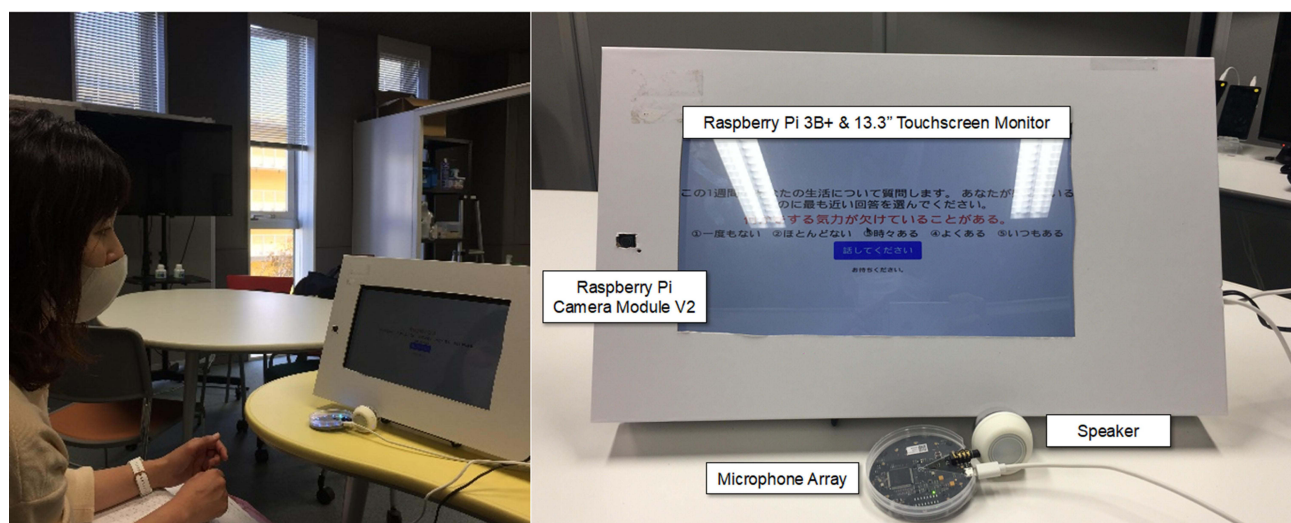

Figure 1 Preliminary experiment setup. The participants in (see Figure 1 have given written consent to be included in the Figure.

symptoms such as blunted affect, emotional withdrawal, and apathy, representing negative symptoms. The general psychopathological subscale examines symptoms such as anxiety, depression, and cognitive disturbances that do not fall under positive or negative categories. The total PANSS score is the sum of all item scores, providing an overall measure of symptom severity. Each item is rated on a scale from 1 to 7, with higher scores indicating more severe symptoms.

Cognitive functions were assessed using a computer-based tool, Cognitrax (CNS Vital Signs, Morrisville, NC, USA).[36] The test battery assesses several cognitive domains, including verbal memory, executive function, social cognition, working memory, sustained attention, and processing speed. Verbal memory was evaluated using the "Verbal Memory Test", executive function using the "Shifting Attention Test", social cognition using the "Perception of Emotions Test", and working memory using the "4-Part Continuous Performance Test". Higher scores generally indicate better cognitive performance; there are no standardized normal ranges for Cognitrax scores.

Statistical Analysis

All statistical analyses were conducted using SPSS software (IBM SPSS Inc., Armonk, NY, USA) and Python 3.0 (https://www.python.org/download/releases/3.0/). Descriptive statistics for demographic data are presented as means (standard deviation) or medians [interquartile range]. Multiple correlation analyses were performed to investigate the relationship between vocal features (chromagram, mel spectrogram, and MFCC) and psychiatric symptoms or cognitive functions, with age and sex included as covariates.

Acoustic features were extracted in small windows of audio data, resulting in multiple values for a given recording. To facilitate data analysis, we computed the average values of each vocal feature across all windows. For instance, the 40 dimensions of MFCC values were averaged across all windows to obtain a single representative value for each dimension.

Finally, correlation analyses were conducted with psychiatric symptoms as the independent variable and cognitive functions as the dependent variable, controlling for age and sex as covariates. P-values less than 0.05 were considered statistically significant.

Given the gender imbalance (92.9% female participants, n = 13), we conducted a secondary analysis using only the female participants to assess whether sex influenced the results. This analysis was executed using the same methods and vocal features as in the main analysis, with age included as a covariate in all statistical analyses.

Results
Demographic and Clinical Characteristics

Table 1 details the demographic and clinical characteristics of the participants (n = 14). The average age of the participants was 37.36 ± 12.08 years, with a female predominance (92.9%). The average age of onset of schizophrenia was 23 ± 7.05 years, with an average disease duration of 14.36 ± 7.12 years. Chlorpromazine equivalents for medication dosages ranged from 378.75 to 575.00 mg. The PANSS evaluation indicated mild to moderate severity of schizophrenia, with the following scores: PANSS positive scores (median: 15.0, IQR: 14.0–15.0), PANSS negative scores (median: 14.0, IQR: 14.0–14.75), PANSS general psychopathological scores (median: 32.0, IQR: 32.0–33.0), and PANSS total scores (median: 62.0, IQR: 61.0–63.0). Cognitive assessment scores varied across domains: verbal memory (median: 53.50, IQR: 50.50–57.50), executive function (median: 37.50, IQR: 27.25–44.00), social cognition (median: 7.00, IQR: 3.50–8.00), working memory (median: 11.00, IQR: 6.50–13.75), sustained attention (median: 32.00, IQR: 28.50–35.75), and processing speed (median: 105.50, IQR: 101.25–119.00).

Associations of Vocal Features with Psychiatric Symptoms and Cognitive Functions

In the correlation analysis, significant negative correlations were observed between vocal features and cognitive functions, whereas no such correlations were found between vocal features and psychiatric symptoms. Specifically, significant negative correlations were found between working memory and both mel spectrogram (β = −0.645, p = 0.023) and MFCC (β = −0.669, p = 0.017). Furthermore, negative correlations were identified between sustained attention and mel spectrogram (β = −0.626, p = 0.029), as well as between sustained attention and MFCC (β = −0.578, p = 0.049), as shown in Table 2. However, no

Table 1 Demographic and Clinical Data of Study Participants

Demographic Data	Participants (n = 14)
Age, years	37.36 ± 12.08
Sex, female	13 (92.9%)
Age of schizophrenia onset	23.00 ± 7.05
Duration of disease	14.36 ± 7.12
Chlorpromazine Equivalents, mg	400.00 [378.75–575.00]
Clinical Data	
PANSS Positive	15.0 [14.0–15.0]
PANSS Negative	14.0 [14.0–14.75]
PANSS General Psychopathological	32.0 [32.0–33.0]
PANSS Total	62.0 [61.0–63.0]
Verbal Memory	53.50 [50.50–57.50]
Executive Function	37.50 [27.25–44.00]
Social Cognition	7.00 [3.50–8.00]
Working Memory	11.00 [6.50–13.75]
Sustained Attention	32.00 [28.50–35.75]
Processing Speed	105.50 [101.25–119.00]

Note: All data are presented as numbers (percentages), or as means ± standard deviations, or as medians [interquartile ranges].
Abbreviation: PANSS, Positive and Negative Syndrome Scale.

Table 2 Correlation of Psychiatric Symptoms and Vocal Features with Cognitive Functions

Variables	Working Memory	Sustained Attention
PANSS Total	−0.604*	−0.650*
PANSS Positive	−0.567	−0.676*
Mel Spectrogram	−0.645*	−0.626*
MFCC	−0.669*	−0.578*

Note: *p<0.05.
Abbreviations: PANSS, Positive and Negative Syndrome Scale; MFCC, mel-frequency cepstral coefficients.

statistically significant correlations were found between any of the vocal features and psychiatric symptoms (PANSS scores). Significant negative correlations of the PANSS total score with both working memory ($\beta = -0.604$, p = 0.037) and sustained attention ($\beta = -0.650$, p = 0.022) were found. Additionally, sustained attention was negatively correlated with the PANSS positive score ($\beta = -0.676$, p = 0.016) (Figures 2 and 3).

A secondary analysis, conducted using only the female participants (n = 13), showed results consistent with those of the entire sample. Specifically, the significant negative correlations between vocal features (mel spectrogram and MFCC) and cognitive functions, such as working memory and sustained attention, remained unchanged. Similarly, the relationship between cognitive functions and psychiatric symptoms remained consistent in this subgroup.
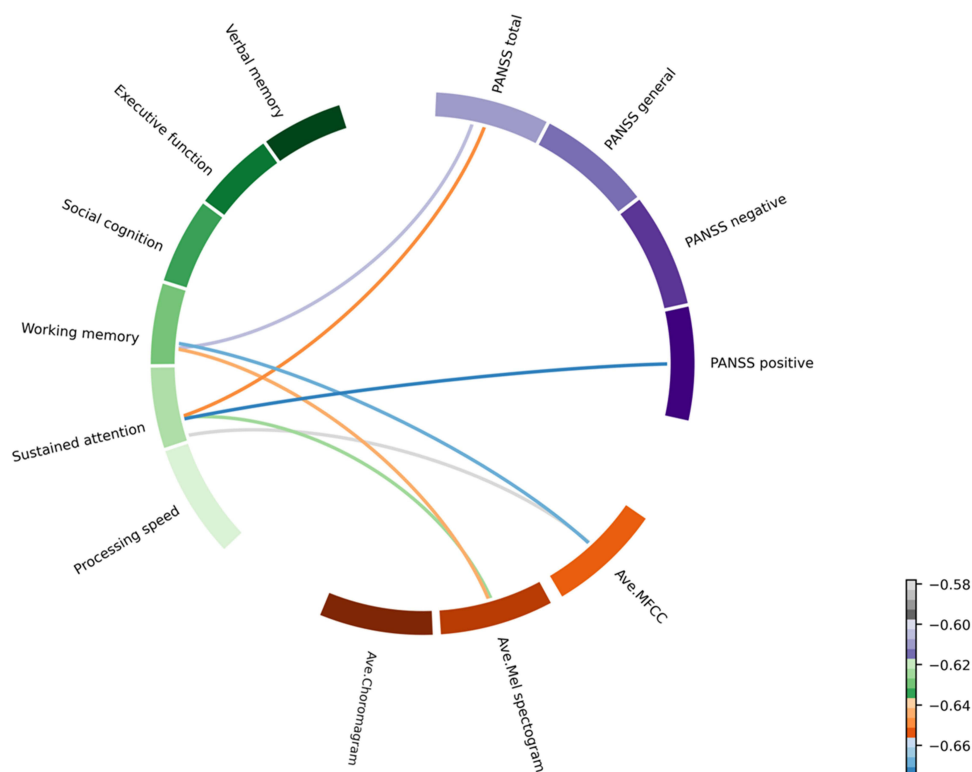
Figure 2 Chord diagram of correlations between voice features, psychiatric symptoms, and cognitive functions.

Discussion

This study examined the relationships between psychiatric symptoms, cognitive functions, and vocal features in patients with schizophrenia. We found significant correlations between vocal features and cognitive functions in the 14 participants. Notably, negative correlations were observed between certain cognitive functions—specifically working memory and sustained attention—and vocal features extracted from mel spectrogram and MFCC analyses. Although no direct correlations between psychiatric symptoms and vocal features were identified, cognitive functions showed negative correlations with the severity of psychiatric symptoms, including both the PANSS total scores and PANSS positive scores.

Cognitive impairment—particularly deficits in attention, memory, and executive function—is often more debilitating than positive and negative symptoms concerning their impact on daily functioning.[37] Identifying objective vocal features that correlate with cognitive functions holds considerable promise for both clinical practice and research in schizophrenia.[38] In our study, the negative correlations between vocal features (mel spectrogram and MFCC) and cognitive functions (working memory and sustained attention) emphasize the potential of vocal features as markers of cognitive impairment in schizophrenia. Previous studies have suggested that abnormalities in specific neural oscillations affect language function. In schizophrenia, beta oscillations are particularly impaired and are closely associated with inhibitory control in the brain. Beta oscillations prevent irrelevant information from interfering with language processing. During working memory maintenance, increased beta activity is thought to protect stored information from external distractions. However, in patients with schizophrenia, the reduction in beta oscillations weakens this inhibitory function, leading to excessive activation of irrelevant information. Moreover, beta oscillations also play a role in motor control, including the coordination of speech production.[39] The observed associations between cognitive functions and vocal features in this study may be attributable to beta oscillation abnormalities.

There is a growing need for objective and reliable methods to assess cognitive impairment in schizophrenia. Traditional cognitive assessments typically rely on performance-based tasks and subjective evaluations, which can be influenced by factors such as patient motivation and examiner bias.[40] Therefore, objective vocal analysis methods, such as those based on mel spectrogram and MFCC, could complement existing cognitive assessments, providing clinicians
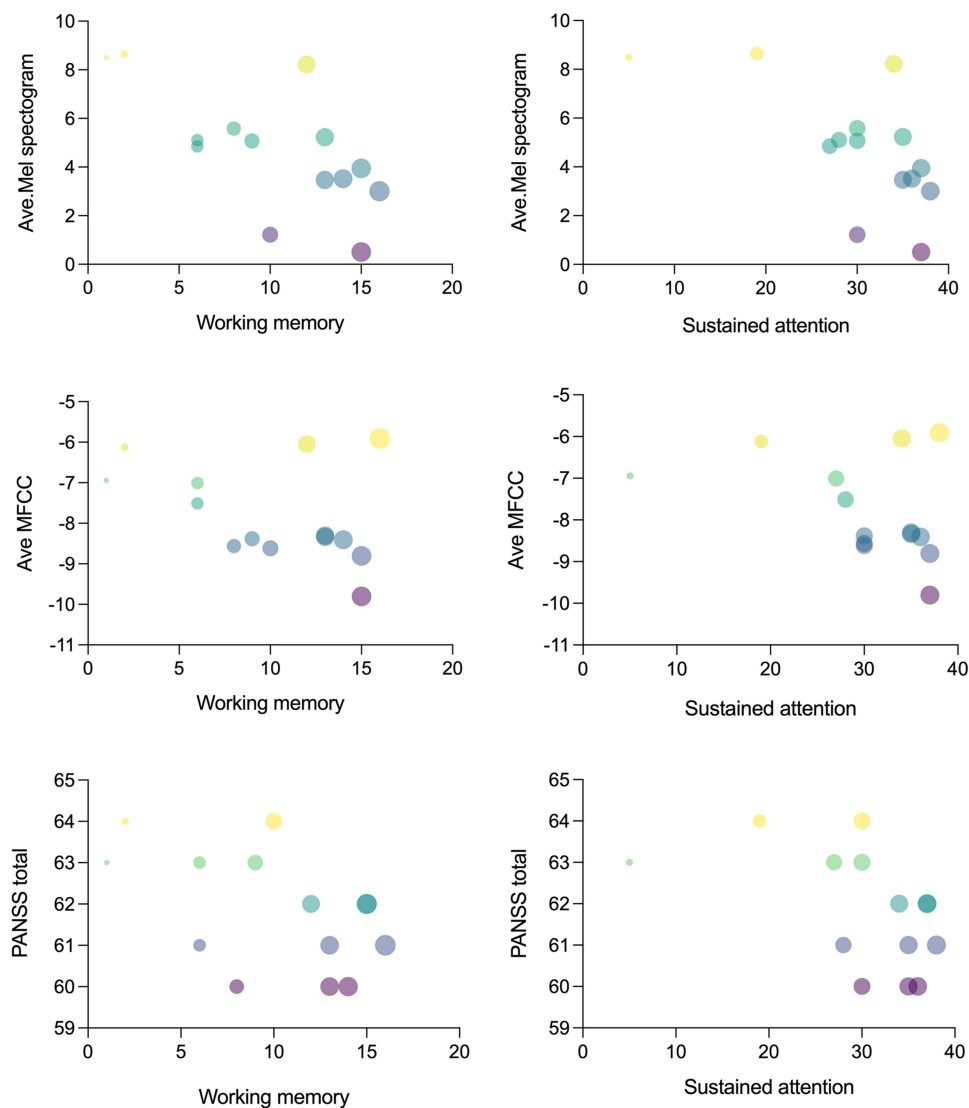
Figure 3 Pearson correlations between vocal features, psychiatric symptoms, and cognitive functions, with bubble size indicating cognitive scores and color representing symptom scores.

with additional non-invasive tools for evaluating cognitive impairment. However, larger-scale studies are necessary to determine whether these objective vocal features can facilitate early detection of cognitive decline and enable more targeted interventions aimed at improving cognitive functions in schizophrenia.

Our study also identified significant negative correlations between cognitive functions and PANSS scores, suggesting that higher levels of psychiatric symptoms are associated with poorer cognitive performance across multiple domains. Specifically, sustained attention was negatively correlated with both PANSS total scores and positive scores, while working memory showed a negative correlation with PANSS total scores. These results align with prior research linking cognitive impairments to traditional psychiatric symptoms, including positive, negative, and general psychopathological symptoms.[41,42] However, previous studies have reported that the disorganization factor in the PANSS—which includes attention deficits, disorientation, impaired conceptualization, and difficulties with abstract thinking—shows the strongest association with cognitive scale scores.[41,42] Moreover, recent studies using network analyses have found no clear correlation between positive and negative symptoms and cognitive scale scores, but have shown associations between cognitive deficits and general psychopathological symptoms.[43,44] These discrepancies highlight the complexity of the relationship between cognitive impairment and psychiatric symptoms in schizophrenia, which may differ across patients.

Despite the absence of direct correlations between psychiatric symptoms and vocal features, our results suggest that cognitive functions may mediate the relationship between psychiatric symptoms and vocal characteristics. This finding contrasts with previous studies that have identified associations between certain psychiatric symptoms, such as disorganization in thought or behavior, and distinct vocal characteristics.[45,46] However, the lack of significant direct correlations does not rule out a potential relationship between psychiatric symptoms and vocal features. Factors such as sample size, symptom heterogeneity, and variability in vocal patterns may have influenced these results. Future studies with larger and more diverse samples are needed to clarify the relationship between psychiatric symptoms and vocal features.

Although cognitive impairments in schizophrenia are likely to contribute to vocal processing differences detected through acoustic analysis, our study has several limitations. First, the small sample size and the lack of a control group reduced the robustness of our findings. With only 14 participants, the statistical power of our analysis was limited, and some relationships may not have been detected due to insufficient sample size. A power analysis revealed a statistical power of 24.71%, which is below the recommended threshold of 80%. Thus, the generalizability of our findings is limited, necessitating future studies with larger and more diverse samples to validate our results and improve statistical robustness. Additionally, the psychiatric symptoms among our sample were relatively mild, as only outpatients were included. This may have introduced sampling bias, limiting generalizability. Furthermore, participant exposure to antipsychotic treatment and the lack of control for educational background may have influenced both vocal patterns and cognitive outcomes.

Second, sex differences in vocal features present a limitation. Males and females exhibit differences in vocal tract length, affecting vocal characteristics such as fundamental frequencies and MFCC. Our sample had a significant sex imbalance, with 92.9% of participants being female, which may have influenced the results. The predominance of female participants could have biased the acoustic measures, limiting the applicability of our findings to the broader schizophrenia population. Future studies should consider sex as a factor in vocal analysis to better understand its impact on the relationship between vocal features, cognitive functions, and psychiatric symptoms in schizophrenia. Ensuring a more balanced sex distribution in future research will be essential to assess whether sex-specific differences affect the observed associations.

Third, our study was limited by the methodology that utilized only single-word responses and short utterances of approximately 1 second. Although this approach was necessary to standardize experimental conditions, it does not capture the full complexity of natural speech. Moreover, it remains unclear whether our findings are applicable to languages other than Japanese. Future research should address these limitations by incorporating multi-word responses, larger and more diverse samples, control groups, and longitudinal designs to track changes in vocal features over time. Additionally, our study primarily focused on acoustic characteristics and did not take into account syntactic and semantic aspects of speech, which could further enhance our understanding of the relationship between vocal features and cognitive and psychiatric functions in schizophrenia.

Finally, recent research highlights the complexity of speech as both a biosocial marker and a neurobiological indicator of psychiatric symptoms. Speech is influenced not only by biological factors but also by social determinants such as socioeconomic background and interpersonal interactions.[47] Additionally, a study has shown that PANSS Disorganization (PANSS-P2) is more closely associated with linguistic style, particularly lower analytic thinking scores, than general cognitive impairments.[48] Another finding links conceptual disorganization to altered effective connectivity between the anterior cingulate cortex and anterior insula, highlighting its neurobiological basis.[49] However, our study did not separately examine the PANSS Disorganization factor, which may provide further insight into the relationship between vocal features and clinical symptoms. Moreover, our analysis was limited to a structured reading task rather than naturalistic or spontaneous speech, which may restrict the generalizability of our findings. Future research should incorporate spontaneous speech analysis to better capture the full scope of linguistic disturbances in schizophrenia and further explore the interplay between speech characteristics, cognitive functions, and psychiatric symptoms.

Conclusion

This study highlights the potential interplay between vocal features, psychiatric symptoms, and cognitive functions in patients with schizophrenia. By demonstrating correlations between vocal features and cognitive impairments, this study contributes to the growing body of research exploring vocal markers as potential indicators of cognitive impairments in schizophrenia. These results highlight the potential of vocal analysis in detecting early signs of cognitive impairments

and psychiatric symptoms, which could lead to more timely and targeted interventions. Future research should expand the scope of vocal analysis to include a broader range of linguistic features such as prosody, syntax, and semantics. Larger, longitudinal studies with control groups are needed to further validate these findings and refine clinical applications, improving diagnostic accuracy and intervention outcomes for patients with schizophrenia.

Data Sharing Statement

The data generated and analyzed during this study are not publicly available due to privacy concerns but are available from the corresponding author on reasonable request.

Ethical Statement

This study was conducted collaboratively by the University of Occupational and Environmental Health and the Kyushu Institute of Technology in accordance with the tenets of the Declaration of Helsinki (1975, revised in 2008) and the Ethical Guidelines for Medical and Health Research Involving Human Participants in Japan. All procedures involving human participants were approved by the Ethics Committee of the University of Occupational and Environmental Health Sciences (approval number UOEHCRB21-090). Written informed consent was obtained from all participants, who were assigned anonymous identification numbers to ensure privacy.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

Disclosure

All authors declare that they have no conflicts of interest in this work.

References

1. Lyu J, He H, Liu Q, et al. Trends in the incidence and DALYs of schizophrenia at the global, regional and national levels: results from the Global Burden of Disease Study 2017. *Epidemiol Psychiatr Sci*. 2020;29:e91. doi:10.1017/S2045796019000891
2. Schulz SC, Michael FG, Katharine JN, editors. *Schizophrenia and Psychotic Spectrum Disorders*. Online ed. New York: Oxford Academic; 2016. doi:10.1093/med/9780199378067.001.0001
3. Jauhar S, Johnstone M, McKenna PJ. Schizophrenia. *Lancet*. 2022;399:473–486. doi:10.1016/S0140-6736(21)01730-X
4. Heinrichs RW, Zakzanis KK. Neurocognitive deficit in schizophrenia: a quantitative review of the evidence. *Neuropsychology*. 1998;12 (3):426–445. PMID: 9673998. doi:10.1037/0894-4105.12.3.426
5. Mohamed S, Paulsen JS, O'Leary D, Arndt S, Andreasen N. Generalized cognitive deficits in schizophrenia: a study of first-episode patients. *Arch Gen Psychiatry*. 1999;56(8):749–754. PMID: 10435610. doi:10.1001/archpsyc.56.8.749
6. Fioravanti M, Carlone O, Vitale B, Cinti ME, Clare L. A meta-analysis of cognitive deficits in adults with a diagnosis of schizophrenia. *Neuropsychol Rev*. 2005;15:73–95. doi:10.1007/s11065-005-6254-9
7. Mcdowd J, Tang T-C, Tsai P-C, Wang S-Y, Su C-Y. The association between verbal memory, processing speed, negative symptoms and functional capacity in schizophrenia. *Psychiatry Res*. 2011;187:329–334. doi:10.1016/j.psychres.2011.01.017
8. Panov G, Dyulgerova S, Panova P. Cognition in patients with schizophrenia: interplay between working memory, disorganized symptoms, dissociation, and the onset and duration of psychosis, as well as resistance to treatment. *Biomedicines*. 2023;11(12):3114. PMID: 38137335; PMCID: PMC10740456. doi:10.3390/biomedicines11123114
9. Green MF. What are the functional consequences of neurocognitive deficits in schizophrenia? *Am J Psychiatry*. 1996;153:321–330. doi:10.1176/AJP.153.3.321
10. Green MF, Kern RS, Braff DL, Mintz J. Neurocognitive deficits and functional outcome in schizophrenia: are we measuring the "right stuff"? *Schizophr Bull*. 2000;26:119–136. doi:10.1093/OXFORDJOURNALS.SCHBUL.A033430
11. Reichenberg A, Harvey PD. Neuropsychological impairments in schizophrenia: integration of performance-based and brain imaging findings. *Psychol Bull*. 2007;133(5):833–858. Erratum in: Psychol Bull. 2008;134(3):382. PMID: 17723032. doi:10.1037/0033-2909.133.5.833
12. Bagney A, Rodriguez-Jimenez R, Martinez-Gras I, et al. Negative symptoms and executive function in schizophrenia: does their relationship change with illness duration? *Psychopathology*. 2013;46:241–248. doi:10.1159/000342345

13. Duțescu MM, Popescu RE, Balcu L, et al. Social functioning in schizophrenia clinical correlations. *Curr Health Sci J*. 2018;44(2):151–156. PMID: 30746163; PMCID: PMC6320466. doi:10.12865/CHSJ.44.02.10

14. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr Bull*. 1987;13:261–276. doi:10.1093/SCHBUL/13.2.261

15. García Molina JT, Gaspar PA, Figueroa-Barra A. Automatic speech recognition in psychiatric interviews: a rocket to diagnostic support in psychosis. *Rev Colomb Psiquiatr*. 2024. doi:10.1016/J.RCP.2023.12.002

16. Elvevåg B, Foltz PW, Rosenstein M, DeLisi LE. An automated method to analyze language use in patients with schizophrenia and their first-degree relatives. *J Neurolinguistics*. 2010;23:270–284. doi:10.1016/J.JNEUROLING.2009.05.002

17. Corcoran CM, Cecchi GA. Using language processing and speech analysis for the identification of psychosis and other disorders. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 2020;5:770–779. doi:10.1016/J.BPSC.2020.06.004

18. Berardi M, Brosch K, Pfarr JK, et al. Relative importance of speech and voice features in the classification of schizophrenia and depression. *Transl Psychiatry*. 2023;13(1):298. PMID: 37726285; PMCID: PMC10509176. doi:10.1038/s41398-023-02594-0

19. Alonso-Sánchez MF, Ford SD, MacKinley M, et al. Progressive changes in descriptive discourse in first episode schizophrenia: a longitudinal computational semantics study. *Schizophrenia*. 2022;8:36. doi:10.1038/s41537-022-00246-8

20. Oomen PP, de Boer JN, Brederoo SG, et al. Characterizing speech heterogeneity in schizophrenia-spectrum disorders. *J Psychopathol Clin Sci*. 2022;131(2):172–181. PMID: 35230859. doi:10.1037/abn0000736

21. Arevian AC, Bone D, Malandrakis N, et al. Clinical state tracking in serious mental illness through computational analysis of speech. *PLoS One*. 2020;15(1):e0225695. PMID: 31940347; PMCID: PMC6961853. doi:10.1371/journal.pone.0225695

22. Goldsmith DR, Crooks CL, Walker EF, Cotes RO. An update on promising biomarkers in schizophrenia. *Focus*. 2018;16(2):153–163. PMID: 31975910; PMCID: PMC6526854. doi:10.1176/appi.focus.20170046

23. Zhang J, Pan Z, Gui C, Zhu J, Cui D. Clinical investigation of speech signal features among patients with schizophrenia. *Shanghai Arch Psychiatry*. 2016;28(2):95–102. PMID: 27605865; PMCID: PMC5004093. doi:10.11919/j.issn.1002-0829.216025

24. Huang J, Zhao Y, Tian Z, et al. Evaluating the clinical utility of speech analysis and machine learning in schizophrenia: a pilot study. *Comput Biol Med*. 2023;164:107359. PMID: 37591160. doi:10.1016/j.compbiomed.2023.107359

25. Huang YJ, Lin YT, Liu CC, et al. Assessing schizophrenia patients through linguistic and acoustic features using deep learning techniques. *IEEE Trans Neural Syst Rehabil Eng*. 2022;30:947–956. PMID: 35358049. doi:10.1109/TNSRE.2022.3163777

26. Birajdar G, Patil M. Speech/music classification using visual and spectral chromagram features. *J Ambient Intell Human Comput*. 2020;11:329–347. doi:10.1007/s12652-019-01303-4

27. Issa D, Fatih Demirci M, Yazici A. Speech emotion recognition with deep convolutional neural networks. *Biomed Signal Process Control*. 2020;59:101894. doi:10.1016/J.BSPC.2020.101894

28. Tracey B, Volfson D, Glass J, et al. Towards interpretable speech biomarkers: explaining MFCCs. *Sci Rep*. 2023;13(1):22787. PMID: 38123603; PMCID: PMC10733367. doi:10.1038/s41598-023-49352-2

29. Hecker P, Steckhan N, Eyben F, Schuller BW, Arnrich B. Voice analysis for neurological disorder recognition—a systematic review and perspective on emerging trends. *Front Digit Health*. 2022;4:842301. doi:10.3389/fdgth.2022.842301

30. Nassif AB, Shahin I, Attili I, Azzeh M, Shaalan K. Speech recognition using deep neural networks: a systematic review. *IEEE Access*. 2019;7:19143–19165. doi:10.1109/ACCESS.2019.2896880

31. Kos M, Vlaj D, Kačič Z. Speaker's gender classification and segmentation using spectral and cepstral feature averaging. In: *2011 18th International Conference on Systems, Signals and Image Processing*. IEEE; 2011:1–4.

32. Dikaios K, Rempel S, Dumpala SH, Oore S, Kiefte M, Uher R. Applications of speech analysis in psychiatry. *Harv Rev Psychiatry*. 2023;31(1):1–13. doi:10.1097/HRP.0000000000000356

33. Zaher F, Diallo M, Achim AM, et al. Speech markers to predict and prevent recurrent episodes of psychosis: a narrative overview and emerging opportunities. *Schizophr Res*. 2024;266:205–215. doi:10.1016/J.SCHRES.2024.02.036

34. Rapcan V, D'Arcy S, Yeap S, Afzal N, Thakore J, Reilly RB. Acoustic and temporal analysis of speech: a potential biomarker for schizophrenia. *Med Eng Phys*. 2010;32(9):1074–1079. PMID: 20692864. doi:10.1016/j.medengphy.2010.07.013

35. Parola A, Simonsen A, Bliksted V, Fusaroli R. Voice patterns in schizophrenia: a systematic review and Bayesian meta-analysis. *Schizophr Res*. 2020;216:24–40. doi:10.1016/J.SCHRES.2019.11.031

36. Gualtieri CT, Johnson LG. Reliability and validity of a computerized neurocognitive test battery, CNS vital signs. *Arch Clin Neuropsychol*. 2006;21:623–643. doi:10.1016/J.ACN.2006.05.007

37. Green MF, Kern RS, Heaton RK. Longitudinal studies of cognition and functional outcome in schizophrenia: implications for MATRICS. *Schizophr Res*. 2004;72:41–51. doi:10.1016/J.SCHRES.2004.09.009

38. Tang SX, Cong Y, Nikzad AH, et al. Clinical and computational speech measures are associated with social cognition in schizophrenia spectrum disorders. *MedRxiv*. 2022. doi:10.1101/2022.03.18.22272633

39. Palaniyappan L. Dissecting the neurobiology of linguistic disorganisation and impoverishment in schizophrenia. *Semin Cell Dev Biol*. 2022;129:47–60. doi:10.1016/j.semcdb.2021.08.015

40. Homayoun S, Nadeau-Marcotte F, Luck D, Stip E. Subjective and objective cognitive dysfunction in schizophrenia is there a link? *Front Psychol*. 2011;2:148. doi:10.3389/fpsyg.2011.00148

41. Bell MD, Lysaker PH, Milstein RM, Beam-Goulet JL. Concurrent validity of the cognitive component of schizophrenia: relationship of PANSS scores to neuropsychological assessments. *Psych Res*. 1994;54:51–58. doi:10.1016/0165-1781(94)90064-7

42. Rodriguez-Jimenez R, Bagney A, Mezquita L, et al. Cognition and the five-factor model of the positive and negative syndrome scale in schizophrenia. *Schizophr Res*. 2013;143:77–83. doi:10.1016/J.SCHRES.2012.10.020

43. Chang WC, Wong CSM, Or PCF, et al. Inter-relationships among psychopathology, premorbid adjustment, cognition and psychosocial functioning in first-episode psychosis: a network analysis approach. *Psychol Med*. 2020;50:2019–2027. doi:10.1017/S0033291719002113

44. Moura BM, Van Rooijen G, Schirmbeck F. A network of psychopathological, cognitive, and motor symptoms in schizophrenia spectrum disorders. *Schizophr Bull*. 2021;47:915–926. doi:10.1093/SCHBUL/SBAB002

45. Merrill AM, Karcher NR, Cicero DC, Becker TM, Docherty AR, Kerns JG. Evidence that communication impairment in schizophrenia is associated with generalized poor task performance. *Psychiatry Res*. 2017;249:172–179. doi:10.1016/j.psychres.2016.12.051

46. Tan EJ, Meyer D, Neill E, Rossell SL. Investigating the diagnostic utility of speech patterns in schizophrenia and their symptom associations. *Schizophr Res*. 2021;238:91–98. doi:10.1016/j.schres.2021.10.003

47. Palaniyappan L. More than a biomarker: could language be a biosocial marker of psychosis? *NPJ Schizophr*. 2021;7(1):42. doi:10.1038/s41537-021-00172-1

48. Silva A, Limongi R, MacKinley M, Palaniyappan L. Small words that matter: linguistic style and conceptual disorganization in untreated first-episode schizophrenia. *Schizophr Bull Open*. 2021;2(1):sgab010. doi:10.1093/schizbullopen/sgab010

49. Limongi R, Silva AM, Mackinley M, Ford SD, Palaniyappan L. Active inference, epistemic value, and uncertainty in conceptual disorganization in first-episode schizophrenia. *Schizophr Bull*. 2023;49(Suppl_2):S115–S124. doi:10.1093/schbul/sbac125