Effects of sensory dominance and gender differences on impaired emotion perception in schizophrenic patients

Yi Lin1, Chuoran Li2, Yueqi Chen1, Jiaqi Zhang1, Hongwei Ding1/
1Speech-Language-Hearing Center, School of Foreign Languages, Shanghai Jiao Tong University, Shanghai, China
2Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, Shanghai, China
carol.y.lin@sjtu.edu.cn hwding@sjtu.edu.cn

Abstract

Patients with schizophrenia have been repeatedly reported to show dysfunctions in emotion perception. However, it remains unclear how contextual (e.g., communication channels) and individual (e.g., gender differences) factors interact to modulate their perceptual performances. The present study examined how emotions conveyed through three different sensory channels (i.e., face, prosody, and semantics) were perceived by schizophrenic patients and healthy controls of the two genders. Fourteen (seven women and seven men) schizophrenic patients and sixteen healthy controls (eight women and eight men) were asked to identify the emotion displayed visually through facial expressions, or auditorily through prosody or semantics in a fixed-choice format. The integration of accuracy and reaction time data revealed that schizophrenic patients showed impaired emotion perception compared to healthy controls. For both patients and controls, emotional faces gained most perceptual salience among the three communicative channels, but patients were more biased towards emotional semantics whereas controls were more dominated by emotional prosody. Additionally, the preserved abilities in semantic processing were more pronounced for male patients compared with their female counterparts. To sum up, schizophrenic patients demonstrate impaired nonverbal emotion perception irrespective of perceivers’ gender, while their intact abilities in verbal emotion perception are restrictive to males.

Index Terms: schizophrenic patients, channel dominance, gender differences, emotion perception, prosody, semantics, face

1. Introduction

Schizophrenia is a serious psychiatric disorder that affects one’s thinking, feelings and behaviors. It is generally agreed that patients with schizophrenia show reduced ability in emotion processing [1-4]. Emotion deficits in schizophrenics can lead to vulnerability and difficulty in various life domains such as interpersonal relationships and communication as well as social cognition and adaptation [5].

During daily interactions, emotional information can be conveyed through various verbal and nonverbal channels, such as semantic content, prosody, facial expressions, body movements etc. These different sensory channels are not always equivalent in emotional communication and a certain channel may hold more perceptual advantages over the others. Some previous studies identified the perceptual advantages of nonverbal cues (i.e., facial expressions and prosody) among healthy individuals in unisensory and multisensory emotional perception settings [6, 7]. Schizophrenic patients, by contrast, have been repeatedly reported to demonstrate deficits in processing facial [8] and prosodic [9, 10] emotional stimuli [3], even though a nonverbal dominance effect can still be observed among the patient group [11]. However, there is a lack of consensus regarding whether there is a visual facial or auditory prosodic dominance effect in emotion perception among schizophrenics [12]. In addition, relatively few studies have compared the role of verbal and nonverbal channels in emotion processing between patients and controls.

Decoders’ gender is another important factor that can affect emotion perception. While a wealth of research revealed gender differences for emotion perception in healthy participants with women outperforming men [13-15], relatively fewer studies examined the role of gender in perceiving emotional stimuli among mentally impaired patients including schizophrenics [16]. Sparse evidence from schizophrenic individuals suggests that the advantage of emotional processing in healthy women is preserved in patients albeit with overall poorer performances relative to controls [17]. Since the ability to decode emotional signals serves as an essential component of social communication, more severe impairments among male schizophrenic patients may explain their worse outcomes in social functioning compared to their female counterparts [18].

Despite considerable research on the effects of channel dominance and gender differences on emotion perception, the two domains have seldom been associated to examine emotion identification performances among mentally impaired patients. It remains to be explored how channel as a contextual factor and gender as an individual characteristic interact in affecting emotion perception among schizophrenics. The present study aims to investigate how emotional cues transmitted through facial, prosodic and semantic channels are processed by female and male schizophrenic and healthy participants. We first predicted that schizophrenic patients would produce worse performances in emotion identification compared to healthy controls (hypothesis 1) [1, 4]. We also hypothesized that both schizophrenics and controls would be more biased towards facial and prosodic cues in emotion perception though patients tended to show reduced sensitivity to the nonverbal dominance effect (hypothesis 2) [11, 29]. Furthermore, women’s superior ability in emotion perception compared to men would be preserved in schizophrenics (hypothesis 3) [17]. We hope to deepen the understanding of how theoretical accounts of channel dominance and gender differences converge in shaping emotion perception among healthy and clinical populations. The present study also serves to provide implications for the clinical diagnosis and
intervention of some gender-related psychiatric affective disorders.

2. Methods

2.1. Participants

Fourteen (7 women and 7 men) schizophrenic patients and sixteen healthy controls (8 women and 8 men) participated in this study. Schizophrenic patients were recruited from the inpatient units of Shanghai Mental Health Center affiliated to Shanghai Jiao Tong University (SJTU). Diagnosis of schizophrenia was confirmed by trained psychiatrists based on International Statistical Classification of Disease (ICD)-10 criteria [19]. Symptom severity was assessed with the Positive and Negative Syndrome Scale (PANSS) [20] and the Scale for the Assessment of Negative Symptoms (SANS) [21]. No significant gender difference was found for patients in age at the onset of the first psychosis, the number of hospitalizations, and the duration of illness (all \( p > .05 \)). Healthy participants were matched to patients based on age, gender and years of education. All participants had normal or corrected to normal vision and hearing, spoke Mandarin Chinese as native language, reported no history of alcohol abuse or history of neurological illness (e.g., stroke), and had a sufficient level of cognitive functioning to complete the required tasks. For healthy controls, an additional included criterion was no reported history of psychiatric illness.

Table 1 shows the demographic and clinical variables related to all study participants. The current study was approved by the Institutional Review Boards (IRB) of School of Foreign Languages and Shanghai Mental Health Center at SJTU. The experimental procedures were fully explained to each participant, and all participants gave written informed consent at the study onset.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Schizophrenic patients (n = 14)</th>
<th>Healthy controls (n = 16)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>29.14 (8.23)</td>
<td>30.88 (9.90)</td>
</tr>
<tr>
<td>Gender (female/male)</td>
<td>7/7</td>
<td>8/8</td>
</tr>
<tr>
<td>Education</td>
<td>13.36 (2.50)</td>
<td>15.16 (2.85)</td>
</tr>
<tr>
<td>Illness duration (years)</td>
<td>3.86 (1.81)</td>
<td>NA</td>
</tr>
<tr>
<td>Age of first psychosis</td>
<td>25.21 (7.49)</td>
<td>NA</td>
</tr>
<tr>
<td>Number of hospitalizations</td>
<td>1.36 (48)</td>
<td>NA</td>
</tr>
<tr>
<td>PANSS total</td>
<td>67.86 (11.63)</td>
<td>NA</td>
</tr>
<tr>
<td>Positive</td>
<td>13.36 (2.87)</td>
<td>NA</td>
</tr>
<tr>
<td>Negative</td>
<td>21.36 (6.37)</td>
<td>NA</td>
</tr>
<tr>
<td>General</td>
<td>33.14 (5.26)</td>
<td>NA</td>
</tr>
<tr>
<td>SANS total</td>
<td>31.43 (17.32)</td>
<td>NA</td>
</tr>
</tbody>
</table>

Notes. All values represent mean ± SD. NA = not applicable.

2.2. Stimuli

Each stimulus expressed one of the three basic emotions (i.e., anger, happiness and sadness) [22] or neutrality through the visual facial channel, or the auditory prosodic or semantic channels. The visual facial stimuli were 64 black-and-white photographs of 16 Chinese actors (eight women and eight men) portraying emotional (i.e., angry, happy and sad) or neutral facial expressions. These stimuli were selected from the Chinese Affective Picture System [23] based on the identification accuracy of emotional valence and ratings of emotional intensity in an earlier norming study, which will be introduced in the next paragraph. The auditory prosodic and semantic stimuli were produced by four Chinese professional speakers (two women and two men), and digitized into a MacBook Pro with AVID Mbox Mini at a sampling rate of 44.100 kHz with a 16-bit resolution. For the prosodic stimuli, the speakers enunciated 64 semantically neutral disyllabic concrete words (e.g., 肥皂 (soap in Chinese), 佛手 (zebra in Chinese)) in angry, happy, sad and neutral prosody. For the semantic stimuli, the speakers uttered 64 disyllabic words suggesting angry (e.g., 哀痛 (wrathful in Chinese), happy (e.g., 兴奋 (excited in Chinese)), sad (e.g., 悲伤 (sorrowful in Chinese)) and neutral (e.g., 平淡 (flat in Chinese)) verbal content in a neutral tone of voice. There was an equal number of stimuli in each of the three sensory channels in terms of emotional category (16 stimuli for each category) and encoders’ gender (8 stimuli for each gender per category). Please see our recently published article [6] for more information about the duration and f0 measures of the prosodic and semantic stimuli used in the current study.

All stimuli used in the current study were applied in our previous investigations on emotional perception among healthy populations [6, 24]. We administered a norming study for stimulus validation, in which 23 Chinese university students (11 females and 12 males, mean age ± SD: 23.70 ± 2.59) who did not participate in the current experiment were invited to identify the emotional category and rate the emotional intensity of each visual or auditory stimulus. The selected stimuli received over 90% identification consistency for emotional categories. All emotional (i.e., happy, sad, and angry) stimuli received an average rating above 3 for emotion intensity on a 7-point Likert scale (0 = not intense, 6 = very intense), and neutral stimuli had an intensity score around the midpoint on the scale. Table 2 indicates the identification accuracy of emotional category and rating of emotional intensity for the included stimuli in the present experiment.

<table>
<thead>
<tr>
<th>Stimulus type</th>
<th>Emotion category</th>
<th>Identification of emotional category</th>
<th>Rating of emotional intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean ± SD</td>
<td>Mean ± SD</td>
</tr>
<tr>
<td>Face</td>
<td>Angry</td>
<td>93.29% ± 2.03%</td>
<td>4.50 ± 0.60</td>
</tr>
<tr>
<td></td>
<td>Happy</td>
<td>93.06% ± 1.96%</td>
<td>4.52 ± 0.62</td>
</tr>
<tr>
<td></td>
<td>Neutral</td>
<td>92.13% ± 1.31%</td>
<td>3.41 ± 0.28</td>
</tr>
<tr>
<td></td>
<td>Sad</td>
<td>92.59% ± 1.73%</td>
<td>4.03 ± 0.50</td>
</tr>
<tr>
<td>Prosody</td>
<td>Angry</td>
<td>95.41% ± 4.22%</td>
<td>4.77 ± 0.32</td>
</tr>
<tr>
<td></td>
<td>Happy</td>
<td>92.51% ± 3.49%</td>
<td>4.23 ± 0.38</td>
</tr>
<tr>
<td></td>
<td>Neutral</td>
<td>91.55% ± 2.70%</td>
<td>3.28 ± 0.22</td>
</tr>
<tr>
<td></td>
<td>Sad</td>
<td>93.72% ± 3.89%</td>
<td>4.40 ± 0.24</td>
</tr>
<tr>
<td>Semantics</td>
<td>Angry</td>
<td>96.74% ± 3.26%</td>
<td>5.05 ± 0.26</td>
</tr>
<tr>
<td></td>
<td>Happy</td>
<td>96.74% ± 3.61%</td>
<td>4.59 ± 0.70</td>
</tr>
<tr>
<td></td>
<td>Neutral</td>
<td>95.92% ± 3.25%</td>
<td>3.05 ± 0.17</td>
</tr>
</tbody>
</table>

Note. Participants identified the emotional category of the stimuli, and rated the intensity on a 7-point scale (0 = not intense, 6 = very intense).

2.3. Procedure

The study was administered in a sound-attenuated booth using E-prime (Version 2.0.8.22) for stimulus presentation. The experiment consisted of three counterbalanced sessions with altogether 192 trials, namely emotional face identification,
emotional prosody identification, and emotional semantics identification. Visual facial stimuli were shown in the center of the LCD screen with a constant white background, and auditory prosodic and semantic stimuli were displayed binaurally through Sennheiser HD280 PRO headphones at 70 dB SPL. Each trial began with a fixation cross in the center of the computer screen for 1100 milliseconds (ms). Then emotional signals were presented from one of the facial, prosodic, or semantic channels, at the onset of which participants were asked to indicate the emotion by pressing the emotion-coded keys (“h” for happy, “b” for sad, “a” for angry, and “u” for neutral) as quickly as possible without sacrificing accuracy. After the response was made, a blank screen was displayed for 1000 ms as an inter-trial interval. The presentation order of trials in each session was randomized across participants. Participants entered the experiment after familiarizing themselves with the procedure by running eight practice trials in each session with at least 75% accuracy. They were allowed to take a short rest after completing every 32 trials during each session.

2.4. Statistical analyses

We employed a series of linear mixed-effects models in R (version 3.6.1) with the lme4 package for data analysis [25]. Accuracy and reaction time were entered as dependent variables in the linear mixed-effects models respectively. The fixed categorical factors included communication channel (i.e., facial, prosodic, and semantic), decoders’ gender (i.e., female and male), and group (i.e., schizophrenic and healthy), in which the semantic channel, male participants and schizophrenic patients were set as the baseline level respectively. Prosody was set as the baseline level when it was compared with face. Random factors for intercepts contained individual subjects and test items. Tukey’s post hoc tests in the emmeans package [26] were applied for pairwise comparison in case of a significant main effect or interaction. The full models with intercepts, coefficients, and error terms for accuracy and response time analyses are shown as follows:

\[
\text{Accuracy}_i = \beta_0 + (\beta_1 \times \text{communication channel}) + (\beta_2 \times \text{decoders’ gender}) + (\beta_3 \times \text{group}) + (\beta_4 \times \text{communication channel} \times \text{group}) + (\beta_5 \times \text{decoders’ gender} \times \text{group}) + (\beta_6 \times \text{communication channel} \times \text{decoders’ gender} \times \text{group}) + b_{0i} + b_{1i} + \epsilon_{0i}
\]

\[
\text{Reaction time}_i = \beta_0 + (\beta_1 \times \text{communication channel}) + (\beta_2 \times \text{decoders’ gender}) + (\beta_3 \times \text{group}) + (\beta_4 \times \text{communication channel} \times \text{decoders’ gender}) + (\beta_5 \times \text{group} \times \text{communication channel} \times \text{decoders’ gender}) + (\beta_6 \times \text{group} \times \text{communication channel} \times \text{decoders’ gender} \times \text{group}) + b_{0i} + b_{1i} + \epsilon_{0i}
\]

In both models, \(\beta_0\) was the intercept, which represented the predicted value when all other factors were equal to 0, \(\beta_1, \beta_2, \ldots, \beta_7\) stood for the coefficients for communication channel, decoders’ gender, group, and their interactions respectively. These coefficients suggested the extent to which the outcome variable changed relative to a unit of change in the corresponding predictors. Since the primary focus of the present study was to compare emotion perception performances of schizophrenic patients and healthy controls, we only reported the coefficients of the main effect and the highest-level interaction effects that contained the factor of group in the following Results section. The random intercepts were represented as \(b_{0i}\) and \(b_{1i}\), where \(i\) changed according to individual subjects and \(j\) changed based on test items. An error term \(\epsilon\) was also added in the two models to indicate the distance between the predicted outcome and the actual data point (i.e., residual).

3. Results

Linear mixed-effects analyses on accuracy revealed no main effect of gender (\(\chi^2 (1) = .06, p > .05\)), but a main effect of channel (\(\chi^2 (2) = 15.86, p < .001\)) and a main effect of group (\(\chi^2 (1) = 9.09, p = .003\)). Patients with schizophrenia made significantly less accurate responses compared with healthy controls (\(\hat{\beta}_1 = -.07, SE = .02, t = -3.14, p = .002\)). There was a significant interaction between channel and gender (\(\chi^2 (2) = 9.95, p = .007\)), between channel and group (\(\chi^2 (2) = 50.10, p < .001\)), and among channel, gender, and group (\(\chi^2 (2) = 10.72, p = .005\)). Channel and group interaction suggested that for schizophrenic patients, emotional face elicited significantly more accurate responses compared to prosody (\(\hat{\beta}_1 = .13, SE = .02, t = 5.90, p < .001\)) and semantics (\(\hat{\beta}_1 = .06, SE = .02, t = 2.57, p = .028\)), and emotional semantics also triggered more accurate responses than prosody (\(\hat{\beta}_1 = -.07, SE = .02, t = -3.37, p = .002\)). For healthy participants, however, while emotional face also triggered significantly more accurate responses than emotional semantics (\(\hat{\beta}_1 = .06, SE = .02, t = 2.82, p = .013\), there was no significant difference between emotional face and prosody (\(\hat{\beta}_1 = .04, SE = .02, t = 1.64, p = .229\)), and between prosody and semantics (\(\hat{\beta}_1 = .03, SE = .02, t = 1.18, p = .463\)). By parsing out the three-way interaction, we found that patients made significantly less accurate responses to emotional prosody compared with healthy controls for both female (\(\hat{\beta}_1 = .11, SE = .03, t = 3.24, p = .001\)) and male (\(\hat{\beta}_1 = .17, SE = .03, t = 5.19, p < .001\)) participants. No significant difference was found between patients and controls of both genders for emotional face and semantics (all \(p > .05\)). See Figure 1 for mean identification accuracy by the two groups.

![Figure 1: Mean identification accuracy for emotional stimuli in three different communication channels by female and male participants in schizophrenic and healthy groups.](image-url)

In terms of reaction time data, we first excluded responses over two standard deviations from the mean and incorrect responses in each of the schizophrenic (outliers: 6.06%; incorrect responses: 11.17%) and healthy group (outliers: 4.9%; incorrect responses: 4.87%) [27, 28]. Linear mixed-effects models on reaction time data showed no main effect of gender (\(\chi^2 (1) = .27, p > .05\)), but a main effect of channel (\(\chi^2 (2) = 11.36, p = .003\)), and a main effect of group (\(\chi^2 (1) = 8.33, p = .004\)). Patients with schizophrenia made significantly slower responses compared with healthy controls (\(\hat{\beta}_1 = 782, SE = 261, t = 2.99, p = .003\)). There was a significant interaction between channel and group (\(\chi^2 (2) = 105.91, p < .001\)) and among channel, gender and group (\(\chi^2 (2) = 16.99, p < .001\)). The interaction between channel and group revealed no significant difference among the three communication channels for schizophrenic patients (all \(p > .05\)). For healthy controls, however, facial
stimuli elicited faster responses than the prosodic (β̂5 = -285, SE = 84, t = -3.39, p = .002) and semantic (β̂5 = -558, SE = 85, t = -6.62, p < .001) ones, and prosodic stimuli also triggered faster responses compared to verbal content (β̂5 = -273, SE = 85, t = -3.23, p = .004). The three-way interaction among channel, group and gender suggested that schizophrenic patients made significantly slower responses to emotional face compared to healthy controls for both female (β̂7 = -1151, SE = 333, t = -3.45, p < .001) and male (β̂7 = -1092, SE = 326, t = -3.35, p = .001) participants. Similar slower responses to emotional prosody were found in both female (β̂7 = -828, SE = 329, t = -2.52, p = .012) and male (β̂7 = -776, SE = 330, t = -2.35, p = .019) patients with schizophrenia. However, for emotional semantics, only female patients showed slower responses compared to controls (β̂7 = -707, SE = 329, t = -2.15, p = .032), and no significant difference between the patients and controls was found for male participants (β̂7 = -153, SE = 329, t = -.47, p = .642). See Figure 2 for mean reaction time by the two groups.

Figure 2: Mean reaction time for emotional stimuli in three different communication channels by female and male participants in schizophrenic and healthy groups

4. Discussion

The present study compared the identification performances of facial, prosodic and semantic emotional cues between schizophrenics and healthy adults in a Mandarin Chinese background. The integration of accuracy and reaction time data has revealed impaired emotion perception in patients with schizophrenia. For both patients and controls, visual facial expressions dominate over auditory prosodic and semantic cues in emotion perception. During auditory processing, however, while healthy controls are more susceptible to the influence of emotional prosody, patients show more biases towards emotional semantics. Such biases towards semantic content are only established to male rather than female schizophrenics. Together, these findings delineate the effects of communication channels and decoders’ gender as well as their reciprocal interactions on dysfunctional emotion perception in patients with schizophrenia.

The present study demonstrated emotion perception impairments in schizophrenic patients by showing the main effect of group in both accuracy and reaction time measures, which confirms our first hypothesis and coincides with previous literature [3, 5, 8-10]. As predicted in our second hypothesis, schizophrenic patients displayed a different profile of channel dominance effects in emotion perception compared with healthy controls. While nonverbal emotional cues (e.g., facial expressions and prosody) held more communicative advantages over emotional semantics among healthy individuals, this nonverbal dominance effect was less distinctive in patients. Schizophrenics showed reduced sensitivity to emotional facial and prosodic cues and demonstrated a perceptual bias towards verbal content during emotion processing. These findings are consistent with previous investigations showing nonverbal perceptual deficits in schizophrenic patients [11, 29]. Such a diminished nonverbal dominance effect could pose challenges to schizophrenics in complex interpersonal interactions, in which patients often have disturbed interpretations of the overall affective message conveyed simultaneously through multisensory verbal and nonverbal channels.

In our study, both female and male patients demonstrated similar deficits in facial and prosodic emotion perception compared to healthy controls, which aligns with previous studies showing gender-independent nonverbal emotion identification among schizophrenics [30-33]. With regard to verbal emotion identification, however, male patients did not present significantly poorer performances as their female counterparts, though patients of both genders produced slower responses relative to controls. This interesting finding was somewhat beyond our expectation in hypothesis 3 and contrasted with Scholten et al.’s study showing preservation of a female advantage in emotion perception among schizophrenics [17]. The inconsistent result can be related to differences in language and cultural background across studies: while the previous study was conducted in western contexts (i.e., the Netherlands), our study involved patients speaking Mandarin Chinese. One can speculate that in a high-context background, where interlocutors generally rely more on nonverbal cues, male schizophrenics might weight verbal emotional cues more intensely than their female counterparts, which can lead to poorer premorbid and social functioning among male patients.

Some limitations should be taken into consideration when interpreting the findings of the current study. On one hand, schizophrenic patients in the present study exhibited diagnostic homogeneity in a small sample size. Future studies can examine to what extent the current findings can be generalized by involving a larger size and wider range of psychotic patients (e.g., patients with different subtypes of schizophrenia or other mental illnesses like depression, autism and bipolar disorders) [12]. On the other hand, due to the limited number of study samples and experimental trials, we did not investigate different categories of emotional cues and different genders of the emotion encoders separately. However, previous studies have revealed that encoders’ gender and emotional category can result in dissimilar perceptual performances in different communication channels and among female and male decoders [6, 34], which deserves further exploration.

5. Conclusion

The present study investigated emotion identification performances in verbal and nonverbal channels by female and male patients with schizophrenia and healthy controls. Results show that patients exhibited more severe deficits in processing emotional face and prosody and less impairments in understanding verbal content compared to healthy controls. In addition, biased perception towards emotional semantics was more pronounced in male patients. To conclude, emotion perception deficits in schizophrenic patients are susceptible to the influence of communication channels, decoders’ gender and their reciprocal interactions, which has important theoretical and clinical implications.

6. Acknowledgements

This research was funded by the major Program of National Social Science Foundation of China (No. 18ZDA293).
7. References


