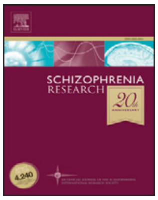
Schizophrenia Research 107 (2009) 286–293

Contents lists available at ScienceDirect

Schizophrenia Research

journal homepage: www.elsevier.com/locate/schres



# Audiovisual emotion recognition in schizophrenia: Reduced integration of facial and vocal affect

J.J. de Jong [a,b], P.P.G. Hodiamont [b,c], J. Van den Stock [a,d], B. de Gelder [a,e,*]

[a] *Cognitive Neuroscience Laboratory, Tilburg University, P.O. Box 90153, 5000 LE, The Netherlands*
[b] *GGZ Midden-Brabant, P.O. Box 770, 5000 AT, The Netherlands*
[c] *Department of Developmental, Clinical and Cross-cultural Psychology, Tilburg University, P.O. Box 90153, 5000 LE, The Netherlands*
[d] *Old Age Psychiatry Department, University Hospitals Leuven, Brusselsestraat 69, 3000 Leuven, Belgium*
[e] *Martinos Center for Biomedical Imaging, Massachussetts General Hospital and Harvard Medical School, Building 36, First Street, Charlestown, MA 02129, USA*

ARTICLE INFO

*Article history:*
Received 3 July 2008
Received in revised form 29 September 2008
Accepted 2 October 2008
Available online 5 November 2008

*Keywords:*
Emotion recognition
Multisensory integration
Schizophrenia

ABSTRACT

Since Kraepelin called *dementia praecox* what we nowadays call schizophrenia, cognitive dysfunction has been regarded as central to its psychopathological profile. Disturbed experience and integration of emotions are, both intuitively and experimentally, likely to be intermediates between basic, non-social cognitive disturbances and functional outcome in schizophrenia. While a number of studies have consistently proven that, as part of social cognition, recognition of emotional faces and voices is disturbed in schizophrenics, studies on multisensory integration of facial and vocal affect are rare.

We investigated audiovisual integration of emotional faces and voices in three groups: schizophrenic patients, non-schizophrenic psychosis patients and mentally healthy controls, all diagnosed by means of the Schedules of Clinical Assessment in Neuropsychiatry (SCAN 2.1). We found diminished crossmodal influence of emotional faces on emotional voice categorization in schizophrenics, but not in non-schizophrenia psychosis patients. Results are discussed in the perspective of recent theories on multisensory integration.

## 1. Introduction

The negative impact of schizophrenia on building and maintaining a social life caused the World Health Organization to call it “youth's greatest disabler” (WHO, 2008). Since the relation between social disabilities and cognitive deficits has been established (Green et al., 2000, 2004; Velligan et al., 1997, 2000), *social* cognition has become a high priority area in schizophrenia research (Couture et al., 2006; Green et al., 2008). Social cognition refers to ‘the mental operations that underlie social interactions, including perceiving, interpreting and generating responses to the intentions, dispositions and behaviors of others’ (Green et al., 2008). Besides social cognition, specific disturbances in the automatic, pre-attentive processing of sensory information are thought to be predictors of social outcome (Braff and Light, 2004; Light and Braff, 2005; Swerdlow et al., 2006). Sensorimotor gating deficits, as assessed by prepulse inhibition (PPI) paradigms, represent robust, trait-like features of schizophrenia (Braff et al., 2001; Grillon et al., 1992; Schall et al., 1996; Swerdlow et al., 2006). Whereas sensorimotor gating is part of the *inhibitory* mechanism that regulates the processing of environmental stimuli, we found the multisensory *integration* (MSI) mechanism of visual (facial) and auditory (vocal) emotional information to be disturbed in schizophrenia as well (de Gelder et al., 2005b). Studies on non-affective MSI in schizophrenia are scarce, but they also suggest impaired integration (de Gelder et al., 2003; Ross et al., 2007; Surguladze et al., 2001).

Environmental stimuli are perceived through separate sensory channels. MSI refers to the neurocognitive operations that underlie the integration of temporally and/or spatially coincident streams of information (Bertelson and de Gelder, 2004; Calvert, 2004; de Gelder and Bertelson, 2003; Hershenson, 1962; Meredith and Stein, 1986; Stein et al., 1988; Voort

---

* Corresponding author. Martinos Centre for Biomedical Imaging, Massachusetts General Hospital, Room 417, Building 36, First Street, Charlestown, Massachussets 02129, USA. Tel.: +1 6177267956; fax: +1 31134662067.
*E-mail address:* degelder@nmr.mgh.harvard.edu (B. de Gelder).

0920-9964/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.schres.2008.10.001

**Table 1**
Demographic, clinical (PANSS) and neuropsychological (CPT) characteristics of the three groups of subjects (schizophrenic patients, non-schizophrenic psychosis patients, controls)

| | Schizophrenic subjects | Non-schizophrenic psychosis subjects | Control subjects | Significance |
|---|---|---|---|---|
| *N* | 55 | 46 | 50 | |
| Age (mean years ± SD) [a] | 33.53 (8.80) [c] | 35.22 (9.04) [c] | 41.16 (12.94) | $p=0.001$ |
| Sex (% men) [b] | 70.9 | 63.0 | 48.0 | $p=0.052$ |
| Handedness (% right-handed) [b] | 85.5 | 84.8 | 88.0 | $p=0.888$ |
| Education (within-group %) [b] | | | | $p=0.079$ |
| 1 | 7.3 | 2.2 | 0.0 | |
| 2 | 18.2 | 21.7 | 6.0 | |
| 3 | 40.0 | 37.0 | 56.0 | |
| 4 | 34.5 | 39.1 | 38.0 | |
| PANSS [a] | | | | |
| Positive | 16.8 | 13.6 | | $p=0.001$ |
| Negative | 20.6 | 16.2 | | $p=0.001$ |
| General | 38.1 | 35.0 | | $p=0.058$ |
| Total | 75.5 | 64.8 | | $p=0.001$ |
| CPT (mean *d'* ± SD) [a] | 3.349 (0.763) [c] | 3.061 (0.894) [c] | 3.749 (0.631) | $p<0.001$ |

[a] ANOVA.

[b] Chi-square.

[c] Significantly different from controls, not the other patient group.

et al., 1972). MSI serves the ability of man and animal alike to experience their environment as one integral occurrence. The MSI process implies an early interaction between sensory channels in the pre-attentive phases of perception (Foxe and Schroeder, 2005; Giard and Peronnet, 1999; Schroeder et al., 2001; Schroeder and Foxe, 2002). This *crossmodal influence* between channels facilitates behavioral and neural performance (Bertelson and de Gelder, 2004; Brancazio and Miller, 2005; Calvert et al., 2000; Calvert, 2001, 2004; de Gelder, 2000; Hershenson, 1962; Jones and Callan, 2003; Macaluso et al., 2004; McGurk and MacDonald, 1976; Meredith and Stein, 1986; Radeau, 1994; Stein et al., 1988). The integration of *emotionally neutral* seen lip movements and heard speech, for example, improves listening to another person, where the binding of facial or bodily (i.e. visual) with prosodic (i.e. auditory) *emotional* information improves understanding the other (de Gelder et al., 2002, 2005a; Dolan et al., 2001; Meeren et al., 2005; Van den Stock et al., 2007). So, the consequences of MSI for interpersonal behavior are obvious.

Using audiovisual emotional categorization tasks, we explored MSI in schizophrenia patients, as compared to healthy controls and non-schizophrenia psychosis patients. As a measure of MSI, crossmodal impact of emotional faces on emotional voices was determined for both emotion-congruent and -incongruent conditions. Regarding our previous study on this topic (de Gelder et al., 2005b), we hypothesized diminished MSI performance in the schizophrenia group. In comparison with our first study, however, larger groups, including a patient-control group, were included, patient characteristics were more extensively described and task materials were simplified. In addition, we explicitly addressed vigilance and use of medication as possible confounds.

## 2. Method

### 2.1. Participants

Three groups participated in our study. The first group consisted of 55 schizophrenic patients (**Sch** group). The second group included 46 patients with non-schizophrenic psychotic illness (**N-Sch-Psy** group). The third group (**Ctrl** group) contained 50 neurologically and psychiatrically normal controls. The study was approved by the regional Medical Ethics Committee. Subjects gave their informed consent and participated for a small fee.

All patients (**Sch** and **N-Psy-Sch**), being outpatients of the local psychiatric hospital and relatively stable with respect to symptomatology, were examined by a trained psychiatrist (JJdJ) using the Schedules of the Assessment in Neuropsychiatry (SCAN) version 2.1. By means of the computerized SCAN 2.1 algorithm, a DSM-IV (APA, 2000) classification was obtained. In case of DSM-IV schizophrenia, patients were assigned to the first group. In case of all other types of (i.e. non-schizophrenic) psychotic illness, patients were assigned to the second group.

**Table 2**
DSM-IV classifications within both patient groups (schizophrenic patients, non-schizophrenic psychosis patients)

| | Schizophrenic subjects | Non-schizophrenic psychosis subjects |
|---|---|---|
| 295.30 Schizophrenia, paranoid type | 53 | |
| 295.90 Schizophrenia, residual type | 2 | |
| 295.40 Schizophreniform disorder | | 1 |
| 295.70 Schizoaffective disorder, bipolar type | | 3 |
| 295.70 Schizoaffective disorder, depressive type | | 5 |
| 297.1 Delusional disorder, persecutory type | | 3 |
| 298.8 Brief psychotic disorder | | 3 |
| 296.44 Bipolar I disorder, last episode manic, with psychosis | | 12 |
| 296.54 Bipolar I disorder, last episode depressed, with psychosis | | 1 |
| 296.24 Depressive disorder, single episode, with psychosis | | 3 |
| 296.34 Depressive disorder, recurrent, with psychosis | | 2 |
| 298.9 Psychosis not otherwise specified | | 13 |
| Total | 55 | 46 |

Inclusion criteria for patients were: DSM-IV diagnosis of psychotic illness, age between 20 and 60 years. Exclusion criteria for patients were: organic or substance-induced psychosis, current substance abuse, serious somatic illness, relevant neurological illness, auditory and/or visual handicap and language problems. All patients were submitted to the Positive and Negative Symptom Scale (Kay et al., 1987).

Controls between 20 and 60 years of age were recruited by advertisement in a local newspaper. The SCAN 2.1 interview was used to screen for psychiatric illness. Exclusion criteria were: current substance abuse, serious somatic illness, relevant neurological illness, auditory and/or visual handicap and language problems.

Highest completed school grade was recorded, according to the usual conventions (Pichot et al., 1993), using four categories being suitable to the Dutch educational system (1 = elementary school; 2 = junior secondary vocational education; 3 = secondary education; 4 = at least higher secondary education).

Groups did not significantly differ on sex ratio ($\chi^2(2,151)=5.90$, $p=n.s.$), educational level ($\chi^2(6,151)=11.32$, $p=n.s.$) and handedness ($\chi^2(2,151)=0.24$, $p=n.s.$) (Table 1). Age differences were, though significant ($F(2,150)=7.59$, $p<0.001$), small (mean ages 33.53, 35.22 and 41.16 years for **Sch**, **N-Sch-Psy** and **Ctrl**, respectively). Positive, negative and total PANSS-scores were significantly higher in **Sch** than in **N-Psy-Sch** ($F(1,100)=258.60$, $p<0.001$; 480.35, $p<0.001$ and 2882.13, $p<0.001$, respectively). Table 2 shows subtypes of schizophrenia (group 1) and non-schizophrenic psychosis (group 2). Table 3 gives an overview of different types of medications used for both patient groups.

### 2.2. Materials and procedure

Two tasks were administered, during which subjects were asked to categorize emotional voices, while watching an emotional face on a computer screen. In task 1, ‘happy’ and ‘fear’ were used as vocal and facial emotions. In task 2, ‘happy’ and ‘sad’ were used.

In addition, all subjects were submitted to a continuous performance test (CPT) to determine pre-existing group differences with respect to vigilance. Using a 3-7-target format, 600 digits (ranging from 1 to 9) were randomly presented with a frequency of 1/s, resulting in a total test

**Table 3**
Use of antipsychotic and mood-stabilizing medication within both patient groups (schizophrenic patients, non-schizophrenia psychosis patients)

| | Schizophrenic subjects | | Non-schizophrenic psychosis patients | |
|---|---|---|---|---|
| | *N* | Mean daily dosage (mg) (minimum–maximum) | *N* | Mean dosage (minimum–maximum) |
| *Antipsychotic medication* | | | | |
| Aripiprazole | 3 | 20.0 (15.0–30.0) | 0 | |
| Broomperidol (decanoate) | 2 | 9.8 (7.1–12.5) | 1 | 1.8 |
| Clozapine | 4 | 537.5 (200.0–800.0) | 2 | 475.0 (250.0–700.0) |
| Flupentixol | 0 | | 1 | 5.0 |
| Fluphenazine | 1 | 3.6 | 0 | |
| Haloperidol | 2 | 5.8 (4.0–7.5) | 1 | 7.5 |
| Olanzapine | 17 | 12.9 (5.0–25.0) | 10 | 10.5 (5.0–20.0) |
| Penfluridol | 1 | 1.4 | 1 | 1.4 |
| Pimozide | 0 | | 2 | 3.0 (2.0–4.0) |
| Quetiapine | 1 | 300.0 | 2 | 600.0 (400.0–800.0) |
| Risperidone | 17 | 2.9 (2.0–4.0) | 16 | 2.9 (0.5–6.0) |
| Risperidone (suspension for injection) | 1 | 1.8 | 0 | |
| Thioridazine | 0 | | 1 | 200 |
| Zuclopentixol | 0 | | 1 | 25.0 |
| No antipsychotic medication | 7 | | 9 | |
| Total | 56 [a] | | 47 [b] | |
| *Mood-stabilizing medication* | | | | |
| Lithium | 1 | 800.0 | 8 | 975.0 (400.0–1400.0) |
| Topimarate | 0 | | 1 | 250.0 |
| Valproate | 1 | 750.0 | 4 | 950.0 (300.0–1500.0) |
| No mood-stabilizing medication | 53 | | 35 | |
| Total | 55 | | 48 [c] | |
| *Antidepressive medication* | | | | |
| Citalopram | 11 | 30.9 (20.0–40.0) | 4 | 35.0 (20.0–40.0) |
| Clomipramine | 2 | 187.5 (150.0–225.0) | 3 | 166.7 (125.0–225.0) |
| Dosulepine | 0 | | 1 | 75.0 |
| Fluoxetine | 0 | | 1 | 20.0 |
| Imipramine | 1 | 50.0 | 1 | 150.0 |
| Mirtazapine | 0 | | 3 | 40.0 (30.0–60.0) |
| Nortryptiline | 1 | 75.0 | 1 | 150.0 |
| Paroxetine | 0 | | 3 | 26.7 (20.0–40.0) |
| Tranylcypramine | 1 | 40.0 | 0 | |
| No antidepressive medication | 39 | | 29 | |
| Total | 55 | | 46 | |

[a] One schizophrenic patient used two antipsychotics: olanzapine and quetiapine.
[b] One non-schizophrenic psychosis patient used two antipsychotics: olanzapine and pimozide.
[c] Two non-schizophrenic psychosis patient used two mood-stabilizors: lithium and valproate.

duration of 10 min (CDLJava, version7.01) (Lezak et al., 2004). Mean CPT *d'* scores are shown in Table 1.

To explore a possible relationship between medication and performance, individual daily doses of antipsychotics (AP) in milligrams were converted into multiples of the Daily Defined Dose (DDD) of each drug (Nose et al., 2008; WHO Collaborating Center for Drug Statistic Methodology, 2003).

#### 2.2.1. Task 1

*2.2.1.1. Auditory material.* Four (2 males and 2 females) professional Dutch actors were instructed to pronounce four different and semantically neutral sentences 'as if he or she was happy' or 'as if he or she was afraid'. Sentences were selected in accordance to the length of the visual stimuli. As a result, the following 4 utterances were obtained: "Auto gekocht", which means "Bought a car"; "Naar Amsterdam", which means "To Amsterdam"; "Kapper geweest", which means "Been to hairdresser" and "Met het vliegtuig", which means "By airplane". The sentences were recorded digitally. The durations of the 36 (4 actors, 4 semantically different sentences and 2 emotions) utterances ranged from 599 to 1265 ms.

*2.2.1.2. Visual material.* Eight black-and-white photographs were drawn from the Ekman and Friesen series (Ekman and Friessen, 1976), representing 4 (2 males and 2 females) actors, each of them expressing happiness or fear. Each photograph occupied a 13.0 × 8.5 cm rectangle on the computer screen, which at the mean viewing distance of 60 cm corresponded to a visual angle of 12.4 × 8.1°. Photographs were showed against a black coloured background.

*2.2.1.3. Design and procedure.* Subjects were tested individually in a quiet room. Stimuli were presented after the participant pushed a response button that subsequently initiated the next trial. All trials consisted of a synchronically presented emotional face and emotional voice. Emotional faces had a fixed duration of 800 ms. Emotional voice utterances started at exactly the same time as emotional faces.

Thirty-two voice utterances were used. Four (2 males and 2 females) actors presented 4 semantically different utterances in 2 different emotional tones. During the task, each utterance was presented twice: one synchronically with an emotion-congruent face (for example a happy utterance with a happy face) and the other with an emotion incongruent face (for example a happy utterance with a fearful face). As a result, the task consisted of 64 stimuli, which were presented in pseudorandom order: vocal and facial actors were fixed couples with the same gender.

Instructions were provided both on the screen and by the experimenter. Participants were instructed to press one of two keys as fast as possible to indicate whether the vocal emotion was 'happy' or 'fear'. Subjects were asked explicitly to watch the computer screen while listening to the voice, preventing them to concentrate on the voices only. They were, however, instructed to ignore the facial emotion while judging the voice. Hereafter, a practice phase of 4 trials was administered.

#### 2.2.2. Task 2

Except for the emotional content of the stimuli, the same materials and procedure as in task 1 were used. This time, target emotions were happiness and sadness, so that visual and auditory stimuli consisted of happy and sad faces and voices.

### 2.3. Statistical procedures and outlier management

We applied univariate analyses to this repeated measures design. To correct for pre-existing differences between groups concerning vigilance, CPT *d'* scores were entered as a covariate. Pearson correlation analyses were performed on the relationships between AP DDDs and performance on the congruent and incongruent conditions as well as their difference (crossmodal impact) scores.

A preliminary analysis of results revealed that some subjects appeared to have judged facial emotion instead of vocal emotion, resulting in normal accuracy rates in the *emotion-congruent* condition and extremely low accuracy rates in the *emotion-incongruent* condition. Furthermore, a few subjects appeared to have reversed response keys throughout the task. Data were considered as outliers and deleted when accuracy rates for the *emotion-congruent* condition were below chance-level (0.5) or when differences between *emotion-congruent* and *emotion-incongruent* conditions amounted more than 0.5. Numbers of outliers did not differ significantly between groups (total outliers in task 1 and/or task 2 were 7, 5 and 3 for **Sch**, **N-Sch-Psy** and **Ctrl**, respectively; $\chi^2(2,151) = 1.39$, $p = n.s.$).

## 3. Results

CPT *d'* scores significantly differed between-groups (see Table 1; $F(2,135) = 8.55$, $p < 0.001$). When groups were compared pairwise, using Tukey's post-hoc analyses, a significant difference was found between **Sch** and **Ctrl** ($p = 0.035$) and between **N-Sch-Psy** and **Ctrl** ($p < 0.001$), not between both patient groups.

### 3.1. Happy–fear (task 1)

Fig. 1 presents proportions of correct responses as a function of facial emotion condition. Facial emotion was either congruent or incongruent with vocal emotion. Categorization of emotional (happy and fear) voices was more accurate when an *emotion-congruent* face as compared to an *emotion-incongruent* face accompanied vocal affect (even when subjects were instructed to ignore facial affect while categorizing voices). Furthermore, this crossmodal influence effect of a face on the voice was obviously larger in **Ctrl** and **N-Sch-Psy** than in **Sch**. Diminished crossmodal impact in **Sch** is illustrated by the flattened slope of the line between the two face conditions as compared to both other groups.

A 2 × 3 ANOVA was performed with Congruence (congruent, incongruent) as within-subject factor and Group (**Sch**, **N-Sch-Psy** and **Ctrl**) as between-subject factor. From these main effects, Congruence was significant ($F(1,135) = 183.99$, $p < 0.001$). This finding reflects the crossmodal effect of the face on the voice: vocal emotion was categorized more accurate when the simultaneously shown face was *emotion-congruent* (0.789) as compared to *emotion-incongruent* (0.645). The fact that the effect of Group was not significant ($F(2,135) = 1.96$, *n.s.*) illustrates that overall-task difficulty was

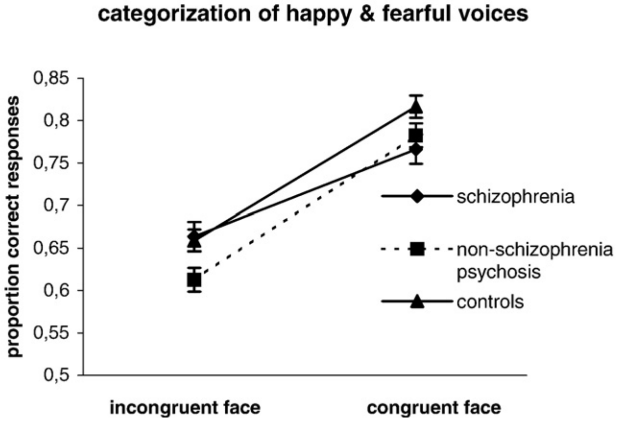


**Fig. 1.** Categorization of happy & fearful voices: the proportions of correct responses as a function of facial emotion (either emotion-incongruent or emotion-congruent) are displayed. As is illustrated by a flattened slope, schizophrenia patients show diminished crossmodal impact of an emotional face on an emotional voice.

the same for all groups. More importantly, we found an interaction effect of Congruence × Group ($F(2,135)=3.97$, $p=0.021$), which statistically illustrates diminished crossmodal impact of facial on vocal emotion in **Sch**, as was already apparent from Fig. 1. Cohen's effect size for this effect, comparing **Sch** to **Ctrl**, was 0.42.

To test the possibility that preexisting differences between groups with respect to vigilance (affecting task performance) confounded the latter effect, we entered CPT *d'* scores in the model as a covariate. While the main effect of Group remained non-significant, the interaction effect of Congruence × Group became *more* significant ($p=0.008$), with Cohen's effect size counting 0.53.

Across both patient groups, no correlation was found between AP DDDs and performance on the congruent and incongruent condition as well as their difference (crossmodal impact) scores (Pearson's correlation coefficient 0.110; 0.073 and 0.026 respectively, $p$=*n.s.* for all).

### 3.2. Happy–sad (task 2)

Fig. 2 (for reasons of clarity plotted with a rescaled Y-axis as compared to Fig. 1) presents proportions of correct responses as a function of facial emotion condition, using facial emotion that was either congruent or incongruent with vocal emotion. Again, categorization of emotional (happy and sad) voices was more accurate when an *emotion-congruent* face as compared to an *emotion-incongruent* face accompanied vocal affect (while subjects were told to ignore facial emotion and categorize vocal emotion only). As in task 1 (Fig. 1), this crossmodal bias effect of a face on the voice seems larger in **Ctrl** and **N-Sch-Psy** than in **Sch**, illustrated once more by the flattened slope between the two face conditions for **Sch** compared to both other groups.

A 2×3 ANOVA was performed with Congruence (congruent, incongruent) as within-subject factor and Group (**Sch**, **N-Sch-Psy** and **Ctrl**) as between-subject factor. From the two main effects, Congruence was significant ($F(1,143)=70.15$, $p<0.001$). This finding results from the expected crossmodal impact of an *emotion-congruent* (0.877) as compared to an *emotion-incongruent* (0.819) face on vocal emotion categorization. The fact that the main effect of Group ($F(2,143)=1.06$, $p$=*n.s.*) was not significant illustrates that, as in task 1, task difficulty was the same for all groups. In contrast to task 1, however, the interaction effect of Congruence × Group did *not* reach statistical significance ($F(2,143)=0.70$, $p$=*n.s.*) this time.

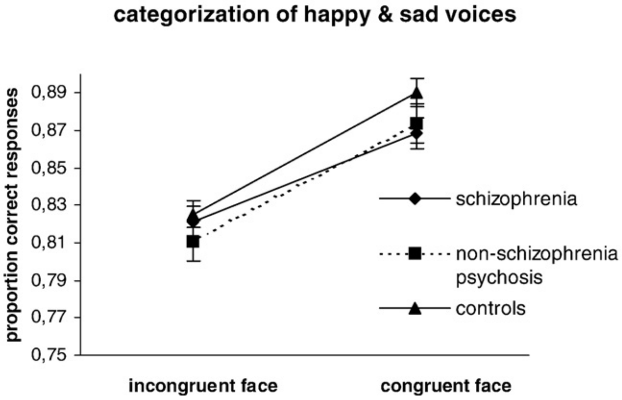


**Fig. 2.** Categorization of happy & sad voices: the proportions of correct responses as a function of facial emotion (whether emotion-incongruent or emotion-congruent) are displayed. While schizophrenia patients show comparable accuracy rates to controls in the case of emotion-incongruent faces, their performance is the lowest when face-voice pairs are emotion-congruent.

After controlling for CPT scores, the main effect of Group and the interaction effect of Congruence × Group both remained non-significant.

As in task 1, no correlation was found between AP DDDs and performance on the congruent and incongruent condition as well as their difference (crossmodal impact) scores (Pearson's correlation coefficient −0.027; −0.005 and −0.050 respectively, $p = n.s.$ for all).

### 3.3. Happy–fear and happy–sad (task1 and task 2)

Because of the resemblance between Figs. 1 and 2 on the one hand, and the different levels of statistical significance with respect to the interaction of Congruence × Group on the other, an omnibus 2 × 2 × 3 ANOVA was performed. In addition to the 2 × 3 ANOVAs in both separate tasks, Emotions (task 1 (happy/fear), task 2 (happy/sad)) was added as a second within-subject factor to the model. As in the separate analyses, main effect of Congruence was highly significant ($F(1,133) = 219.87$, $p < 0.001$) and the main effect of Group was non-significant. The main effect of Task showed also high statistical significance ($F(1,133) = 332.20$, $p < 0.001$), reflecting the substantial difference in overall performance between task 1 (0.720) and task 2 (0.851). Apparently, happy and sad voices were more easily classified than happy and fearful voices, probably resulting in *ceiling effects* in task 2. Most importantly, the Congruence × Group interaction *across* both tasks reached the same statistical significance as was shown at task 1 ($F(2,133) = 3.93$, $p = 0.022$). Cohen's effect size for this effect, comparing **Sch** to **Ctrl**, was 0.44.

Correction for preexisting differences in CPT *d'* scores again, as in task 1, amplified the uncorrected level of significance ($F(2,116) = 4.69$, $p = 0.011$), while the main effect of Group remained non-significant. After this correction, Cohen's effect size was 0.52.

Again, no correlation was found between AP DDDs and performance on the congruent and incongruent condition as well as their difference (crossmodal impact) scores (Pearson's correlation coefficient 0.042; 0.025 and 0.010 respectively, $p = n.s.$ for all).

## 4. Discussion

In sum, the combined data of task 1 and task 2 indicate that the crossmodal influence of emotional faces on emotional voice categorization is diminished in schizophrenics, but not in non-schizophrenia psychosis patients. Specific exploration of the role of vigilance and AP medication use indicated that these factors did not induce or influence our findings.

The most straightforward explanation for this result is a diminished crossmodal impact of an emotional face on an emotional voice due to deficits in MSI in schizophrenia. However, some possible confounds may have affected the results.

First, *response bias* has been mentioned as a possible confounder (Vroomen et al., 2001). Subjects could, to some extent, have relied on the non-target source of information (facial emotion). Because **Sch** subjects performed somewhat better (task 1) than or comparable (task 2) to both other groups in the incongruent condition, response bias would have affected **Ctrl** and **N-Sch-Psy** subjects more frequently than **Sch** patients. Although contra-intuitive, this theoretical possibility cannot be ruled out. Second, a different but related issue refers to diminished task-vigilance. We statistically controlled for preexisting between-group differences with respect to CPT *d'* scores. This *strengthened* the pattern of diminished crossmodal impact in **Sch** in task 1 and in the omnibus analysis, while the resembling pattern in task 2 remained preserved. So, reduced task-vigilance in **Sch** and **N-Sch-Psy** reduced, rather than caused the finding of diminished MSI in **Sch**. Third, differential task-difficulty between groups might be a factor to consider. However, no main effects of Group were found, whether correction for task-vigilance was used or not. Fourth, use of medication or psychopathological 'load' may be a relevant factor. However, **N-Sch-Psy** subjects showed a crossmodal influence comparable to **Ctrl** while being on antipsychotic medication and having severe symptoms. In sum, disturbed MSI of affective information in schizophrenia is the most likely explanation for our findings. The only preceding study on this topic is, as far as we know, our initial investigation of this issue (de Gelder et al., 2005b). Then, we found possible deficits in affective MSI, but could not draw robust conclusions because groups were small, no non-schizophrenia psychosis group was included and description of patient characteristics was incomplete. Furthermore, our present investigation was aimed at reducing attentional factors and post-perceptual strategies by presenting emotions in their most explicit, dichotomous form.

In daily life, information concerning our fellow human being reaches us through different sensory channels. Outside the laboratory, these various inputs are mostly semantically congruent. Therefore, our *laboratory* finding of, at least, normal performance in **Sch** in case of emotion-*incongruent* faces does not seem to reflect on adaptive social interaction in daily life. But schizophrenia patients do lag behind when *congruent* information is deficiently integrated. Given the highly complex reality of interpersonal behavior, this might add to the pervasive social disabilities that characterize schizophrenia.

Another motive for further research on MSI in schizophrenia is provided in the general neuropsychological literature by the growing evidence for its neural basis. Cortical regions such as the anterior cingulate cortex, the fusiform gyrus, the middle temporal gyrus and the superior temporal sulcus are probably implicated in the MSI of affective information (de Gelder et al., 1999; Dolan et al., 2001; Kreifelts et al., 2007; MacLeod and MacDonald, 2000; Pourtois et al., 2000, 2005). In addition, amygdala activation is specifically described when fearful faces and fearful voices are combined (as in task 1) (Dolan et al., 2001; Pourtois et al., 2005). Several of these regions, potentially critical to MSI, have been mentioned by imaging studies on schizophrenia because of their functional deficits, which in turn might be explained by micro-architectural and neural migration disturbances (Brunet-Gouet and Decety, 2006).

Whether schizophrenia, classically described as a disorder of disintegration of mental functions, is also characterized by disintegration of what are normally automatic, mandatory processes of sensory integration will be a challenging topic for future research.

## Role of the funding source

Research was funded by a ZonMw grant (Dutch Science Foundation; 100-002-015). Furthermore, the study was partly supported by Human Frontier Science Program HFSP - RGP0054/2004-C and EC-contract number FP6-NEST-2005-Path-IMP-043403. Neither ZonMw nor HFSP had a role in study design; in the collection, analysis and interpretation of data; in the writing the report; and in the decision to submit the paper for publication.

## Contributors

Authors J.J. de Jong, P.P.G. Hodiamont and B. de Gelder designed the study and wrote the protocol. J.J. de Jong managed the literature searches, the data collection and interviewed all subjects using PANSS and SCAN 2.1. J.J. de Jong, J. Van den Stock and B. de Gelder undertook the statistical analysis. J.J. de Jong and P.P.G. Hodiamont wrote the first draft of the manuscript. All authors contributed to and have approved the final manuscript.

## Conflict of interest

All authors declare that they have no conflicts of interest.

## Acknowledgement

We thank all participants, especially the patients, for their willingness to contribute to the study.

## References

APA, 2000. Diagnostic and Statistical Manual of mental disorders, DSM-IV-TR. American Psychiatric Association, Washington, DC.

Bertelson, P., de Gelder, B., 2004. The psychology of multimodal perception. In: Spence, C., Driver, J. (Eds.), Crossmodal space and crossmodal attention. Oxford University Press, Oxford, pp. 151–177.

Braff, D.L., Light, G.A., 2004. Preattentional and attentional cognitive deficits as targets for treating schizophrenia. Psychopharmacology (Berl) 174, 75–85.

Braff, D.L., Geyer, M.A., Light, G.A., Sprock, J., Perry, W., Cadenhead, K.S., Swerdlow, N.R., 2001. Impact of prepulse characteristics on the detection of sensorimotor gating deficits in schizophrenia. Schizophr. Res. 49, 171–178.

Brancazio, L., Miller, J.L., 2005. Use of visual information in speech perception: evidence for a visual rate effect both with and without a McGurk effect. Percept. Psychophys. 67, 759–769.

Brunet-Gouet, E., Decety, J., 2006. Social brain dysfunctions in schizophrenia: a review of neuroimaging studies. Psychiatry Res. 148, 75–92.

Calvert, G.A., 2001. Crossmodal processing in the human brain: insights from functional neuroimaging studies. Cereb. Cortex 11, 1110–1123.

Calvert, G.A., 2004. Handbook of multisensory processes. MIT, Cambridge, MA.

Calvert, G.A., Campbell, R., Brammer, M.J., 2000. Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Curr. Biol. 10, 649–657.

Couture, S.M., Penn, D.L., Roberts, D.L., 2006. The functional significance of social cognition in schizophrenia: a review. Schizophr. Bull. 32 (Suppl. 1), S44–S63.

de Gelder, B., 2000. Neuroscience. More to seeing than meets the eye. Science 289, 1148–1149.

de Gelder, B., Bertelson, P., 2003. Multisensory integration, perception and ecological validity. Trends Cogn. Sci. 7, 460–467.

de Gelder, B., Bocker, K.B., Tuomainen, J., Hensen, M., Vroomen, J., 1999. The combined perception of emotion from voice and face: early interaction revealed by human electric brain responses. Neurosci. Lett. 260, 133–136.

de Gelder, B., Pourtois, G., Weiskrantz, L., 2002. Fear recognition in the voice is modulated by unconsciously recognized facial expressions but not by unconsciously recognized affective pictures. Proc. Natl. Acad. Sci. U. S. A. 99, 4121–4126 %19.

de Gelder, B., Vroomen, J., Annen, L., Masthof, E., Hodiamont, P., 2003. Audio-visual integration in schizophrenia. Schizophr. Res. 59, 211–218.

de Gelder, B., Morris, J.S., Dolan, R.J., 2005a. Unconscious fear influences emotional awareness of faces and voices. Proc. Natl. Acad. Sci. U. S. A. 102, 18682–18687 %20.

de Gelder, B., Vroomen, J., de Jong, S.J., Masthoff, E.D., Trompenaars, F.J., Hodiamont, P., 2005b. Multisensory integration of emotional faces and voices in schizophrenics. Schizophr. Res. 72, 195–203.

Dolan, R.J., Morris, J.S., de Gelder, B., 2001. Crossmodal binding of fear in voice and face. Proc. Natl. Acad. Sci. U. S. A. 98, 10006–10010.

Ekman, P., Friessen, W.V., 1976. Measuring facial movement. J. Environ. Psychol. Non-Verbal Behav. 1, 56–75.

Foxe, J.J., Schroeder, C.E., 2005. The case for feedforward multisensory convergence during early cortical processing. Neuroreport 16, 419–423.

Giard, M.H., Peronnet, F., 1999. Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. J. Cogn. Neurosci. 11, 473–490.

Green, M.F., Kern, R.S., Braff, D.L., Mintz, J., 2000. Neurocognitive deficits and functional outcome in schizophrenia: are we measuring the "right stuff"? Schizophr. Bull. 26, 119–136.

Green, M.F., Nuechterlein, K.H., Gold, J.M., Barch, D.M., Cohen, J., Essock, S., Fenton, W.S., Frese, F., Goldberg, T.E., Heaton, R.K., Keefe, R.S., Kern, R.S., Kraemer, H., Stover, E., Weinberger, D.R., Zalcman, S., Marder, S.R., 2004. Approaching a consensus cognitive battery for clinical trials in schizophrenia: the NIMH-MATRICS conference to select cognitive domains and test criteria. Biol. Psychiatry 56, 301–307.

Green, M.F., Penn, D.L., Bentall, R., Carpenter, W.T., Gaebel, W., Gur, R.C., Kring, A.M., Park, S., Silverstein, S.M., Heinssen, R., 2008. Social cognition in schizophrenia: an NIMH workshop on definitions, assessment, and research opportunities. Schizophr. Bull.

Grillon, C., Ameli, R., Charney, D.S., Krystal, J., Braff, D., 1992. Startle gating deficits occur across prepulse intensities in schizophrenic patients. Biol. Psychiatry. 32, 939–943.

Hershenson, M., 1962. Reaction time as a measure of intersensory facilitation. J. Exp. Psychol. 63, 289–293.

Jones, J.A., Callan, D.E., 2003. Brain activity during audiovisual speech perception: an fMRI study of the McGurk effect. Neuroreport 14, 1129–1133.

Kay, S.R., Fiszbein, A., Opler, L.A., 1987. The positive and negative syndrome scale (PANSS) for schizophrenia. Schizophr. Bull. 13, 261–276.

Kreifelts, B., Ethofer, T., Grodd, W., Erb, M., Wildgruber, D., 2007. Audiovisual integration of emotional signals in voice and face: an event-related fMRI study. Neuroimage 37, 1445–1456.

Lezak, M.D., Howieson, D.B., Loring, D.W., 2004. Orientation and attention. In: Lezak, M.D., Howieson, D.B., Loring, D.W. (Eds.), Neuropsychological Assessment. Oxford University Press, New York, pp. 337–374.

Light, G.A., Braff, D.L., 2005. Mismatch negativity deficits are associated with poor functioning in schizophrenia patients. Arch. Gen. Psychiatry 62, 127–136.

Macaluso, E., George, N., Dolan, R., Spence, C., Driver, J., 2004. Spatial and temporal factors during processing of audiovisual speech: a PET study. Neuroimage 21, 725–732.

MacLeod, C.M., MacDonald, P.A., 2000. Interdimensional interference in the Stroop effect: uncovering the cognitive and neural anatomy of attention. Trends Cogn. Sci. 4, 383–391.

McGurk, H., MacDonald, J., 1976. Hearing lips and seeing voices. Nature 264, 746–748.

Meeren, H.K., van Heijnsbergen, C.C., de Gelder, B., 2005. Rapid perceptual integration of facial expression and emotional body language. Proc. Natl. Acad. Sci. U. S. A. 102, 16518–16523.

Meredith, M.A., Stein, B.E., 1986. Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. J. Neurophysiol. 56, 640–662.

Nose, M., Tansella, M., Thornicroft, G., Schene, A., Becker, T., Veronese, A., Leese, M., Koeter, M., Angermeyer, M., Barbui, C., 2008. Is the defined daily dose system a reliable tool for standardizing antipsychotic dosages? Int. Clin. Psychopharmacol. 23, 287–290.

Pichot, P., Lebeaux, M.O., Penhouët, C., Simon, M., 1993. Comparaison entre la méthode de Barona et l'Index de détérioration de Wechsler. Revue Européenne de Psychologie Appliquée 43, 291–301.

Pourtois, G., de Gelder, B., Vroomen, J., Rossion, B., Crommelinck, M., 2000. The time-course of intermodal binding between seeing and hearing affective information. Neuroreport 11, 1329–1333.

Pourtois, G., de Gelder, B., Bol, A., Crommelinck, M., 2005. Perception of facial expressions and voices and of their combination in the human brain. Cortex 41, 49–59.

Radeau, M., 1994. Auditory-visual spatial interaction and modularity. Curr. Psychol. Cogn. 13, 117–123.

Ross, L.A., Saint-Amour, D., Leavitt, V.M., Molholm, S., Javitt, D.C., Foxe, J.J., 2007. Impaired multisensory processing in schizophrenia: deficits in the visual enhancement of speech comprehension under noisy environmental conditions. Schizophr. Res. 97, 173–183.

Schall, U., Schon, A., Zerbin, D., Eggers, C., Oades, R.D., 1996. Event-related potentials during an auditory discrimination with prepulse inhibition in patients with schizophrenia, obsessive-compulsive disorder and healthy subjects. Int. J. Neurosci. 84, 15–33.

Schroeder, C.E., Foxe, J.J., 2002. The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. Brain Res. Cogn. Brain Res. 14, 187–198.

Schroeder, C.E., Lindsley, R.W., Specht, C., Marcovici, A., Smiley, J.F., Javitt, D.C., 2001. Somatosensory input to auditory association cortex in the macaque monkey. J. Neurophysiol. 85, 1322–1327.

Stein, B.E., Huneycutt, W.S., Meredith, M.A., 1988. Neurons and behavior: the same rules of multisensory integration apply. Brain Res. 448, 355–358.

Surguladze, S.A., Calvert, G.A., Brammer, M.J., Campbell, R., Bullmore, E.T., Giampietro, V., David, A.S., 2001. Audio-visual speech perception in schizophrenia: an fMRI study. Psychiatry Res. 106, 1–14.

Swerdlow, N.R., Light, G.A., Cadenhead, K.S., Sprock, J., Hsieh, M.H., Braff, D.L., 2006. Startle gating deficits in a large cohort of patients with schizophrenia: relationship to medications, symptoms, neurocognition, and level of function. Arch. Gen. Psychiatry 63, 1325–1335.

Van den Stock, J., Righart, R., de Gelder, B., 2007. Body expressions influence recognition of emotions in the face and voice. Emotion 7, 487–494.

Velligan, D.I., Mahurin, R.K., Diamond, P.L., Hazleton, B.C., Eckert, S.L., Miller, A.L., 1997. The functional significance of symptomatology and cognitive function in schizophrenia. Schizophr. Res. 25, 21–31.

Velligan, D.I., Bow-Thomas, C.C., Mahurin, R.K., Miller, A.L., Halgunseth, L.C., 2000. Do specific neurocognitive deficits predict specific domains of community function in schizophrenia? J. Nerv. Ment. Dis. 188, 518–524.

Voort, L.V., Senf, G.M., Benton, A.L., 1972. Development of audiovisual integration in normal and retarded readers. Child Dev. 43, 1260–1272.

Vroomen, J., Driver, J., de Gelder, B., 2001. Is cross-modal integration of emotional expressions independent of attentional resources? Cogn. Affect. Behav. Neurosci. 1, 382–387.

WHO, 2008. http://www.searo.who.int/en/section1174/section1199/section1567_6744.htm. internet. Ref Type: Electronic Citation.

WHO Collaborating Center for Drug Statistic Methodology, 2003. Guidelines for ATC Classification and DDD Assignment. Oslo.