Categorization and timbre perception of environmental sounds in schizophrenia

Jean-Arthur Micoulaud-Franchi a,b,c, Mitsuko Aramaki b,c, Adrien Merer d, Michel Cermolacce a,b,c, Solvi Ystad d, Richard Kronland-Martinet d, Jean Vion-Dury a,b,c

Laboratoire de Mécanique et d’Acoustique, LMA-CNRS UPR 7051, 31 Chemin Joseph Aiguier 13402 Marseille cedex 20, France

Institut de Neurosciences Cognitives de la Méditerranée, INCM-CNRS UMR 6193, 31 Chemin Joseph Aiguier 13402 Marseille cedex 20, France

Pôle de Psychiatrie

Accepted 2 March 2011

1. Introduction

Schizophrenia is associated with complex cognitive (Frith, 1987) and perceptual (Holzman, 1972) impairments. An extensive literature analyzed perceptual impairments that were mainly interpreted as consequences of sensory processing dysfunctions (Javitt, 2009a, 2009b). In the auditory modality (Rabinowicz et al., 2000), anomalies were found in tone matching (Javitt et al., 1997), pitch identification (Holcomb et al., 1995), or sound duration discrimination (Todd et al., 2003). Leitman et al. (2008) evaluated the contribution of these underlying dysfunctions, especially tone matching, to emotion recognition from vocal expressions in schizophrenia. Further on, Bach et al. (2009) suggested that perceptual impairments can be related to a more general dysfunction in cortical brain networks according to the Signal Detection Theory (SDT) that allowed the authors to explain the shallower response curves in schizophrenia found by Kee et al. (2006) in an identification task on emotional facial continua (i.e., two facial expressions morphed between two material categories). Results highlighted shallower response curves in subjects with schizophrenia than healthy participants, and are discussed in the framework of Signal Detection Theory and in terms of impaired perception of specific timbre features in schizophrenia.

2. Methods

2.1. Participants

We included 20 chronic inpatients and outpatients from the Department of Psychiatry, Marseille University Hospital, France on the Structured Clinical Interview for DSM-IV Axis I Disorders: Clinical Version (SCID-CV) confirmed the diagnosis of schizophrenia (First et al., 1997). The Positive and Negative Syndrome Scale (Kay et al., 1987) assessed severity of
symptoms. All patients were medicated. CTL constituted 20 healthy subjects screened for any current or lifetime history of a DSM-IV axis I disorder based on MINI interviews (Sheehan et al., 1998) (Table 1).

The exclusion criteria were presence of neurological illness, brain injury or other significant medical illness, current or past substance abuse or dependence, and any kind of auditory impairment. In particular, all participants presented a normal screening audiogram. They gave written informed consent, according to the guidelines established by the institutional review board.

2.2. Stimuli

Stimuli were selected from the sound data bank used in the previous study (Aramaki et al., 2011). To design these sounds, 15 impact sounds from everyday life objects (wood beams, glass bowls and metallic plates) that were 5 sounds per material, were recorded, analyzed and re-synthesized with an additive synthesis model. The synthetic sounds were then tuned to the same chroma and equalized by gain adjustments to minimize influence of pitch and loudness variations in the categorization judgments. From these 15 reference sounds, 15 continua were constructed to simulate progressive transitions between material categories (i.e., 5 Wood – Metal, 5 Glass – Metal, and 5 Glass – Wood continua). Each continuum was composed of 20 hybrid sounds that were obtained by interpolating the synthesis parameters of the 2 reference sounds located at the extreme positions. A total of 300 hybrid sounds (without reference sounds) were created.

To shorten the duration of the experiment for SCZ, we reduced the previous sound corpus to 3 continua (one per transition): based on the response curves obtained in the previous study, we selected the continua presenting curves that were most centered on the middle sound position. We also verified that all reference sounds differed from each other (e.g. the reference sound of Metal in Wood continuum). Thus, 60 impact sounds1 i.e. 3 continua of 20 hybrid sounds, were kept for the current study (Fig. 1A). The mean sound duration was 875 ms for Wood – Metal, 1044 ms for Glass – Metal, and 653 ms for Glass – Wood continuum.

2.3. Procedure

Impact sounds were presented through the internal sound card of a PC computer in a quiet room, using open headphones (HD650 Sennheiser) ampli

Table 1 Demographic and clinical characteristics, behavioral data and values of parameter d (slope) and c (shift point) obtained from bootstrap procedure. Statistical differences were tested with Mann–Whitney U-test.

<table>
<thead>
<tr>
<th></th>
<th>SCZ (n = 20)</th>
<th>CTL (n = 20)</th>
<th>Differences (SCZ-CTL)</th>
<th>Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Demographic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age, mean (S.D.)</td>
<td>31.90 (6.95)</td>
<td>29.65 (10.87)</td>
<td>+2.25</td>
<td>0.440</td>
</tr>
<tr>
<td>Gender (male)</td>
<td>43%</td>
<td>43%</td>
<td>0</td>
<td>1.000</td>
</tr>
<tr>
<td>Education, mean (S.D.)</td>
<td>6.40 (3.12)</td>
<td>7.80 (2.31)</td>
<td>−1.40</td>
<td>0.124</td>
</tr>
<tr>
<td>Hearing threshold (DB)</td>
<td>14.88 (4.42)</td>
<td>13.08 (3.85)</td>
<td>+1.80</td>
<td>0.347</td>
</tr>
<tr>
<td>Left, mean (S.D.)</td>
<td>14.94 (4.43)</td>
<td>12.89 (4.25)</td>
<td>+2.05</td>
<td>0.231</td>
</tr>
<tr>
<td><strong>Clinical</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PANSS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive, mean (S.D.)</td>
<td>17.45 (6.47)</td>
<td>17.45 (6.47)</td>
<td>0</td>
<td>1.000</td>
</tr>
<tr>
<td>Negative, mean (S.D.)</td>
<td>20.4 (7.62)</td>
<td>20.4 (7.62)</td>
<td>0</td>
<td>1.000</td>
</tr>
<tr>
<td>Age 1st hospitalization, mean (S.D.)</td>
<td>20.80 (2.67)</td>
<td>20.80 (2.67)</td>
<td>0</td>
<td>1.000</td>
</tr>
<tr>
<td>Lifetime hospitalizations, mean (S.D.)</td>
<td>11.0 (6.74)</td>
<td>11.0 (6.74)</td>
<td>0</td>
<td>1.000</td>
</tr>
<tr>
<td>Chlorpromazine equivalents, mean (S.D.)</td>
<td>368.58 (218.54)</td>
<td>368.58 (218.54)</td>
<td>0</td>
<td>1.000</td>
</tr>
<tr>
<td><strong>Behavioral data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RT (s), mean (S.D.)</td>
<td>4.49 (1.34)</td>
<td>4.26 (1.25)</td>
<td>+0.23</td>
<td>0.310</td>
</tr>
<tr>
<td>Unexpected responses (%)</td>
<td>4.49 (1.34)</td>
<td>4.49 (1.34)</td>
<td>0</td>
<td>1.000</td>
</tr>
<tr>
<td>Glass (in Wood–Metal),</td>
<td>5.75 (10.5)</td>
<td>1.25 (3.2)</td>
<td>+4.50</td>
<td>0.357</td>
</tr>
<tr>
<td>Wood (in Glass–Metal),</td>
<td>3 (5.5)</td>
<td>0.75 (2.45)</td>
<td>+2.25</td>
<td>0.296</td>
</tr>
<tr>
<td>Metal (in Glass–Wood),</td>
<td>25.5 (18.2)</td>
<td>29.75 (15.75)</td>
<td>−4.25</td>
<td>0.670</td>
</tr>
<tr>
<td><strong>Bootstrap data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slope, median (Quartile 25, Quartile 75)</td>
<td>0.53 (0.45, 0.66)</td>
<td>0.93 (0.81, 1.04)</td>
<td>−0.40</td>
<td>≤0.001</td>
</tr>
<tr>
<td>Glass–Metal,</td>
<td>0.70 (0.58, 0.86)</td>
<td>0.88 (0.75, 1.07)</td>
<td>−0.18</td>
<td>≤0.001</td>
</tr>
<tr>
<td>Glass–Wood,</td>
<td>0.72 (0.59, 0.82)</td>
<td>0.95 (0.69, 2.99)</td>
<td>−0.23</td>
<td>≤0.001</td>
</tr>
<tr>
<td>Shift point, median (Quartile 25, Quartile 75)</td>
<td>9.91 (9.46, 10.19)</td>
<td>9.49 (9.13, 9.73)</td>
<td>+0.42</td>
<td>≤0.001</td>
</tr>
<tr>
<td>Glass–Metal,</td>
<td>9.25 (8.74, 9.71)</td>
<td>8.78 (8.48, 9.09)</td>
<td>+0.53</td>
<td>≤0.001</td>
</tr>
<tr>
<td>Glass–Wood,</td>
<td>8.75 (8.33, 9.14)</td>
<td>8.94 (8.59, 9.44)</td>
<td>−0.19</td>
<td>0.157</td>
</tr>
</tbody>
</table>

2 A pretest was conducted to evaluate the handiness of the interface with 14 healthy subjects (5 women and 9 men; mean age: 34 years old, S.D. = 13.75) and 3 schizophrenia patients. They did not belong to the groups that were tested in the formal experiment. We controlled: i) that they well understood the task, and that they easily answered with the interface, and ii) that data were recorded, correctly stored, and easy to export toward statistic software.

3. Results

3.1. Behavioral data

Individual responses were averaged so that each sound was associated with a percentage of categorization in Wood, Metal and Glass category for each group. Then, for each material continuum, a response curve was constructed from the percentage of categorization in the second category of material (e.g., Metal for Wood–Metal) as a function of sound position along the continuum (Fig. 1B). Percentages of unexpected responses (e.g. Glass in Wood–Metal continuum) and type amp). Subjects were free to adjust the intensity level of the sounds, once at the beginning of the test. The experimenter controlled that this level was high enough to provide a comfortable listening condition for each subject and that this level was approximately set to the same value across participants in both groups.

The experimenter was present with the participant during the entire test and followed the process through second headphones connected on the same Samsom amplifier. The experiment began with a 4-trials training session to familiarize participants with the task. Then, all sounds were randomly presented in a single session. Participants could listen to the sound as many times as they wanted, and were asked to categorize each sound once as Wood, Glass or Metal, by selecting with a mouse the material category that was the closest to the sound.

Table 1: Demographic and clinical characteristics, behavioral data and values of parameter d (slope) and c (shift point) obtained from bootstrap procedure. Statistical differences were tested with Mann–Whitney U-test.
RT values are shown in Table 1. No significant differences in unexpected responses and RT were found between SCZ and CTL. Note that RTs were quite short (about 4 s) in both groups, suggesting that subjects in SCZ did not experience the test as too difficult.

3.2. Data modeling

Since the response curves presented a sigmoid shape as previously found in Aramaki et al. (2011), they were modeled by the following logistic function:

\[ y = \frac{a + b}{1 + e^{-(x-c) \cdot d}} \]

where \( y \) is the probability to categorize a sound in the second category of material in the continuum, \( x \) the sound position, \( a \) the lower asymptote, \( b \) the difference between upper and lower asymptotes, \( c \) the sound position at midpoint (shift point) and \( d \) the slope. The logistic parameters were estimated by fitting data at best in a least-squares sense (Fig. 1B). Note that Kee et al. (2006) and Pollak and Kistler (2002) used this modeling in studies based on an emotional continuum paradigm.

Since participants categorized each sound only once, we obtained one single response curve for each continuum and for each group. Therefore, a bootstrap randomization method (Efron, 1979; Kee et al., 2006) was used to generate empirical standard error estimates and to evaluate statistical group differences in shift point and slope. In practice, sets of 20 records (i.e. individual responses) were chosen randomly (with replacement) from all available records for each continuum and for each group. The logistic parameters \( c \) and \( d \) were calculated from the resulting response curve and the procedure was repeated 100 times. Finally, a Mann–Whitney U-test was conducted on the two series of 100 values to evaluate differences between groups.

In SCZ, the slopes were significantly smaller for all continua and the shift points moved towards the Metal category in Wood–Metal and Glass–Metal continua (Table 1). The same shift towards Metal was obtained in SCZ by considering the percentages of categorization in the first category of material (i.e., Wood in Wood–Metal and Glass in Glass–Metal continuum).

4. Discussion

As expected, data analysis highlighted shallower curves for all transitions with significantly smaller slopes in SCZ than in CTL. The slope indicates how abrupt the categorization between materials is. Thus, this suggests that SCZ perceive category transitions in a more smooth way than CTL, similar to studies investigating facial emotional continua in visual modality (Kee et al., 2006) and Ba–Da continuum with phonetic sounds (Cienfuegos et al., 1999).

Since damping and roughness were previously shown to be relevant timbre features for the material perception (Aramaki et al., 2011), this result lead us to assume that SCZ are less able than CTL to perceive these specific acoustic features. This assumption is in line with the literature that considers perceptual impairments as consequences of sensory processing dysfunctions (Cienfuegos et al., 1999; Javitt, 2009a, 2009b), such as Leitman et al. (2008) who related the prosodic emotion recognition impairment in schizophrenia to pitch processing dysfunction.

However, based on computational models of schizophrenia (Rolls et al., 2008), we cannot discard a more general deficit in the sensory information processing across modalities. In line with this consideration, Bach et al. (2009) hypothesized that their findings in emotional
Prosodic recognition impairment in schizophrenia can be related to global increased fluctuations in cerebral networks. The authors discussed this hypothesis in the SDT framework, as a broader distribution of noise in the internal representation of the signal (i.e., a reduced signal-to-noise ratio) leading to shallower response curves observable in different modalities, for example in visual modality in Kee et al.’s (2006) study. Hence, for our concern, we can interpret our findings as the fact that the internal representation of material categories might be less distinctive in SCZ.

Going further, data analysis highlighted additional results specifically related to the Metal perception. Indeed, we found that the shift point was moved towards the Metal endpoint category in Wood–Metal and Glass–Metal continua in SCZ. The shift point indicates the perceptual boundary between categories. Thus, in the SDT framework, we can conclude on a biased response criterion for the Metal category perceptual boundary between categories. Thus, in the SDT framework, we can conclude on a biased response criterion for the Metal category perception of facial emotion in schizophrenia. Schizophrenia Research 87, 228–234.

In conclusion, this is the first study to our knowledge that used calibrated and parameterized sounds to assess environmental sound perception in schizophrenia. Results highlighted disturbances of sound categorization that we supposed to be induced by impairments in the perception of specific timbre features and in the internal representation of material category in the SDT framework. This study confirmed the importance of investigating timbre perception in schizophrenia and constitutes a new step to better understand distortions related to this disorder, which induce deep handicap in the everyday situation.

Acknowledgments
The authors thank Dr. Laurent Boyer for providing help with statistical analysis.

References


