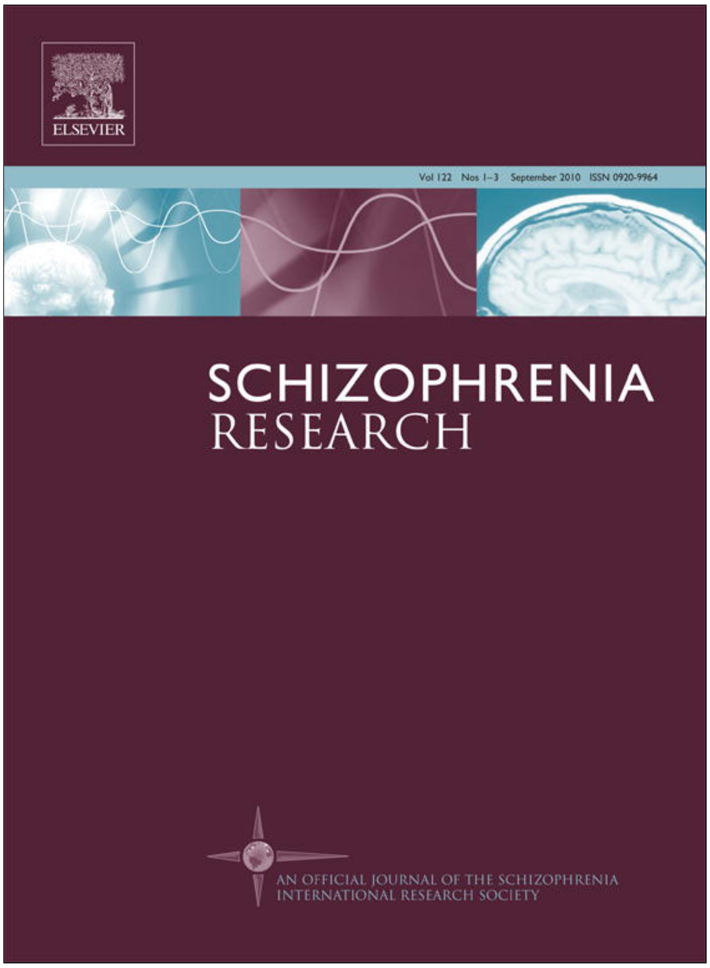
ELSEVIER
Vol 122 Nos 1–3 September 2010 ISSN 0920-9964
SCHIZOPHRENIA
RESEARCH
AN OFFICIAL JOURNAL OF THE SCHIZOPHRENIA
INTERNATIONAL RESEARCH SOCIETY

Schizophrenia Research 122 (2010) 136–143

Contents lists available at ScienceDirect

Schizophrenia Research

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

Modality-specific attention and multisensory integration of emotions in schizophrenia: Reduced regulatory effects

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

[a] Cognitive Neuroscience Laboratory, Tilburg University, P.O. Box 90153, 5000 LE, Tilburg, The Netherlands
[b] GGz Breburg Groep, P.O. Box 770, 5000 AT, Tilburg, The Netherlands
[c] Department of Developmental, Clinical and Cross-cultural Psychology, Tilburg University, P.O. Box 90153, 5000 LE, The Netherlands
[d] Martinos Center for Biomedical Imaging, Massachusetts General Hospital and Harvard Medical School, Building 36, First Street, Charlestown, MA 02129, USA

A R T I C L E I N F O

Article history:
Received 16 February 2010
Received in revised form 19 April 2010
Accepted 21 April 2010

Keywords:
Emotion recognition
Multisensory integration
Selective attention
Modality-specific attention
Cognition
Schizophrenia

A B S T R A C T

Background: Deficits in emotion perception are a well-established phenomenon in schizophrenic patients and studies have typically used *unimodal* emotion tasks, presenting either emotional faces or emotional vocalizations. We introduced *bimodal* emotion conditions in two previous studies in order to study the process of multisensory integration of visible and audible emotion cues. We now build on our earlier work and address the regulatory effects of selective attention mechanisms on the ability to integrate emotion cues stemming from multisensory channels.
Methods: We added a neutral secondary distractor condition to the original bimodal paradigm in order to investigate modality-specific selective attention mechanisms. We compared schizophrenic patients ($n = 50$) to non-schizophrenic psychotic patients ($n = 46$), as well as to healthy controls ($n = 50$). A trained psychiatrist used the Schedules of Clinical Assessment in Neuropsychiatry (SCAN 2.1) to diagnose the patients.
Results: As expected, in healthy controls, and to a lesser extent in non-schizophrenic psychotic patients, modality-specific attention attenuated multisensory integration of emotional faces and vocalizations. Conversely, in schizophrenic patients, auditory and visual distractor conditions yielded unaffected and even exaggerated multisensory integration.
Conclusions: These results suggest that schizophrenics, as compared to healthy controls and non-schizophrenic psychotic patients, have modality-specific attention deficits when attempting to integrate information regarding emotions that stem from multichannel sources. Various explanations for our findings, as well as their possible consequences, are discussed.

1. Introduction

Schizophrenia is characterized by pervasive cognitive and social impairments, which has led to the concept of *social* cognition deficit (Couture et al., 2006; Green et al., 2000, 2004, 2005; Velligan et al., 1997, 2000). Emotion perception represents an essential part of this concept. Numerous studies have demonstrated deficits in emotion perception in schizophrenia and these deficits have been proposed as intermediates between neurocognitive disturbances and impaired social outcomes (Edwards et al., 2002; Kohler et al., 2000; Mandal et al., 1998).

Studies that have investigated the effects of emotion perception have almost exclusively used *unimodal* tasks in which participants were presented with either visual *or* auditory emotionally laden stimuli (Edwards et al., 2002). In contrast, de Gelder et al. (2005b) and de Jong et al. (2009) used *bimodal* tasks in which emotionally laden visual and auditory stimuli were presented simultaneously. The emotions in the two stimuli were either matched or unmatched.

* 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 addresses: s.dejong@ggzbreburggroep.nl (J.J. de Jong), p.hodiamont@uvt.nl (P.P.G. Hodiamont), degelder@nmr.mgh.harvard.edu (B. de Gelder).

0920-9964/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.schres.2010.04.010

Doing so provided the opportunity to study the naturally occurring process of *multisensory integration* of stimuli. The results of the two studies showed that schizophrenic patients were significantly less able to integrate facial and vocal emotions compared to non-schizophrenic psychotics, and to healthy people.

However, neither of these two studies addressed the important issue of selective attention mechanisms. Normally, spatial and modality-specific selective attention will affect multisensory integration in a top–down manner. Whereas the integration of matching vs. non-matching information has resulted in behavioral and electrophysiological enhancements and decrements, such consequences have typically been attenuated when attention was selectively directed towards a specific location or sensory channel (Alsius et al., 2005; Mozolic et al., 2008; Turatto et al., 2002). Therefore, the purpose of the current study was to investigate how modality-specific attention regulates the multisensory integration of emotion in schizophrenic patients. We compared schizophrenic to non-schizophrenic psychotic patients and to healthy controls.

Crossmodal influence between sensory channels is a natural and automatic part of multisensory integration, and behavioral and neural performances are facilitated by this phenomenon (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). For example, integrating visible lip movements with audible speech improves *listening*, whereas integrating facial with prosodic emotion improves *understanding* (Dolan et al., 2001; de Gelder et al., 2002, 2005a; Meeren et al., 2005; Pourtois et al., 2000, 2005; van den Stock et al., 2007). These findings underscore the importance of multisensory integration for adaptive behavior.

The most compelling task for the brain, when combining separate sensory stimuli into a single event, is to discard the overabundance of irrelevant multisensory stimuli. Selective attention mechanisms enhance integrating stimuli that are biologically relevant while suppressing stimuli that do not convey the same event (de Gelder, 2000). We presented participants with emotionally laden faces and vocalizations in order to study multisensory integration. We then used visual and auditory distractors to create different conditions of modality-specific attention. We hypothesized that the regulatory effects of modality-specific attention on multisensory integration of emotions would discriminate between schizophrenics, non-schizophrenic psychotic patients, and healthy people.

2. Materials and methods

2.1. Participants

One hundred and one outpatients from a local psychiatric hospital and 50 neurologically and psychiatrically healthy controls (*Ctrl*) participated in the study. A trained psychiatrist examined the outpatients using the Schedules of the Clinical Assessment in Neuropsychiatry (SCAN 2.1) and diagnosed 55 patients as having schizophrenia (*Sch*), and 46 patients as having a non-schizophrenic psychosis (*N-Sch-Psy*). Participants gave their informed consent and were paid for their participation, and the study was approved by the regional Medical Ethics Committee. Participants in this study were identical to those used by de Jong et al. (2009) and the data for this study were collected concurrently with those of the de Jong study. However, whereas the de Jong et al. study focused exclusively on the multisensory processing of facial and vocal emotions, the current study focused on how modality-specific attention regulates the multisensory integration of emotions.

DSM-IV classifications are displayed in Table 1. Table 2 shows demographic and clinical variables. The three groups did not differ significantly on sex ratio, $\chi^2(2,151) = 5.90$, $p = 0.052$; educational level, $\chi^2(6,151) = 11.32$, $p = 0.079$; nor handedness $\chi^2(2,151) = 0.24$, $p = 0.888$. Age differences were, however, significant, $F(2,150) = 7.59$, $p = 0.001$, and mean ages were as follows: *Sch* (33.53), *N-Sch-Psy* (35.22) and *Ctrl* (41.16). Moreover, *Sch* scored significantly higher than did *N-Psy-Sch* on positive, $F(1,100) = 258.60$, $p = 0.001$; negative, $F(1,100) = 480.35$, $p = 0.001$; and total PANSS-scores $F(1,100) = 2882.13$, $p = 0.001$.

2.2. Materials and procedure

Participants were presented simultaneously with a face on a computer screen and a short vocalization, and then asked to rate the emotion of the vocalization. The faces, as well as the

Table 1
DSM-IV classifications within both patient groups (schizophrenic patients and non-schizophrenia psychosis patients).

	Schizophrenia subjects	Non-schizophrenia 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

Table 2
Demographic, clinical (PANSS) and neuropsychological (CPT) characteristics of the three groups of subjects (schizophrenic patients, non-schizophrenia psychosis patients and controls).

	Schizophrenia subjects	Non-schizophrenia 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		$p = 0.001$

[a] ANOVA.
[b] Chi-square.
[c] Significantly different from controls, not the other patient group.

vocalizations, depicted either a happy, sad, or fearful emotion, and were extensively tested in a pilot-study to ensure that the emotions were unambiguous and easily discernible from one another. The vocalizations consisted of one of the following four short phrases: "bought a car"; "to Amsterdam"; "been to hair dresser"; and "by airplane". Two task sets were employed: Task Set 1 and Task Set 2. The target emotions in Task Set 1 were happy or sad, and the emotions in the visual and auditory stimuli were either matched or mismatched. Matching meant combining a happy face with a happy vocalization or a sad face with a sad vocalization and mismatching meant combining a sad face with a happy vocalization or happy face with a sad vocalization. The target emotions in Task Set 2 were happy or fearful. Participants pressed one of two buttons as fast as possible to indicate whether the emotion in the vocalization was happy or sad, or happy or fearful. Participants were asked explicitly to watch the computer screen while rating the vocalization, but to ignore the facial emotion.

Faces were taken from Ekman and Friessen (1976) and were presented for 800 ms. Vocalizations were semantically neutral and were obtained by instructing four professional actors (two males and two females) to pronounce the phrases as if they were happy, sad, or afraid; and durations ranged from 599 to 1265 ms.

Whereas the original study by de Jong et al. (2009) included only paired combinations of different facial and auditory emotions, the current study included secondary visual and auditory distractors, thus creating four additional trials (two for each Task Set) that represented conditions of modality-specific attention. The 64 stimuli for each trial were presented in pseudorandom order — 32 trials used matched visual/auditory stimuli and 32 trials used mismatched stimuli.

The visual distractor consisted of two black squares that were 1.0×1.0 cm ($2.0 \times 2.0°$) with a white digit (randomly presented as either a '6' or an '8') in the centre. The digits were projected randomly for 800 ms between the eyebrows of either a happy, sad, or fearful face. Participants pressed a button to rate the emotion of the vocalization, and were then immediately prompted on-screen by the question: "was there an 8 in the face?" The auditory distractor consisted of a pair of tones, each 300 ms in duration and separated by an interstimulus duration of 200 ms. In the first pair, each tone had a frequency of 500 Hz, whereas in the second pair the first tone had a frequency of 500 Hz and the second tone had a frequency of 540 Hz. The pairs of tones were presented randomly and their onset coincided with that of the faces and vocalizations. Participants first rated the emotion in the vocalization and were then asked to respond to the question, "Did you hear a high tone?" by pressing a 'yes' or a 'no' button as fast as possible.

To summarize, each participant rated the emotion of a vocalization in the presence of a matching or mismatching facial expression and either with or without a secondary distracting condition that was presented concurrently. Each participant worked through six different trials in a fixed order. Trial 1 used happy vs. sad emotions in visual/auditory combinations (Task Set 1). Trial 2 did the same, but pitted happy vs. fearful emotions against one another (Task Set 2). Trial 3 combined a visual distractor under Task Set 1 conditions whereas trial 4 combined a visual distractor under Task Set 2 conditions. Trials 5 and 6 combined an auditory distractor under Task Sets 1 and 2, respectively. A graphic illustration of the stimuli used within the different Task conditions is presented in Table 3.

2.3. Statistical procedures and outlier management

Performance accuracy was the proportion of correct responses. Accuracy was determined for matched and mismatched face/voice pairs respectively, yielding two within-subject variables for each trial. De Jong et al. (2009) indicated a *Matching*×*Group* interaction and reduced emotional multisensory integration in schizophrenics. *Matching* was defined by within-subject accuracy rates on matched and mismatched emotion items, reflecting the crossmodal influence of the facial emotions on the perceptions of emotion in the vocalizations. In this study, *Matching*×*Task*×*Group* interactions were explored to test our hypothesis and the *Task* variable reflected different conditions of modality-

Table 3
Graphical illustration of presented stimuli (columns) within each Task condition (rows).

Emotional voice[1]	Emotional face[1]	Visual/Auditory distractor	Task-condition
"bought a car"		None	Without distractor
"been to hair-dresser"		6 or 8 (presented between the eyebrows)	Visual distractor
"by airplane"		♪ ♪ 500 Hz 500 Hz or ♪ ♪ 500 Hz 540 Hz	Auditory distractor

[1]In Task Set 1, target emotions were Happy and Sad, in Task Set 2, target emotions were Happy and Fear.

specific attention. We used SPSS 15.0 and univariate analyses with our repeated measures design.

Outliers were managed with the same procedure and criteria as described in de Jong et al. (2009). Some subjects appeared to have judged facial emotion instead of the required vocal emotion, resulting in normal accuracy rates for matched, but extremely low accuracy rates for mismatched face/voice pairs. In such cases, differences between *emotion-congruent* and *emotion-incongruent* conditions amounted to more than 0.5, which was considered to be an appropriate cutoff for outlier data. Furthermore, a few subjects appeared to have reversed the yes/no response keys throughout the task. The accuracy rates for matching face/voice pairs in these instances also fell below the chance-level (0.5) and such data were also considered as outliers. The number of discarded outliers across the six trials was 16, 15, and 6 for schizophrenics, non-schizophrenic psychotics, and controls, respectively $\chi^2(2,151) = 6.49$, $p = 0.039$. Outliers were treated on a task-by-task basis during the analysis and no participant data was excluded entirely.

We conducted three main analyses: within-group effects of *Task* were plotted and analyzed for Task Set 1 and Task Set 2 (see Sections 3.1 and 3.2), after which an omnibus *between*-group analysis across both Task Sets was conducted (see Section 3.3).

3. Results

3.1. Happy/sad

Scores for Task Set 1 have been plotted in Fig. 1 in order to display the within-group effects of varying levels of modality-specific attention. Crossmodal impact was affected by Task condition only for the control group, whereas the schizophrenic and psychosis groups each showed parallel lines across the different conditions. The slope of the visual distractor condition was attenuated for the control group. A 2×3 ANOVA for each group, with *Matching* and *Task* as within-subject factors, was conducted to explore these findings further. *Matching* and *Task* interacted significantly $F(2,92) = 5.39$, $p = 0.006$ for the control group, however this interaction was non-significant for the schizophrenic, $F(2,90) = 0.07$, $p = 0.93$ and psychosis, $F(2,72) = 0.29$, $p = 0.75$ groups.

3.2. Happy/fear

Scores for the various Task Set 2 distractor conditions have been plotted in Fig. 2. In *Ctrl* and *N-Sch-Psy*, crossmodal impact was attenuated in the visual distractor condition. Again, adding a visual distractor did not alter the original

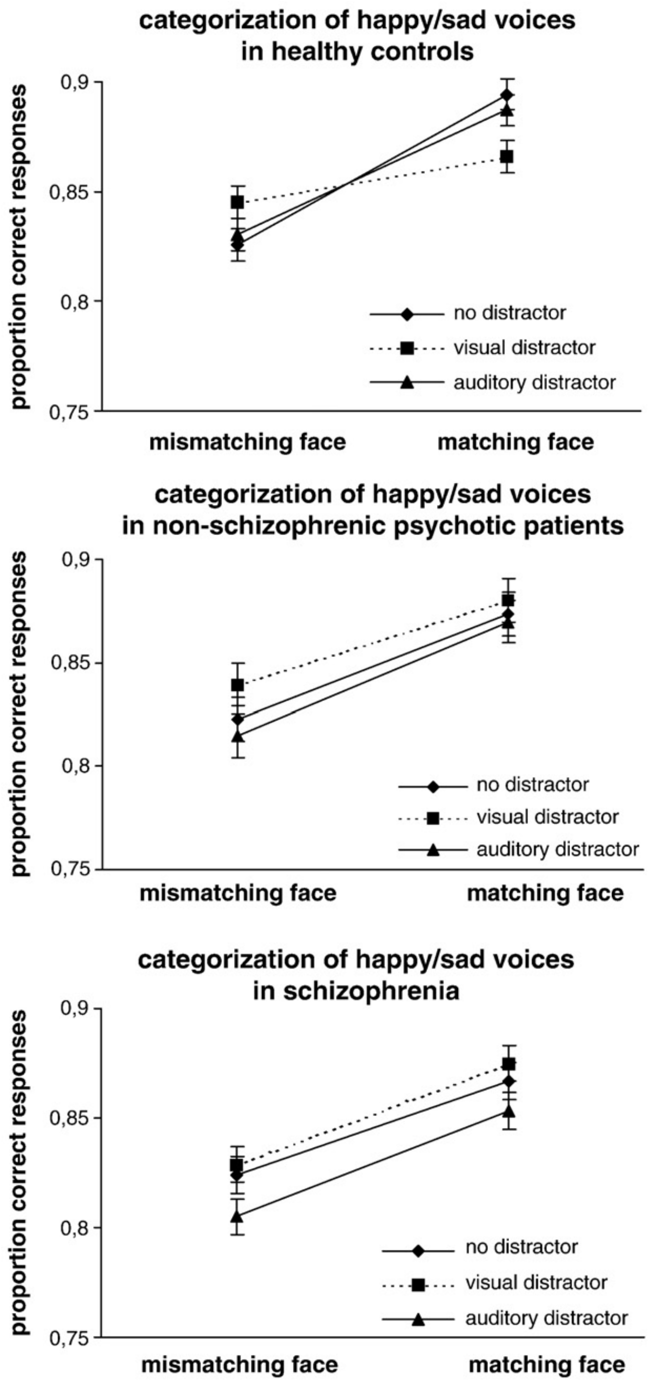

Fig. 1. *Ctrl* categorization of happy and sad voices in healthy controls, displayed for each task condition. The flattened slope of the line that represents the visual distractor condition illustrates a diminished cross-modal impact of facial on vocal emotion. *N-Sch-Psy* categorization of happy and sad voices in non-schizophrenic psychotic patients, displayed for each task condition. The nearly parallel slopes of the three lines indicate that the distractor conditions do not affect initial crossmodal impact of facial on vocal emotion. *Sch* categorization of happy and sad voices in schizophrenia, displayed for each task condition. The parallel slopes of the three lines indicate that the crossmodal impact of facial on vocal emotion does not vary with conditions of modality-specific attention.

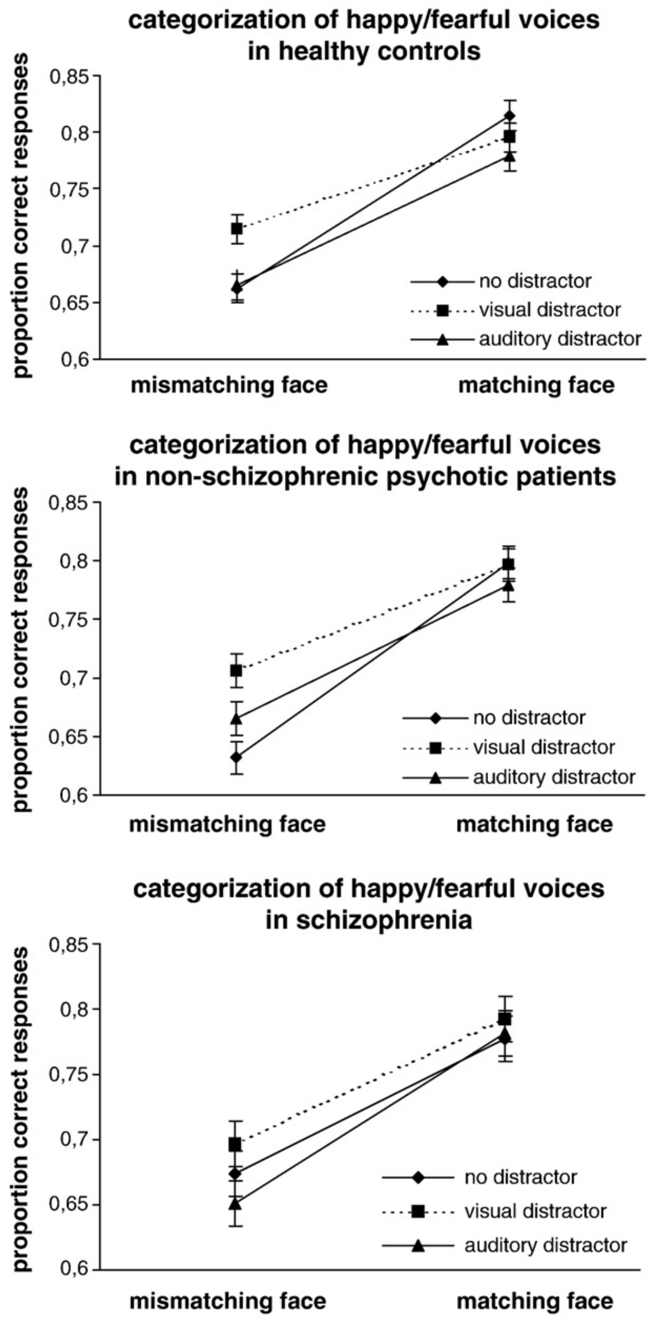

Fig. 2. *Ctrl* categorization of happy and fearful voices in healthy controls, displayed for each task condition. The flattened slopes of the lines that represent visual and auditory distractor conditions illustrate attenuated crossmodal impact of facial on vocal emotion. *N-Sch-Psy* categorization of happy and fearful voices in non-schizophrenic psychotic patients, displayed for each Task condition. The slopes of the lines that reflect distractor conditions are more horizontal than the line that represents the no-distractor condition. *Sch* categorization of happy and fearful voices in schizophrenia, displayed for each task condition. The slopes of the lines that reflect the no-distractor and the visual distractor conditions are parallel. In contrast to the other groups, the line that reflects the auditory distractor condition is steeper, thereby illustrating an exaggerated crossmodal impact of facial on vocal emotion.

MSI pattern in *Sch*. In contrast, the line that reflected the auditory distractor condition was steeper for *Sch* but not for *Ctrl* nor for *N-Sch-Psy*. Further analyses revealed significant interaction effects for *Ctrl*, $F(2,90) = 5.24$, $p = 0.007$ and for *N-Sch-Psy*, $F(2,64) = 6.85$, $p = 0.002$. Crossmodal impact did not interact significantly with the Task condition for schizophrenics $F(2,88) = 1.55$, $p = 0.22$. Therefore, an exaggerated multisensory integration effect, as seen in Fig. 2-*Sch* when schizophrenic participants were distracted by an auditory stimulus, was not confirmed here.

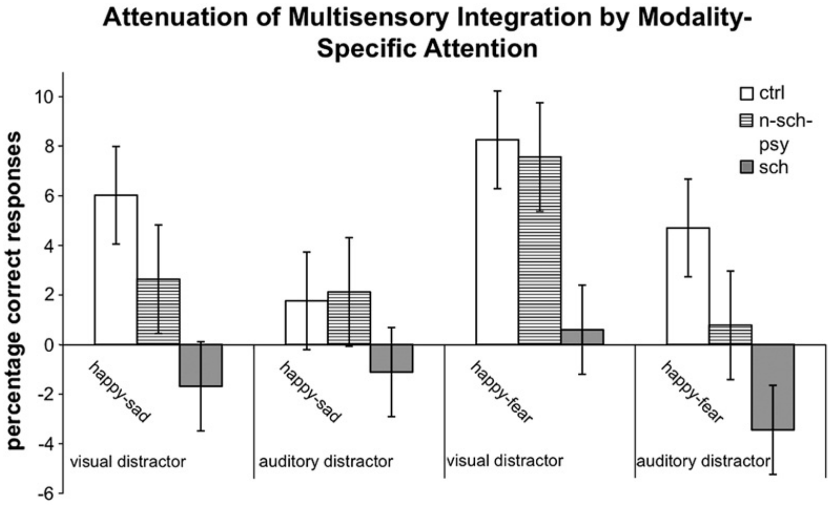

Fig. 3. Regulation of multisensory integration of emotions by modality-specific attention conditions is measured as *crossmodal impact without distractor* minus *crossmodal impact with auditory/visual distractor*, displayed for happy/sad and happy/fear.

3.3. Omnibus analyses

To explore our hypothesis of between-group differences more directly, we performed an omnibus analysis across both Task Sets (happy/sad and happy/fear) with *Matching* and *Task* as within-subject factors and *Group* as a between-subject factor. The main factor, *Group*, was not statistically significant, $F(2,111)=0.48$, $p=0.62$, which reflected the use of unambiguous emotional stimuli. After all, our purpose was not to measure between-group differences in emotion perception abilities. The interaction among *Matching*, *Task*, and *Group* was significant, $F(4,222)=2.72$, $p=0.031$, and revealed that multisensory integration varied differentially between groups across the modality-specific attention conditions. We repeated this analysis with covariates of *Age*, *Sex*, and *Education* entered as a single set and discovered that the interaction among *Matching*, *Task* and *Group* became more significant, $F(4,220)=3.64$, $p=0.007$.

To explore further the interaction among *Matching*, *Task* and *Group*, we repeated the analysis twice with two instead of three distractor conditions: first targeting the effect of the visual distractor, by removing the auditory distractor effect; and second targeting the effect of the auditory distractor, by removing the visual distractor. The interaction among *Congruence*, *Task*, and *Group*, controlling for the set of three covariates, was significant for the visual $F(2,120)=4.06$, $p=0.020$, as well as the auditory $F(2,120)=4.21$, $p=0.017$ distractors. This time, the impression from Fig. 2-*Sch* of an exaggerated integration in *Sch* during auditory distraction was confirmed statistically. The lack of statistical significance mentioned above, when all *Task*-levels were included, was probably a result of opposing trends between the auditory and visual distractor effects.

Fig. 3 shows how visual and auditory distractors differentially affected initial crossmodal impact among groups. The visual distractor attenuated multisensory integration in *Ctrl* and in *N-Sch-Psy*, but not in *Sch*, for the happy/sad and the happy/fear Task Sets. Moreover, within the auditory distractor condition, the difference in response set between *Ctrl* and *Sch* becomes even larger as one moves from the happy/sad Task set to the happy/fear Task Set, namely, multisensory integration becomes even more attenuated in *Ctrl* and even more exaggerated in *Sch*.

Finally, performance on the distraction tasks was analyzed. Mean accuracy rates, when responding to the visual distractor ("Was there an 8 in it?"), registered 0.97, 0.96, and 0.95 for *Ctrl*, *N-Sch-Psy*, and *Sch* respectively, and $F(2,141)=1.09$, $p=0.339$. Accuracy rates, when responding to the auditory distractor ("Did you hear a high tone?"), were 0.91, 0.87, and 0.82 respectively, and $F(2,145)=3.72$, $p=0.027$. A post-hoc Tukey test showed that *Sch* scores were significantly lower than *Ctrl* scores, and that *N-Sch-Psy* scores were not significantly different from either *Sch* or *Ctrl* scores.

4. Discussion

Our results confirm the hypothesis that the regulatory effects of modality-specific attention on the multisensory integration of facial and vocal affects discriminate among schizophrenics, non-schizophrenic psychotic patients, and healthy people. The visual distractor attenuates integration patterns for *Ctrl* and for *N-Sch-Psy*, but not for *Sch*, in the happy/sad as well as the happy/fear task sets. Moreover, the auditory distractor also diminishes integration patterns for *Ctrl*, but not for *N-Sch-Psy*, in the happy/fear task set. In contrast with *Ctrl*, a stronger impact of facial on vocal emotion perception occurs in *Sch*. That the trend is in the same direction for happy/sad suggests that this finding is consistent.

When we attempt to explain the results of happy/sad, the concept of 'competition for attentional resources' comes to mind. Research on modality-specific attention shows diminished processing of stimuli from the unattended modality (Johnson and Zatorre, 2006; Laurienti et al., 2002; Macaluso et al., 2000; Spence et al., 2001). When, in our trials, a visual distractor competes with facial stimuli for visual resources, facial emotions become less available for crossmodal binding. A decrease of the detrimental effect of mismatching faces, as well as a reduction in initial performance gains by matching faces can be expected (see Fig. 1-*Ctrl*). That the visual distractor had no effect in *N-Sch-Psy* and *Sch* may be explained by the fact that basic task demands are already quite high for patients. This would result in a *high-load* driven, diminished attention situation to emotional faces as compared to *Ctrl*. *Ctrl* participants *require* a distractor to ignore facial emotion.

One would expect the same patterns in the happy/fear Task Set. Fig. 2-*Ctrl* and Fig. 2-*N-Sch-Psy* indeed show diminished detrimental effects when visual and auditory cues are mismatched, but unaffected performance gains when the cues are matched. These results require an alternative explanation that should take into account the multidimensionality of the emotions involved. Selective attention studies describe a fronto-parietal brain network that is differentially activated depending on whether modality-specific attention cues are semantically congruent with target stimuli (Talsma et al., 2006; Talsma et al., 2008). One might expect that different emotions also activate different response patterns, representing a higher-order model of the supramodal, regulatory mechanism of selective attention, rather than, or along with, the lower-order model of attentional resources.

A second explanation for our results considers, paradoxically, unisensory perception deficits. Whereas we investigated *multisensory* integration of emotions, early *unisensory* processing deficits cannot be ruled out. Numerous studies, using stimuli that are comparable to our distractors, indicate early perception deficits in schizophrenia. In a pitch discrimination study, mean required difference in pitch (Δf) ranged from 8%, for a subgroup of long-term outpatients, to 20% for residential-care patients (Rabinowicz et al., 2000). We use auditory distractor stimuli with a Δf of 8% (500 and 540 Hz) in a demanding dual-task design. This may explain the significantly diminished performance of *Sch* when rating auditory distrators. More importantly, our finding of exaggerated crossmodal influence during auditory distraction for *Sch*, as opposed to attenuated integration for *Ctrl*, can be linked to fundamental auditory processing deficits. This result might translate to common subjective experiences of patients when they feel simultaneously overpowered by environmental information and by disturbing encounters with others.

Early visual dysfunction, including impairments in magnocellular/parvocellular interactions, is also reported in schizophrenia (Butler et al., 2003; Butler et al., 2007). One might expect that simultaneously presenting a *whole* emotional face, and distractors *within* the face will result in an aberrant response in *Sch*. This may explain the diminished regulatory effects for *Sch*, as compared to *Ctrl* and *N-Sch-Psy*, in both task sets during the visual distractor condition.

This is the first study that considers the top–down effects of modality-specific attention on the multisensory integration of emotions in schizophrenia. Although the ability to integrate matching information from different channels serves an adaptive purpose, the top–down, selective attention mechanisms that suppress multisensory processing, when there is too much, or conflicting information, are also crucial. Our data indicate that the regulatory effects of modality-specific attention on the integration of facial and vocal affects are deficient in schizophrenia, whereas intermediate deficits are shown in non-schizophrenic psychotic patients. Although paradigms like the one used in the present study advance our knowledge, task-oriented approaches with electrophysiological and imaging methods would shed further light on the neural basis of schizophrenia.

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 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.

Acknowledgments

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

Furthermore, we wish to thank Uco Wiersma for his comments and insights.

References

Alsius, A., Navarra, J., Campbell, R., Soto-Faraco, S., 2005. Audiovisual integration of speech falters under high attention demands. Curr. Biol. 15, 839–843.

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.

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.

Butler, P.D., DeSanti, L.A., Maddox, J., Harkavy-Friedman, J.M., Amador, X.F., Goetz, R.R., Javitt, D.C., Gorman, J.M., 2003. Visual backward-masking deficits in schizophrenia: relationship to visual pathway function and symptomatology. Schizophr. Res. 59, 199–209.

Butler, P.D., Martinez, A., Foxe, J.J., Kim, D., Zemon, V., Silipo, G., Mahoney, J., Shpaner, M., Jalbrzikowski, M., Javitt, D.C., 2007. Subcortical visual dysfunction in schizophrenia drives secondary cortical impairments. Brain 130, 417–430.

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., 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.

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. 20 (102), 18682–18687.

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

de Jong, J.J., Hodiamont, P.P.G., van den Stockde, J., Gelder, B., 2009. Audiovisual emotion recognition in schizophrenia: reduced integration of facial and vocal affect. Schizophr. Res. 107, 286–293.

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.

Edwards, J., Jackson, H.J., Pattison, P.E., 2002. Emotion recognition via facial expression and affective prosody in schizophrenia: a methodological review. Clin. Psychol. Rev. 22, 789–832.

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

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., Olivier, B., Crawley, J.N., Penn, D.L., Silverstein, S., 2005. Social cognition in schizophrenia: recommendations from the measurement and treatment research to improve cognition in schizophrenia new approaches conference. Schizophr. Bull. 31, 882–887.
Hershenson, M., 1962. Reaction time as a measure of intersensory facilitation. J. Exp. Psychol. 63, 289–293.
Johnson, J.A., Zatorre, R.J., 2006. Neural substrates for dividing and focusing attention between simultaneous auditory and visual events. Neuroimage 31, 1673–1681.
Jones, J.A., Callan, D.E., 2003. Brain activity during audiovisual speech perception: an fMRI study of the McGurk effect. NeuroReport 14, 1129–1133.
Kohler, C.G., Bilker, W., Hagendoorn, M., Gur, R.E., Gur, R.C., 2000. Emotion recognition deficit in schizophrenia: association with symptomatology and cognition. Biol. Psychiatry 48, 127–136.
Laurienti, P.J., Burdette, J.H., Wallace, M.T., Yen, Y.F., Field, A.S., Stein, B.E., 2002. Deactivation of sensory-specific cortex by cross-modal stimuli. J. Cogn. Neurosci. 14, 420–429.
Macaluso, E., Frith, C.D., Driver, J., 2000. Modulation of human visual cortex by crossmodal spatial attention. Science 289, 1206–1208.
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.
Mandal, M.K., Pandey, R., Prasad, A.B., 1998. Facial expressions of emotions and schizophrenia: a review. Schizophr. Bull. 24, 399–412.
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.
Mozolic, J.L., Hugenschmidt, C.E., Peiffer, A.M., Laurienti, P.J., 2008. Modality-specific selective attention attenuates multisensory integration. Exp. Brain Res. 184, 39–52.
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.
Rabinowicz, E.F., Silipo, G., Goldman, R., Javitt, D.C., 2000. Auditory sensory dysfunction in schizophrenia: imprecision or distractibility? Arch. Gen. Psychiatry 57, 1149–1155.
Radeau, M., 1994. Auditory–visual spatial interaction and modularity. Curr. Psychol. Cogn. 13, 117–123.
Spence, C., Nicholls, M.E., Driver, J., 2001. The cost of expecting events in the wrong sensory modality. Percept. Psychophys. 63, 330–336.
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.
Talsma, D., Doty, T.J., Woldorff, M.G., 2006. Selective attention and audiovisual integration: is attending to both modalities a prerequisite for early integration? Cereb. Cortex 17, 679–690.
Talsma, D., Kok, A., Slagter, H.A., Cipriani, G., 2008. Attentional orienting across the sensory modalities. Brain Cogn. 66, 1–10.
Turatto, M., Benso, F., Galfano, G., Umilta, C., 2002. Nonspatial attentional shifts between audition and vision. J. Exp. Psychol. Hum. Percept. Perform. 28, 628–639.
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.