TPM

# VOCAL EMOTION RECOGNITION TEST: RELIABILITY, VALIDITY, AND CLINICAL UTILITY

HONGZHEN FAN[1]
BEIJING HUILONGGUAN HOSPITAL, PEKING UNIVERSITY, CHINA

RONGHUAN JIANG
CHINESE PEOPLE'S LIBERATION ARMY (PLA) GENERAL HOSPITAL,
CHINESE PLA MEDICAL SCHOOL, BEIJING, CHINA

YANLI ZHAO
DONG LI
YAJUN LI
YUNHUI WANG
JING YAO
YANLAN LIU
XIAOLIN ZHU
SHUPING TAN
BEIJING HUILONGGUAN HOSPITAL, PEKING UNIVERSITY, CHINA

Individuals suffering from schizophrenia demonstrate deficits in recognizing emotion based on voice. The present study aimed to develop a vocal emotion recognition test with good psychometric properties and use it for evaluation in schizophrenia. Data from 82 healthy participants and 161 stable inpatients diagnosed with schizophrenia were collected. The internal consistency and test-retest reliability of the vocal emotion recognition test (VER-42) were acceptable. Patients were less accurate than the healthy group in identifying emotion based on the VER-42 total score and were less accurate than healthy individuals across all emotions. The VER-42 total score correlated with cognitive functions and social cognition or function; and was associated negatively with negative symptoms of patients suffering from schizophrenia. The VER-42 was shown to be a valid and reliable instrument for evaluating vocal emotion recognition ability and could thus contribute to the screening of patients in the future.

Keywords: Vocal emotion recognition; Reliability; Validity; Schizophrenia; Instrumental study.

*Correspondence concerning this article should be addressed to Shuping Tan, Psychiatry Research Center, Beijing Huilongguan Hospital, Peking University, Nandian Road 7, Beijing, 100096, China. Email: shupingtan@126.com*

Social cue perception represents the initial stage of social cognitive processing, involving both facial expressions and voice cues (Green et al., 2015), which activate distinct brain mechanisms (Schirmer, 2018; Young et al., 2020). The perception of facial emotions has been typically associated with limbic regions, inferior frontal gyrus (IFG), and medial prefrontal gyrus; whereas the perception of voice emotions has been related to the superior temporal gyrus (STG) and inferior frontal gyrus (IFG) (Green et al., 2015). Individuals suffering from schizophrenia exhibit significant deficits in emotion perception tasks

(Barkl et al., 2014; Gong et al., 2021; Pinkham et al., 2018; Savla et al., 2013), which could result in maladaptive daily social functioning and interpersonal relationships (Irani et al., 2012). Facial-emotion tasks have been the most commonly used measure for evaluating emotion recognition (Ekman & Frieser, 1975; Kerr & Neale, 1993; Silver et al., 2002) and have shown a large overall effect size on facial identification impairment in schizophrenia (Chan, Li, et al., 2010; Kohler et al., 2010). Some voice emotional databases have also been established (Bänziger et al., 2011; Bellack et al., 1996; Bryson et al., 1997; Gold et al., 2012; Kerr & Neale, 1993; Lassalle et al., 2019; Pawełczyk et al., 2021), but they were rarely used for schizophrenia (Gold et al., 2012; Kerr & Neale, 1993; Pawełczyk et al., 2021) and were more often employed as a cognitive task rather than a standardized instrument (Kerr & Neale, 1993).

Moreover, vocal emotion perception impairments might represent specific deficits across emotions, although these remain unclear at present. The most widely used facial emotion recognition test has demonstrated that fear, disgust, and surprise are differential emotions for which individuals suffering from demonstrate specific processing deficits (Kohler et al., 2003). However, no specific deficits have been observed for vocal emotion recognition in conditions involving controlled stimuli (Gold et al., 2012). Furthermore, emotions are usually easily recognized when the intensity is extreme (Kohler et al., 2003), but it is not clear whether individuals suffering from schizophrenia have impairments in terms of recognizing the intensity of emotion, in addition to recognizing the type of emotion.

## The Present Study

The purpose of the present study was to develop a vocal emotion recognition test, evaluate its psychometric properties, and explore error patterns based on different emotions and intensity in patients suffering from schizophrenia. The study (1) investigated the internal consistency and test-retest reliability of the instrument, (2) established its relationship to cognitive function, emotional capacity, and clinical symptoms — cognitive function was assessed through the MATRICS Consensus Cognitive Battery (MCCB), while emotional perception, experience, and expression were evaluated using the Facial Emotion Recognition Test (FERT), the Temporal Experience of Pleasure Scale (TEPS), and the Toronto Alexithymia Scale-20 (TAS-20), respectively, and the clinical feature of schizophrenia was measured using the Positive and Negative Syndrome Scale (PANSS) — (3) determined its clinical utility in patients suffering from schizophrenia.

# METHOD

## Participants

Data were collected at the Beijing Huilongguan Hospital. Participants were stable inpatients diagnosed with schizophrenia ($N$ = 161) and investigated through a convenient sampling method, as well as healthy controls ($N$ = 82) matched by age, gender, and education. For patients, positive and negative symptom severity was low and stable. The severity of clinical symptoms was measured with the Positive and Negative Syndrome Scale (PANSS; Kay et al., 1987), administered by a well-trained psychiatrist. All participants also completed a questionnaire that obtained general information, demographic characteristics and neurocognitive and emotional capacity measures, as well as the vocal emotion recognition test. Eighty-one healthy participants completed two study examinations: a baseline and a retest conducted 1-4 weeks after the initial examination (mean interval = 14.37 days). Furthermore, 46 of the healthy participants completed neurocognitive assessments during the study period. All subjects provided written informed consent after demonstrating their understanding of the consent documents. This study was approved by the institutional review board of the Beijing Huilongguan Hospital (number 2015-90), and performed in accordance with the guidelines and regulations of the Helsinki Declaration.

Eighty-two healthy participants were involved in the reliability analysis of the VER-42, while its applicability was assessed in 161 chronic inpatients suffering from schizophrenia. The average age of onset for patients suffering from schizophrenia was 23.90 years, with a standard deviation of 7.74 years. The duration of the illness varied from 0.4 to 39 years, with an average of 17.07 years and a standard deviation of 10.91 years. The mean scores for positive symptoms, negative symptoms, and general psychopathology among patients were 12.68 ($SD$ = 5.28), 18.00 ($SD$ = 6.51), and 29.64 ($SD$ = 7.98), respectively. All patients had an average of 60.33 ($SD$ = 15.14) on the PANSS total score. Demographic and clinical characteristics are provided in Table 1. Groups did not differ in terms of gender or age ($p$ > .05), but they differed marginally in years of education ($p$ = .058).

TABLE 1
Demographic variables in clinical samples and the healthy group

| | Schizophrenia ($N$ = 161) | Healthy group ($N$ = 82) | $t/\chi^2$ | $p$ |
|---|---|---|---|---|
| | $M$ ($SD$)/$N$ | $M$ ($SD$)/$N$ | | |
| Age | 41.21 (11.96) | 39.85 (9.78) | –.95 | .342 |
| Gender (male/female) | 74/87 | 29/53 | 2.50 | .114 |
| Education (years) | 13.52 (2.40) | 14.18 (2.85) | 1.91 | .058 |
| Age of onset (years) | 23.90 (7.74) | – | – | – |
| Duration (years) | 17.07 (10.91) | – | – | – |
| PANSS | | | | |
| Positive symptom score | 12.68 (5.28) | – | – | – |
| Negative symptom score | 18.00 (6.51) | – | - | – |
| General psychopathology score | 29.64 (7.98) | – | – | – |
| Total score | 60.33 (15.14) | – | – | – |

*Note*. PANSS = Positive and Negative Syndrome Scale.

## Measures

*The MATRICS Consensus Cognitive Battery*. The MATRICS Consensus Cognitive Battery (MCCB) was used to provide a relatively brief evaluation of seven cognitive domains: speed of processing, attention/vigilance, working memory, verbal learning, visual learning, reasoning and problem solving, and social cognition (Mayer-Salovey-Caruso Emotional Intelligence Test, MSEIT). The test-retest reliabilities were generally good ($r$ > .70) (Nuechterlein et al., 2008).

*The Temporal Experience of Pleasure Scale*. The Temporal Experience of Pleasure Scale (TEPS) was used to measure experiences of pleasure (Chan, Wang, et al., 2010). It has four basic factors: abstract anticipatory, contextual anticipatory, abstract consummatory, and contextual consummatory. The internal consistency was .83 and the test-retest reliabilities were good (*r* = .79 to .81).

*Toronto Alexithymia Scale-20.* The Toronto Alexithymia Scale-20 (TAS-20) was a self-report measure designed to assess alexithymia symptoms and used to assess nonclinical variability in emotional awareness. The questionnaire consists of 20 items and was rated on a 5-point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). The internal consistency was .65 and the test-retest reliability was .78 (Bagby et al., 1994).

*Facial Emotion Recognition Test.* The Facial Emotion Recognition Test (FERT) was used to measure facial emotional recognition ability (Yanli et al., 2020). It consists of 42 items and includes seven types of emotion, with six items each. The participants were asked to judge the emotional expression type by forced choice (eight emotion options). Each correct response was scored as 1, and the participant might score 0 to 42 on the test. The internal consistency was .77 and the test-retest reliability was .81.

*Positive and Negative Syndrome Scale.* The Positive and Negative Syndrome Scale (PANSS) was used to measure patients' symptoms (Kay et al., 1987). The scale consists of 30 items and was rated on a 7-point Likert scale ranging from 1 (*none*) to 7 (*extremely severe*). The PANSS had a good reliability (coefficients > .60).

## Procedure

### *Vocal Emotion Material Library Building*

Serial vocal emotion discrimination items were developed with reference to the emotion portrayals described by Juslin and Laukka (2001); these included eight emotion domains: happiness, sadness, anger, fear, disgust, surprise, sarcasm, and neutral. Each emotion contained three intensity levels: *high*, *medium*, and *low* (with voice intensities in place of emotion intensities for "neutral"). Two neutral content items in Chinese ("What time is it now?" or "Now it is eight o'clock") were used to avoid providing additional cues that would help in discriminating the emotion (Kerr & Neale, 1993). Four professional drama actors (two males and two females) were used as speakers of the vocal items, and the inclusion of two age groups (20~30- and 40~50-year-olds) was intended to increase the generalizability of the results. All the actors performed a 1-week off-site technical training to meet professional project requirements before making the recording and were paid for their anonymous participation. The recordings took place in a professional recording studio. The vocal portrayals were stored in computer memory with a 48-kHz sampling frequency. Two sentences, four actors, eight emotions, and three emotion intensities yielded a total of 192 vocal portrayals. The participants wore headphones (SHM1900), sat alone at the computer desk in a hospital room, and used the mouse to give a response on the computer screen.

### *Pretest*

The 192 items from the vocal emotion material library were placed in a random order to generate the beta version of the vocal emotion discrimination test. In total, 138 people were recruited through advertisements to evaluate the beta version of the test. The participants were asked to judge the emotional expression type by forced choice (eight emotion options) and strength (from 1 to 99 percent) for each item.

Participants could listen to each item as many times as needed. The experiment took approximately 40 minutes, including a short break. The formal test had to be carried out after passing a practice test (correct rate > 70%), which consisted of 10 easy, high-intensity facial emotion recognition tasks, and each participant had two opportunities to pass. Of the 138 individuals, five refused, 15 failed to pass the practice test, and 21 failed to complete the entire test; thus, eventually 97 healthy persons (39 men and 58 women; aged 25-62 years, mean = 40.34; mean number of years of education = 14.41, *SD* = 3.10) completed the beta version evaluation. These participants were either paid or given course credits for their anonymous and voluntary participation.

### *Item Selection, and Initial and Final Version*

Items were selected according to the following three criteria: (1) items for which more than half of the participants selected the labeled emotion; (2) item discrimination indicating the extent to which success on an item corresponds to success on the whole test of more than 0.2 (or more than 0.1 if the item difficulty was close to 0.9 or 0.1); (3) an equal number of items on each emotion type and same gender of the speaker. The initial version was created using eight items for each of the seven emotions assessed. All items relating to happiness were deleted because the average selection rate was 21%, similar to the mean correct rate in a previous study (Gold et al., 2012), and it seems like the happy tone of voice could be limited by the neutral content. The final version of the test was validated with a cross-validation cohort of 82 healthy subjects. The final version of the VER-42 was reduced to six items for each emotion, with the speaker's gender being equal (Table 2).

TABLE 2
Characteristics of the voice on each emotion

| Emotion | *N* of items | Speaker's gender (male/female) | Mean intensity of actor's play (%)[a] |
|---|---|---|---|
| Neutral | 6 | 2/4 | 1.50 (50.0) [b] |
| Anger | 6 | 4/2 | 2.67 (88.9) |
| Sadness | 6 | 3/3 | 2.00 (66.7) |
| Fear | 6 | 2/4 | 1.67 (55.6) |
| Surprise | 6 | 3/3 | 1.83 (61.1) |
| Sarcasm | 6 | 4/2 | 1.83 (61.1) |
| Disgust | 6 | 3/3 | 2.17 (72.2) |
| Total | 42 | 21/21 | 2.03 (67.6) [c] |

*Note*. [a] three intensity levels (*high* = 3, *medium* = 2, and *low* = 1), percentage in parentheses was mean intensity divided by 3 and then times 100; [b] voice intensity; [c] except neutral.

### Data Analyses

For continuous variables, *t*-tests were used to compare differences between groups. For categorical variables, differences between groups were compared using $\chi^2$ tests. Item discrimination used the point bi-serial correlation coefficient for the pretest sample. Kuder and Richardson Formula 20 (KR20) was used to evaluate internal consistency, which was more suitable for dichotomous items than Cronbach's α. Test-retest reliability was computed using Pearson's *r* as the correlation coefficient in the healthy group. Utility paired-samples *t*-tests with Cohen's *d* and intraclass correlation coefficient (ICC) were assessed to evaluate longitudinal stability in the healthy group. Correlations between VER-42 and functional outcomes/symptoms were computed using Pearson's *r* as the correlation coefficient for each group. Bonferroni correction was conducted for multiple correlations between VER-42 scores with MCCB seven domains scores. Independent-samples *t*-test with Cohen's *d* was used to examine group differences in the total score. The vocal emotion recognition score or emotion intensity (only when the emotion type was correct) was assessed for each emotion using repeated-measures analysis of variance (ANOVA), with emotion as the within-subject factor and diagnostic group as a between-subject factor, and years of education was considered as covariate. A $p$-value < .05 was considered statistically significant. Statistical analyses were performed in the IBM SPSS Statistics 20.0 program and the plot was prepared with the R 3.5.3 "ggplot2 3.1.0" package (https://www.r-project.org/).

## RESULTS

### Reliability and Longitudinal Stability of the VER-42

The mean item difficulty and discrimination for each emotion in the healthy group ranged from .52 to .96 and from .24 to .37, respectively (Table 3). The internal consistency of the VER-42 was .77, while that of the seven subscales varied between .60 and .74. The split-half coefficient of the VER-42 was .83, while that of the seven subscales ranged from .48 to .62. The test-retest reliability was .69, and the ICC was .80 (Table 4).

TABLE 3
Item difficulty and discrimination on each emotion in the healthy group

| Emotion | Item | Difficulty | Discrimination |
|---|---|---|---|
| Neutral | 1 | 1.00 | – |
| | 2 | .99 | .43 |
| | 3 | .99 | .29 |
| | 4 | .94 | .25 |
| | 5 | .93 | .25 |
| | 6 | .90 | .23 |
| | Mean | .96 | .24 |
| Sadness | 1 | .90 | .37 |
| | 2 | .89 | .27 |
| | 3 | .88 | .13 |
| | 4 | .78 | .32 |
| | 5 | .74 | .50 |
| | 6 | .43 | .21 |
| | Mean | .77 | .30 |

(table 3 continues)

Table 3 (continued)

| Emotion | Item | Difficulty | Discrimination |
|---|---|---|---|
| Surprise | 1 | .78 | .35 |
| | 2 | .72 | .22 |
| | 3 | .65 | .47 |
| | 4 | .54 | .44 |
| | 5 | .52 | .43 |
| | 6 | .52 | .31 |
| | Mean | .62 | .37 |
| Disgust | 1 | .63 | .41 |
| | 2 | .63 | .37 |
| | 3 | .54 | .33 |
| | 4 | .51 | .37 |
| | 5 | .45 | .22 |
| | 6 | .33 | .28 |
| | Mean | .52 | .33 |
| Anger | 1 | .96 | .38 |
| | 2 | .95 | .40 |
| | 3 | .91 | .09 |
| | 4 | .87 | .40 |
| | 5 | .74 | .28 |
| | 6 | .60 | .33 |
| | Mean | .84 | .31 |
| Fear | 1 | .78 | .20 |
| | 2 | .71 | .44 |
| | 3 | .67 | .35 |
| | 4 | .65 | .36 |
| | 5 | .60 | .27 |
| | 6 | .54 | .22 |
| | Mean | .66 | .31 |
| Sarcasm | 1 | .56 | .32 |
| | 2 | .55 | .47 |
| | 3 | .54 | .34 |
| | 4 | .51 | .20 |
| | 5 | .50 | .30 |
| | 6 | .44 | .28 |
| | Mean | .52 | .32 |
| Total | Mean | .70 | .31 |

TABLE 4
Test-retest coefficients of voice emotion recognition test in the healthy group

| Emotion | *N* | First time (T1) | | Second time (T2) | | Difference (T2-T1) | *t* | *p* | *r* | ICC |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Mean | *SD* | Mean | *SD* | | | | | |
| Neutral | 81 | 5.75 | 0.64 | 5.46 | 1.04 | –.30 | –3.22 | .002 | .60 | .70 |
| Anger | 81 | 5.02 | 1.12 | 4.84 | 1.24 | –.19 | –1.46 | .148 | .54 | .69 |
| Sadness | 81 | 4.60 | 1.28 | 4.43 | 1.51 | –.17 | –1.25 | .214 | .61 | .75 |
| Fear | 81 | 3.95 | 1.67 | 3.77 | 1.98 | –.19 | –.99 | .325 | .59 | .73 |
| Surprise | 81 | 3.72 | 1.64 | 3.91 | 1.86 | .20 | 1.00 | .318 | .49 | .66 |
| Sarcasm | 81 | 3.09 | 1.74 | 3.43 | 1.70 | .35 | 2.21 | .030 | .67 | .80 |
| Disgust | 81 | 3.09 | 1.70 | 2.94 | 1.86 | –.15 | –.73 | .467 | .48 | .64 |
| Total | 81 | 29.22 | 5.48 | 28.78 | 6.97 | –.44 | –.78 | .435 | .69 | .80 |

Note. ICC = intraclass correlation coefficient.

## Relationship to Functional Outcome and Symptoms

In the healthy group, correlations of VER-42 scores with social (MSEIT) and neurocognitive tasks (MCCB total score) were significant — $r(44) = .39$, $p < .05$ and $r(44) = .41$, $p < .01$, respectively —, and the VER-42 score was mainly related to reasoning and problem solving, visual learning, and the speed of processing in MCCB cognitive domains — $r(46) = .31$-$.33$, $p < .05$. Note that all these correlations would survive Bonferroni corrections for multiple (seven) correlations. Correlations with TEPS and TAS-20 were significant — $r(81) = .24$, $p < .05$ and $r(82) = –.40$, $p < .01$, respectively; correlations with FERT was significant — $r(77) = .55$, $p < .01$. In patients, correlations of VER-42 scores with social (MSEIT) and neurocognitive tasks (MCCB total score) were significant — $r(155) = .27$, $p < .01$ and $r(154) = .49$, $p < .01$, respectively — and the VER-42 score was mainly related to reasoning and problem solving, visual learning, verbal learning, attention/vigilance, working memory and the speed of processing in MCCB cognitive domains — $r(156) = .31$~$.47$, $p < .01$. Note that all these correlations would survive Bonferroni corrections for multiple (seven) correlations. Correlations with TEPS and TAS-20 were significant — $r(134) = .11$, $p > .05$ and $r(132) = –.33$, $p < .01$, respectively; correlation with FERT was significant — $r(161) = .70$, $p < .01$. Negative correlations with PANSS negative score was significant — $r(129) = –.20$, $p < .05$ —, and was still significant — $r(121) = –.19$, $p < .05$ — after controlling for MCCB total score. All others were not significant ($p > .05$).

## Group Differences

Patients were less accurate in identifying emotion type than the healthy group, as determined by the VER-42 total score — 22.81 ± 7.80 vs. 29.27 ± 5.46; $t(218) = 7.49$, $p < .001$, effect size = .91 —, and the significance remains unchanged, $F(1, 211) = 43.22$, $p < .001$, after controlling for years of education. The repeated-measures ANOVA across emotions showed a significant group × emotion interaction, $F(6, 1446) = 3.21$, $p = .005$, for accuracy, suggesting statistically different deficits across emotions (Figure 1), and the significance remains unchanged, $F(5, 1159) = 3.38$, $p = .004$, after controlling for years of education. A further simple-effects test showed patients were less accurate than healthy individuals across all emotions ($p < .05$), and the maximal/minimal differences were obtained with sadness and sarcasm, respectively. In terms of the intensity of emotion (excluding neutral intensity), the group × emotion interaction, $F(5,745) = 4.34$, $p = .001$, was significant, and the significance remains unchanged, $F(5, 617) = 5.44$, $p < .001$, after controlling for years of education. Further simple-effects tests indicated that patients showed significant deficits in the recognition of the intensity of anger ($p < .001$) and sadness ($p = .047$), with trends also for impairments in the recognition of the intensity of disgust ($p = .090$).

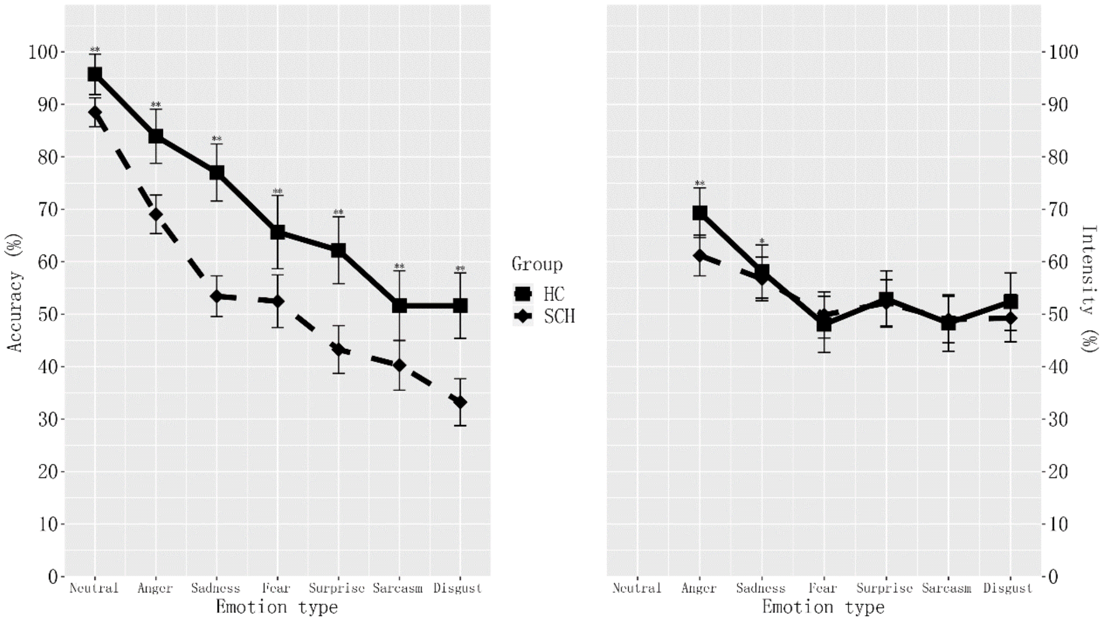


FIGURE 1

Group differences in identifying emotion type and the intensity of emotion

*Note*. HC = healthy control; SCH = schizophrenia. Left = accuracy in identifying emotion type; Right = recognition of intensity. The horizontal axis represents seven different emotions.

## DISCUSSION

In this study, we developed a new instrument for the assessment of vocal emotion recognition ability and evaluated its validity in a group of individuals suffering from schizophrenia or schizoaffective disorder. The internal reliability, test-retest reliability, and validity of the VER-42 were acceptable. It has appropriate difficulty and good discrimination to optimize the distinction of the vocal emotion recognition ability. The VER-42 total score correlated with scores on the MCCB, especially for speed of processing, reasoning, and memory. These results were in general accord with the concept that vocal affect recognition is correlated with cognitive functions, such as speed of processing (Gold et al., 2012). It seemed that an interference of cognitive deficits in patients' early emotional scene processing (Trotti et al., 2021), which further affected vocal emotional recognition. It also correlated with scores on the MSEIT, TEPS, and TAS, some proxy measures for social cognition or function, which agreed with the findings of some previous studies on social function or functional capacity (Giannitelli et al., 2015; Irani et al., 2012; Pinkham et al., 2018). The VER-42 total score was moderately related to facial emotion recognition. It has shown that VER-42 was similar to but different from facial emotion recognition, and the same as previous results (Kerr & Neale, 1993). Moreover, the VER-42 total score was associated with negative symptoms of patients suffering from schizophrenia. There might be some connection between emotion recognition deficits and negative symptoms in schizophrenia (Jung et al., 2021). More diverse cross-linguistic datasets could be required to confirm because these associations might vary across languages (Parola et al., 2023).

The vocal emotion recognition deficit in patients suffering from schizophrenia was strongly confirmed, as previous studies had indicated, and revealed a large effect size (Barkl et al., 2014; Gong et al., 2021; Savla et al., 2013). It showed sensitivity to group differences, and can thus be used clinically. Moreover, this impairment of nonverbal perception could further lead to a loss of social expression (Parola et al., 2023). Patients were less accurate than the healthy controls in recognizing all emotions, and the deficits ranged pattern was similar to that described by Bonfils et al. (2019), Giannitelli et al. (2015), Gong et al. (2021), and was slightly different from that described in other previous studies, which found only deficits in the vocal emotion recognition of sadness, fear, disgust, and neutral (Gold et al., 2012), and in facial emotion recognition of fear, disgust (Barkl et al., 2014; Kohler et al., 2003). This suggests that different modal stimuli could have different mechanisms (Bonfils et al., 2019; Giannitelli et al., 2015; Schirmer, 2018; Young et al., 2020). It presented the smallest difference in sarcasm, which is less effective than previous literature reports (Kantrowitz et al., 2014), possibly because of a lack of context, such as additional information about the speaker’s true beliefs or intentions. Furthermore, this impairment in individuals suffering from schizophrenia existed not only in identifying emotional types but also in identifying emotional intensity. Patients showed more deficits in vocal perception of intensity of anger and sadness than other emotions in vocal stimuli. This suggests that patients suffering from schizophrenia might have more difficulty in identifying anger and sadness. Patients suffering from schizophrenia could experience a reduced intensity of anger expression compared to healthy controls, and this might cause differences in behavioral response (Navalón et al., 2022).

## Limitations

Firstly, the sample size of the study might not be large enough. Secondly, there was a methodological drawback as happy expressions were the only positive universal emotion, but this seemed to be limited by vocally neutral content. The length of stimulus presentation, for example, the number of replays, and the reaction time could also be considered limitations. Also, more general measures of social function could be implemented to address the validity of the VER-42. Fourth, the stability of the VER-42 in clinical patients, especially during an acute episode, needs to be verified. Moreover, other variables might also influence the VER-42 score, such as duration of illness, drugs used and their dosage, and some sociodemographic factors. Further studies are desirable.

## NOTE

1. Hongzhen Fan and Ronghuan Jiang contributed equally to this work.

## Fundings

This work was supported by the Research Project for the Application of Clinical Characteristics in Capital (Z141107002514016), Beijing Hospitals Authority Youth Programme (QML20202001), and the Natural Science Foundation of Beijing Municipality (7162087).

## Acknowledgments

We would like to thank all of the participants and actors who took part in the study.

## References

Bänziger, T., Scherer, K. R., Hall, J. A., & Rosenthal, R. (2011). Introducing the MiniPONS: A short multichannel version of the Profile of Nonverbal Sensitivity (PONS). *Journal of Nonverbal Behavior*, *35*(3), 189-204. https://doi.org/10.1007/s10919-011-0108-3

Bagby, R. M., Parker, J. D., & Taylor, G. J. (1994). The twenty-item Toronto Alexithymia Scale-I. Item selection and cross-validation of the factor structure. *Journal of Psychosomatic Research*, *38*(1), 23-32. https://doi.org/10.1016/0022-3999(94)90005-1

Barkl, S. J., Lah, S., Harris, A. W., & Williams, L. M. (2014). Facial emotion identification in early-onset and first-episode psychosis: A systematic review with meta-analysis. *Schizophrenia Research*, *159*(1), 62-69. https://doi.org/10.1016/j.schres.2014.07.049

Bellack, A. S., Blanchard, J. J., & Mueser, K. T. (1996). Cue availability and affect perception in schizophrenia. *Schizophrenia Bulletin*, *22*(3), 535-544. https://doi.org/10.1093/schbul/22.3.535

Bonfils, K. A., Ventura, J., Subotnik, K. L., & Nuechterlein, K. H. (2019). Affective prosody and facial emotion recognition in first-episode schizophrenia: Associations with functioning & symptoms. *Schizophrenia Research: Cognition*, *18*, Article 100153. https://doi.org/10.1016/j.scog.2019.100153

Bryson, G., Bell, M., & Lysaker, P. (1997). Affect recognition in schizophrenia: A function of global impairment or a specific cognitive deficit. *Psychiatry Research*, *71*(2), 105-113. https://doi.org/10.1016/S0165-1781(97)00050-4

Chan, R. C., Li, H., Cheung, E. F., & Gong, Q. Y. (2010). Impaired facial emotion perception in schizophrenia: A meta-analysis. *Psychiatry Research*, *178*(2), 381-390. https://doi.org/10.1016/j.psychres.2009.03.035

Chan, R. C., Wang, Y., Huang, J., Shi, Y., Wang, Y., Hong, X., Ma, Z., Li, Z., Lai, M. K., & Kring, A. M. (2010). Anticipatory and consummatory components of the experience of pleasure in schizophrenia: Cross-cultural validation and extension. *Psychiatry Research*, *175*(1-2), 181-183. https://doi.org/10.1016/j.psychres.2009.01.020

Ekman, P. E., & Frieser, W. (1975). *Pictures of facial affect*. Consulting Psychologists Press.

Giannitelli, M., Xavier, J., Francois, A., Bodeau, N., Laurent, C., Cohen, D., & Chaby, L. (2015). Facial, vocal and cross-modal emotion processing in early-onset schizophrenia spectrum disorders. *Schizophrenia Research*, *168*(1-2), 252-259. https://doi.org/10.1016/j.schres.2015.07.039

Gold, R., Butler, P., Revheim, N., Leitman, D. I., Hansen, J. A., Gur, R. C., Kantrowitz, J. T., Laukka, P., Juslin, P. N., Silipo, G. S., & Javitt, D. C. (2012). Auditory emotion recognition impairments in schizophrenia: Relationship to acoustic features and cognition. *The American Journal of Psychiatry*, *169*(4), 424-432. https://doi.org/10.1176/appi.ajp.2011.11081230

Gong, B., Li, Q., Zhao, Y., & Wu, C. (2021). Auditory emotion recognition deficits in schizophrenia: A systematic review and meta-analysis. *Asian Journal of Psychiatry*, *65*, Article 102820. https://doi.org/10.1016/j.ajp.2021.102820

Green, M. F., Horan, W. P., & Lee, J. (2015). Social cognition in schizophrenia. *Nature Reviews Neuroscience*, *16*(10), 620-631. https://doi.org/10.1038/nrn4005

Irani, F., Seligman, S., Kamath, V., Kohler, C., & Gur, R. C. (2012). A meta-analysis of emotion perception and functional outcomes in schizophrenia. *Schizophrenia Research*, *137*(1-3), 203-211. https://doi.org/10.1016/j.schres.2012.01.023

Jung, S., Kim, J. H., Kang, N. O., Sung, G., Ko, Y. G., Bang, M., Park, C. I., & Lee, S. H. (2021). Fusiform gyrus volume reduction associated with impaired facial expressed emotion recognition and emotional intensity recognition in patients with schizophrenia spectrum psychosis. *Psychiatry Research: Neuroimaging*, *307*, Article 111226. https://doi.org/10.1016/j.pscychresns.2020.111226

Juslin, P. N., & Laukka, P. (2001). Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. *Emotion*, *1*(4), 381-412. https://doi.org/10.1037/1528-3542.1.4.381

Kantrowitz, J. T., Hoptman, M. J., Leitman, D. I., Silipo, G., & Javitt, D. C. (2014). The 5% difference: Early sensory processing predicts sarcasm perception in schizophrenia and schizo-affective disorder. *Psychological Medicine*, *44*(1), 25-36. https://doi.org/10.1017/s0033291713000834

Kay, S. R., Fiszbein, A., & Opler, L. A. (1987). The Positive and Negative Syndrome Scale (PANSS) for schizophrenia. *Schizophrenia Bulletin*, *13*(2), 261-276. https://doi.org/10.1093/schbul/13.2.261

Kerr, S. L., & Neale, J. M. (1993). Emotion perception in schizophrenia: Specific deficit or further evidence of generalized poor performance? *Journal of Abnormal Psychology*, *102*(2), 312-318. https://doi.org/10.1037//0021-843x.102.2.312

Kohler, C. G., Turner, T. H., Bilker, W. B., Brensinger, C. M., Siegel, S. J., Kanes, S. J., Gur, R. E., & Gur, R. C. (2003). Facial emotion recognition in schizophrenia: Intensity effects and error pattern. *The American Journal of Psychiatry*, *160*(10), 1768-1774. https://doi.org/10.1176/appi.ajp.160.10.1768

Kohler, C. G., Walker, J. B., Martin, E. A., Healey, K. M., & Moberg, P. J. (2010). Facial emotion perception in schizophrenia: A meta-analytic review. *Schizophrenia Bulletin*, *36*(5), 1009-1019. https://doi.org/10.1093/schbul/sbn192

Lassalle, A., Pigat, D., O'Reilly, H., Berggen, S., Fridenson-Hayo, S., Tal, S., Elfström, S., Råde, A., Golan, O., Bölte, S., Baron-Cohen, S., & Lundqvist, D. (2019). The EU-emotion voice database. *Behavior Research Methods*, *51*(2), 493-506. https://doi.org/10.3758/s13428-018-1048-1

Navalón, P., Sahuquillo-Leal, R., Moreno-Giménez, A., Salmerón, L., Benavent, P., Sierra, P., Cañada, Y., Cañada-Martínez, A., Berk, M., & García-Blanco, A. (2022). Attentional engagement and inhibitory control according to positive and negative symptoms in schizophrenia: An emotional antisaccade task. *Schizophrenia Research*, *239*, 142-150. https://doi.org/10.1016/j.schres.2021.11.044

Nuechterlein, K. H., Green, M. F., Kern, R. S., Baade, L. E., Barch, D. M., Cohen, J. D., Essock, S., Fenton, W. S., Frese, F. J., 3rd, Gold, J. M., Goldberg, T., Heaton, R. K., Keefe, R. S., Kraemer, H., Mesholam-Gately, R., Seidman, L. J., Stover, E., Weinberger, D. R., Young, A. S., Zalcman, S., & Marder, S. R. (2008). The MATRICS Consensus Cognitive Battery, Part 1: Test selection, reliability, and validity. *The American Journal of Psychiatry*, *165*(2), 203-213. https://doi.org/10.1176/appi.ajp.2007.07010042

Parola, A., Simonsen, A., Lin, J. M., Zhou, Y., Wang, H., Ubukata, S., Koelkebeck, K., Bliksted, V., & Fusaroli, R. (2023). Voice patterns as markers of schizophrenia: Building a cumulative generalizable approach via a cross-linguistic and meta-analysis based investigation. *Schizophrenia Bulletin*, *49*(Suppl_2), S125-S141. https://doi.org/10.1093/schbul/sbac128

Pawełczyk, A., Łojek, E., Radek, M., & Pawełczyk, T. (2021). Prosodic deficits and interpersonal difficulties in patients with schizophrenia. *Psychiatry Research*, *306*, Article 114244. https://doi.org/10.1016/j.psychres.2021.114244

Pinkham, A. E., Harvey, P. D., & Penn, D. L. (2018). Social cognition psychometric evaluation: Results of the final validation study. *Schizophrenia Bulletin*, *44*(4), 737-748. https://doi.org/10.1093/schbul/sbx117

Savla, G. N., Vella, L., Armstrong, C. C., Penn, D. L., & Twamley, E. W. (2013). Deficits in domains of social cognition in schizophrenia: A meta-analysis of the empirical evidence. *Schizophrenia Bulletin*, *39*(5), 979-992. https://doi.org/10.1093/schbul/sbs080

Schirmer, A. (2018). Is the voice an auditory face? An ALE meta-analysis comparing vocal and facial emotion processing. *Social Cognitive and Affective Neuroscience*, *13*(1), 1-13. https://doi.org/10.1093/scan/nsx142

Silver, H., Shlomo, N., Turner, T., & Gur, R. C. (2002). Perception of happy and sad facial expressions in chronic schizophrenia: Evidence for two evaluative systems. *Schizophrenia Research*, *55*(1-2), 171-177. https://doi.org/10.1016/s0920-9964(01)00208-0

Trotti, R. L., Abdelmageed, S., Parker, D. A., Sabatinelli, D., Tamminga, C. A., Gershon, E. S., Keedy, S. K., Keshavan, M. S., Pearlson, G. D., Sweeney, J. A., McDowell, J. E., & Clementz, B. A. (2021). Neural processing of repeated emotional scenes in schizophrenia, schizoaffective disorder, and bipolar disorder. *Schizophrenia Bulletin*, *47*(5), 1473-1481. https://doi.org/10.1093/schbul/sbab018

Yanli, Z., Hongzhen, F., Yuanyuan, Z., Wei, Q., Dong, L., Yajun, L., & Shuping, T. (2020). Chinese facial emotion recognition test: Reliability and validity. *Chinese Journal of Psychiatry*, *6*(53), 512-519.

Young, A. W., Frühholz, S., & Schweinberger, S. R. (2020). Face and voice perception: Understanding commonalities and differences. *Trends in Cognitive Sciences*, *24*(5), 398-410. https://doi.org/10.1016/j.tics.2020.02.001