Deficient multisensory integration in schizophrenia: An event-related potential study

Jeroen J. Stekelenburg a,⁎, Jan Pieter Maes b, Arthur R. Van Gool b, Margriet Sitskoorn a, Jean Vroomen a

a Tilburg University, Department of Cognitive Neuropsychology, P.O. Box 90153, Warandelaan 2, 5000 LE Tilburg, The Netherlands
b Yulius Academy, Mathenesserlaan 202, 3014 RH Rotterdam, The Netherlands

ABSTRACT

Background: In many natural audiovisual events (e.g., the sight of a face articulating the syllable /ba/), the visual signal precedes the sound and thus allows observers to predict the onset and the content of the sound. In healthy adults, the N1 component of the event-related brain potential (ERP), reflecting neural activity associated with basic sound processing, is suppressed if a sound is accompanied by a video that reliably predicts sound onset. If the sound does not match the content of the video (e.g., hearing /ba/ while lipreading /fu/), the later occurring P2 component is affected. Here, we examined whether these visual information sources affect auditory processing in patients with schizophrenia.

Methods: The electroencephalography (EEG) was recorded in 18 patients with schizophrenia and compared with that of 18 healthy volunteers. As stimuli we used video recordings of natural actions in which visual information preceded and predicted the onset of the sound that was either congruent or incongruent with the video.

Results: For the healthy control group, visual information reduced the auditory-evoked N1 if compared to a sound-only condition, and stimulus-congruency affected the P2. This reduction in N1 was absent in patients with schizophrenia, and the congruency effect on the P2 was diminished. Distributed source estimations revealed deficits in the network subserving audiovisual integration in patients with schizophrenia.

Conclusions: The results show a deficit in multisensory processing in patients with schizophrenia and suggest that multisensory integration dysfunction may be an important and, to date, under-researched aspect of schizophrenia.

© 2013 Published by Elsevier B.V.

1. Introduction

One of the principal functions of the brain is to integrate signals from multiple modalities into coherent multisensory representations of objects and events. Intact multisensory integration implies that distant cortical and subcortical brain areas interact with each other, and for this reason investigators have begun to explore whether there is a specific role for multisensory integration in some of the perceptual deficits seen in disorders such as autism (Iarocci and McDonald, 2006; Kern et al., 2007) and schizophrenia (de Gelder et al., 2003). It has been hypothesized that disruptions in the communication between distant brain areas may hamper maintaining a coherent perception and understanding of the world, which in schizophrenia patients might contribute to the emergence of psychotic experience (Friston and Frith, 1995; Ford et al., 2002; Stephan et al., 2009).

There is a growing body of behavioral evidence showing that patients with schizophrenia do indeed have deficits in multisensory integration if compared to healthy controls. For example, viewing a speaker’s articulatory movements normally improves a listener’s ability to understand spoken words, especially under noisy environmental conditions (Sumby and Pollack, 1954). Patients with schizophrenia, however, show deficits in their ability to derive benefit from visual articulatory motion and are less influenced by lipread information when processing auditory speech (de Gelder et al., 2003; Ross et al., 2007; Pearl et al., 2009). Other abnormalities are found in the multisensory integration of emotions in face and voice (de Gelder et al., 2003; Ross et al., 2007; van Wassenhove et al., 2005; Stekelenburg and Vroomen, 2004; van Wassenhove et al., 2005; Stekelenburg and Vroomen, 2007, 2012). We recorded the electroencephalography (EEG) in patients with schizophrenia and healthy controls using an experimental paradigm specifically designed to tap audiovisual integration. As is commonly practiced in multisensory perception research, multisensory interactions are examined by comparing event-related potentials (ERPs) evoked by the bimodal stimuli with the sum of the neural activity of the unisensory stimuli. This additive model assumes that the neural activity evoked by audiovisual (AV) stimuli is equal to the sum of activities of the auditory (A) and visual (V) activity and its
associated audiovisual interactions (\(AV = A + V + [A \times V\) interactions]) (Gard and Peronnet, 1999). If the unimodal signals are processed independently, then the bimodal response equals the sum of unimodal responses (\(AV = A + V\)). If, however, the bimodal response differs (supra-additive or sub-additive) from the sum of the two unimodal responses, this is attributed to the interaction between the two modalities (Gard and Peronnet, 1999; Molholm et al., 2002; Klucharev et al., 2003; Besle et al., 2004; Teder-Sälejärvi et al., 2005; Stekelenburg and Vroomen, 2007; Vroomen and Stekelenburg, 2010). The stimuli in the current experiment were short video clips of natural human actions like a face speaking a syllable or a clap of two hands. Critical for our purpose is that in these stimuli – as in many natural audiovisual events – the visual signal precedes the sound for several tens up to hundreds of milliseconds. This allows observers to predict when and what sound will occur, and it thus allows us to examine audiovisual integration based on temporal information and informational content. Integration of audiovisual informational based of temporal information (when) has been associated with dampened, and some cases, speeded-up auditory-evoked N1 amplitude (Klucharev et al., 2003; Besle et al., 2004; van Wassenhove et al., 2005; Stekelenburg and Vroomen, 2007, 2012). The auditory N1 is a neural response elicited by audible transient auditory stimuli and reflects the basic encoding of acoustic information in the auditory cortex (Näätänen and Picton, 1987). The N1 has a negative deflection that peaks between 80 and 120 ms after sound onset and reaches its maximal value at the frontocentral electrodes. The N1 is generated by multiple brain areas of which the most prominent are located in the auditory cortex (Näätänen and Picton, 1987). The N1 is followed by the P2 component that can be functionally dissociated from the N1 (Crowley and Colrain, 2004). The functional interpretation of this suppression of the auditory N1 may be a reduction of auditory signal uncertainty, dampened sensation of loudness, or lowered computational demands for auditory brain areas.

Integration of audiovisual informational content (what) has been found to occur at later processing stages. More specifically, it has been found that the auditory-evoked P2 is more negative for incongruent (e.g., hearing /ba/ while lipreading /fu/) than congruent audiovisual pairings (hearing /ba/ and lipreading /ba/) (Stekelenburg and Vroomen, 2007). This distinction in audiovisual integration mechanisms allowed us to examine whether patients with schizophrenia show deficits at these early or late stages of audiovisual integration over and above well-known unimodal processing deficits like a reduced N1 amplitude to auditory stimuli (see for a review, Rosburg et al., 2008).

### 2. Method

#### 2.1. Participants

Eighteen patients with schizophrenia (1 female, mean age 38, SD 9) and 18 healthy control volunteers matched for gender and age (mean age 39, SD 8.1) participated after given written informed consent. Both groups did not significantly differ on educational level (Mann–Whitney U = 146.50, p = 0.63). Inclusion criteria for both psychiatric and non-psychiatric participants were: 18–55 years of age, no history of electroconvulsive treatment, no history of neurological illness, no history of alcohol or drug dependence or abuse within the last year, or long duration (>1 year) of past abuse, no medications which would grossly affect the EEG (e.g., barbiturates), normal hearing and normal or corrected-to-normal vision, an ability and desire to cooperate with our experimental procedures as evidenced by giving informed consent. Inclusion criteria for the schizophrenia participants were: patients who met DSM-IV-TR (APA, 2000) criteria for schizophrenia (\(n = 17\)) and schizoaffective disorder (\(n = 1\)), based both on chart information and on the relevant module of the Mini-International Neuropsychiatric Interview (M.I.N.I.), which is a short, structured diagnostic interview for DSM-IV-TR and ICD-10 disorders that is designed to perform a short but accurate structured psychiatric interview (Sheehan et al., 1998). Severity of the symptoms was assessed using the Dutch 24-item version of the Brief Psychiatric Rating Scale (BPRS). Patients scored on average 41.4 (SD 11.5) on the BPRS scale. The illness duration was 16.2 years (SD 5.6 years). All patients were receiving antipsychotic medication at the time of the study: sixteen were receiving atypical antipsychotics, and two a combination of two atypical antipsychotics (Table 1). All of them were naive to the purpose of the study. They received 35 Euro for their participation. The study was approved by the Medical Ethics Committee of the St. Elisabeth Hospital in Tilburg, the Netherlands, and was conducted in accordance with the Declaration of Helsinki.

#### 2.2. Stimuli and procedure

The experiment took place in a dimly-lit room. Visual stimuli were presented on a 17-inch monitor positioned at eye-level, 70 cm from the participant’s head. The sound came from a loudspeaker directly below the monitor. There were four different video clips: a speaking face articulating the syllables /bi/ or /fu/, a clap of two hands, and a tap of a spoon against a cup (for a full description of the stimuli see Stekelenburg and Vroomen, 2007). The inter-stimulus interval, measured from auditory onsets, was on average 3.7 s. The experimental conditions comprised of visual-only (V), auditory-only (A), audiovisual congruent (AVC), and audiovisual incongruent (AVI) stimuli presentations. The V condition showed one of the four videos, but without sound; the A condition presented one of the four sounds against a black background; the AVC condition showed the video recording with the original sound synchronized to the video. For incongruent AV pairings in the AVI condition, auditory /fu/ was combined with visual /bi/, auditory /bi/ with visual /fu/, auditory hand clapping with visual tapping of a spoon, and auditory tapping of a spoon with visual hand clapping. Note that the onset of the sound in the incongruent stimuli was synchronized to the onset of the sound in the original recordings, so the onset time was predictable in AVI, but not the content. For each condition (A, V, AVC, and AVI), 60 randomized trials for each of the 4 different stimuli were administered across 12 blocks. Testing lasted about 90 min (including short breaks between the blocks). To ensure that participants were looking at the video during stimulus presentation, they had to detect, by key press, the occasional occurrence of catch trials (13% of total number of trials). Catch trials contained a superimposed small plus sign (+) spot either between the lips and nose for speech stimuli, or at the collision site for

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Demographic and clinical characteristics of schizophrenia patients and healthy controls.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schizophrenia patients</td>
<td>Healthy controls</td>
</tr>
<tr>
<td>N</td>
<td>18</td>
</tr>
<tr>
<td>Gender</td>
<td>Male 17, Female 1</td>
</tr>
<tr>
<td>Age (years)</td>
<td>39.1 (8.2)</td>
</tr>
<tr>
<td>Handedness</td>
<td>15 R, 3 L</td>
</tr>
<tr>
<td>Education level</td>
<td>Elementary 1, Middle 13, Higher 4</td>
</tr>
<tr>
<td>Illness duration (years)</td>
<td>16.2 (5.6)</td>
</tr>
<tr>
<td>BPRS total score</td>
<td>41.4 (11.5)</td>
</tr>
<tr>
<td>Antipsychotic medication</td>
<td>None</td>
</tr>
<tr>
<td>Clozapine</td>
<td>7</td>
</tr>
<tr>
<td>Olanzapine</td>
<td>6</td>
</tr>
<tr>
<td>Risperidone</td>
<td>1</td>
</tr>
<tr>
<td>Quetiapine</td>
<td>1</td>
</tr>
<tr>
<td>Zuclopenthixol</td>
<td>1</td>
</tr>
<tr>
<td>Bromperidol</td>
<td>1</td>
</tr>
<tr>
<td>Haloperidol</td>
<td>1</td>
</tr>
</tbody>
</table>
non-speech stimuli for 120 ms or art the middle of the screen in the A-only condition. Catch trials occurred with equal probability across conditions.

2.3. ERP recording and analysis

The EEG was recorded at a sample rate of 512 Hz from 64 locations using active Ag-AgCl electrodes (BioSemi, Amsterdam, The Netherlands) mounted in an elastic cap and two mastoid electrodes. Electrodes were placed according the extended International 10–20 system. Two additional electrodes served as reference (Common Mode Sense [CMS] active electrode) and ground (Driven Right Leg [DRL] passive electrode). EEG was referenced offline to an average of left and right mastoids and band-pass filtered (0.5–30 Hz, 24 dB/octave). The 50 Hz interference was removed by a 50 Hz notch filter. The raw data were segmented into epochs of 600 ms, including a 100-ms prestimulus baseline. ERPs were time-locked to the sound onset in the AV and A conditions, and to the corresponding time stamp in the V condition. After EOG correction (Gratton et al. 1983), epochs with an amplitude change exceeding ± 100 μV at any EEG channel were rejected. ERPs of the non-catch trials were averaged per condition (A, V, AVC and AVI), separately for each of the speech and non-speech stimuli. Subsequently, from both congruent and incongruent AV ERPs the corresponding V-only ERPs were subtracted, yielding AVC − V and AVI − V difference waves. The auditory N1 and P2 had a central maximum for A-only and AV difference waves, and analyses were therefore conducted at the central electrode Cz. The N1 was scored in a window of 70–150 ms, the P2 was scored in a window of 120–250 ms. N1 and P2 amplitude and latency scores were separately entered in a repeated measures MANOVA with the between-subject factor Group (schizophrenia vs. healthy controls), the within-subject factors Stimtype (speech vs. non-speech) and Modality (A, AVC − V, AVI − V). Topographic analysis of N1 and P2 − testing differences in scalp distribution between conditions – comprised vector-normalized amplitudes (McCarthy and Wood, 1985) of the 64 electrodes that were subsequently collapsed into 9 electrode clusters: prefrontal–frontal, central and parietal–occipital for left, middle and right side. Topography was tested by a repeated measures MANOVA with Group (schizophrenia vs. healthy controls), Stimtype (speech vs. non-speech), Modality (A, AVC − V, AVI − V) and Electrode (9 electrode clusters) as factors.

An additional analysis involved the spatiotemporal dynamics of the difference of AV interaction between the two groups by conducting point-by-point two-tailed t-tests at each electrode in a 1–300 ms window. Using a procedure to minimize type I errors (Guthrie and Buchwald, 1991), differences in audiovisual interactions were considered significant when at least 12 consecutive points (i.e., 24 ms when the signal was resampled at 500 Hz) were significantly different between groups. This analysis allowed for the exploration of the exact time-course and location on the scalp where differences in AV interactions occurred. We compared two different interactions (collapsed over stimulus category) between groups. In the first analysis the intersensory effect (A = AVC − V) was investigated. We computed the (AVC − [A + V]) difference wave for both groups and performed the running t-tests between both difference waves. In the second analysis we examined the AV congruency effect by comparing the (AVC − AVI) difference wave between groups.

As a final step, the neural sources underlying audiovisual interactions were estimated by using a linear distributed inverse solution based on a Local Auto-Regressive Average (LAURA) model of the unknown current density in the brain (Grave de Peralta Menendez et al., 2001). LAURA estimates three-dimensional current density distributions calculated on a realistic head model with 5017 solution nodes equally distributed in the gray matter of the average MNI (Montreal Neurological Institute) brain. LAURA makes no a priori assumptions regarding the number of sources or their locations and can deal with multiple simultaneously active sources. This analysis was performed using the Cartool software by Denis Brunet (brainmapping.unige.ch/cartool). In the current article the resulting source estimations provide visualization of the likely underlying sources and do not represent a statistical analysis. It also should be noted that the spatial resolution of distributed source estimation is much lower than the spatial resolution of fMRI.

3. Results

Participants detected 98.4% of the catch trials. Detection rates did not significantly differ between the two groups, t(34) = 1.69, p = 0.1. This indicates that both groups did indeed watch the video with attention during stimulus presentation.

3.1. N1

Fig. 1 shows the group-averaged ERPs for each condition (A, V, AVC and AVI) and stimulus type (speech and non-speech). Fig. 2 depicts the A-only ERP and AVC − V and AVI − V difference waves. As is clearly visible in Fig. 2, only in the control group, but not in the patient group, the amplitude of the auditory N1 was reduced in the AVC − V ERP relative to A. Figs. 1 and 2 also indicate that the amplitude of auditory-only N1 for non-speech stimuli and N1 latency for the speech stimuli differed between groups. The MANOVA produced two interactions, Stimtype × Group, F(1,34) = 15.83, p = 0.001, and Modality × Group, F(1,34) = 7.33, p < 0.01, which confirmed the informal observation made from Fig. 2. To further test these interactions, separate MANOVAS for Stimtype and Modality were conducted for each group. For the control group, N1 amplitude across modality to non-speech stimuli was larger than for speech stimuli, F(1,17) = 36.92, p = 0.01, and there was a main effect of Modality F(2,16) = 13.85, p = 0.001. Post hoc pairwise comparisons (Bonferroni adjusted) for Modality showed that for both speech and non-speech stimuli the audiovisual N1 for both congruent and incongruent presentations were reduced in amplitude compared to auditory only N1 (p-values < 0.001). N1 amplitude to congruent and incongruent AV presentations did not differ between each other. For the schizophrenia group there was no effect of Stimtype, F(1,17) = 1.26, p = 0.28, and Modality, F(2,16) = 1.07, p = 0.37, for N1 amplitude.

N1 latency was tested using the same MANOVA as for N1 amplitude. Because the Stimtype × Modality × Group approached the significance level, F(2,33) = 3.23, p = 0.05, separate MANOVAS (Stimtype × Modality) were conducted for both groups. For the control group there was a Stimtype × Modality interaction, F(2,16) = 6.93, p < 0.01. For speech stimuli there was no significant main effect of Modality on N1 latency, F(2,16) = 3.25, p = 0.07. For non-speech stimuli there was a significant main effect of Modality, F(2,16) = 6.60, p = 0.05, because the N1 peaked 10 ms earlier for both congruent and incongruent AV non-speech presentations than for A-only stimuli (p-values < 0.005). No difference in N1 latency was found between congruent and incongruent AV presentations. For the schizophrenia group there was a main effect of Stimtype, F(1,17) = 95.99, p < 0.001 (N1 of speech stimuli was 35 ms later than for non-speech stimuli), but no main effect of Modality, F(2,16) = 3.05, p = 0.08, and no Stimtype × Modality interaction (F < 1).

We also tested whether the scalp distribution of the N1 suppression differed between groups (Fig. 3). Testing N1 topography resulted in a Group × Modality × Electrode interaction, F(8, 27) = 2.62, p = 0.05, indicating that modality affected the N1 topography in both groups differently. Separate MANOVAS for the two groups revealed that for patients with schizophrenia, there was no effect of modality on N1 topography, while for the control group the AV − V and AVI − V had a slightly more anterior distribution than the A-only N1.
3.2. P2

Similar analyses as for the N1 were performed for the P2. Figs. 1 and 2 suggest that in both schizophrenia and control groups auditory P2 amplitude is decreased in the AV conditions. This was confirmed by a repeated measures MANOVA showing a main effect of Modality on P2 amplitude, F(2,33) = 40.76, p < 0.001, that depended on Stimtype, F(2,34) = 20.85, p < 0.001, F(2,34) = 35.16, p < 0.001. Post hoc pairwise comparisons for Modality showed that audiovisual P2 for both congruent and incongruent presentations was reduced in amplitude compared to auditory-only P2 (p-values < 0.001). P2 amplitude to incongruent AV presentations was smaller than for congruent AV presentations (p < 0.01). There was a Stimtype × Group interaction, F(1,34) = 9.48, p < 0.01 indicating that in the control group P2 amplitude to non-speech stimuli was larger (1.4 μV) than to speech stimuli, F(1,17) = 6.07, p < 0.05. In the schizophrenia group Stimtype was marginally significant, F(1,17) = 3.64, p = 0.07, with larger amplitudes (0.6 μV) for speech than for non-speech stimuli.

For P2 latency, no within-subject variable interacted with Group (F-values < 1). There were main effects of Modality, F(2,33) = 8.12, p < 0.01, and Stimtype F(1,34) = 140.18, p < 0.001. P2 latency was shortened for both congruent (11 ms) and incongruent presentations (10 ms) as opposed to A-only (p-values < 0.05), but did not differ between different audiovisual stimuli, F(2,33) = 1.07, p = 0.35. P2 latency was 36 ms shorter for non-speech stimuli compared to speech stimuli.

P2 topography was tested by a Group × Stimtype × Modality × Electrode repeated measures MANOVA. All interactions between Electrode and Group with other variables were non-significant (all p-values at least >0.24). Although the interactions between Stimtype × Electrode, F(1,34) = 3.04 p < 0.05, and Modality × Electrode, F(1,34) = 2.41, p < 0.05 were significant, visual inspection of Fig. 3 points out that P2 topography hardly deviated between these conditions.

3.3. Point-wise t-test

Fig. 4 shows the running t-tests between the groups separately for the AVC − A and AVC − AVI difference waves. Reliable differences in AV interactions started at about 85 ms and lasted approximately 60 ms, which corresponds to the difference in N1 suppression. Although late interactions were highly similar for both groups, a late fronto-central differential effect was observed in a window of 250–285 ms signaling that the difference between A and AV − V was more prolonged for the schizophrenia group. Neural activity associated with AV congruency differed reliably across several electrodes between the groups at the fronto-central sites in a 200–260 ms window. This corresponds to the finding that AV congruency affected (the rising flank of) the P2 in the control group but not in the schizophrenia group.

3.4. Source localization

Source analysis was applied to estimate the neural source of the auditory-only N1, the N1-suppression to bimodal presentations, and the congruency effect at the P2 component. First, the data were re-referenced to the average reference. For every participant the LAURA inverse solutions for the relevant components were computed separately for the A, V, AVC and AVI conditions. For audiovisual interactions at the N1 component we calculated on a node-by-node basis the difference between the AVC and the sum of the unisensory conditions (AVC − [A + V]) over a period of 80–130 ms for non-speech stimuli.
for both groups, 104–128 ms for speech stimuli for the control group and 124–148 ms for speech stimuli for the schizophrenia group, and subsequently averaged across participants. For the control group the main neural generators of the auditory-only N1 were broadly distributed bilaterally in the posterior auditory cortex (BA 22) for both speech and non-speech stimuli. Fig. 5 shows that the strength of these sources was reduced for the schizophrenia group. Source estimations of the audiovisual interaction (AVC − [A + V]) at the N1 revealed interactions at the right posterior superior temporal gyrus (pSTG) for non-speech stimuli and at the right posterior middle temporal gyrus (pMTG) for speech stimuli for the control group. For the schizophrenia patients no neural source was associated with audiovisual integration for non-speech stimuli, whereas for speech stimuli interactions occurred more posteriorly in the middle temporal gyrus than for the control group.

The congruency effect on the P2 component was calculated as the difference in inverse solution between the congruent and incongruent conditions (AVC − AVI) over a period of 170–230 ms for non-speech stimuli and 184–234 ms for speech stimuli. For the control group the congruency effects were predominantly found in the right pSTG for non-speech stimuli and in the left inferior frontal gyrus (BA 9) for speech stimuli. As shown in Fig. 5, none of these activations associated with congruency effects were found for the schizophrenia group.

3.5. The relationship between symptom severity and multisensory integration

Correlations were tested for the schizophrenia group between the BPRS total score, BPRS positive and negative symptoms, with the amplitude and latency of the auditory-only N1 and P2, audiovisual interactions and congruency effects at the N1 and P2 components at electrode Cz, separately for speech and non-speech stimuli. The correlation between the BPRS total score and the auditory-only N1 amplitude for speech stimuli was marginally significant (r = 0.45, p = 0.06). Patients with higher BPRS scores showed smaller auditory-only N1 amplitudes. All other correlations were non-significant.

As audiovisual integration could be influenced by medication, we calculated the correlation between dosage – separately for Clozapine (7 participants) and Olanzapine (6 participants) – and audiovisual integration. All correlations were non-significant (all p-values at least >0.11).

4. Discussion

The most important finding of the current study was that visual predictive information about sound onset did not suppress or speed-up the auditory-evoked N1 in patients with schizophrenia as it did in normal controls. In the patient group there was also no audiovisual congruency effect on the frontal P2 amplitude, thus indicating deficient audiovisual integration at the neural level in these patients. Source estimations of the neural generators further identified abnormalities in the neural networks subserving audiovisual integration. Besides deviations in audiovisual integration, we found differences in auditory-only processing between the groups, as the N1 amplitude for auditory-only stimuli was smaller (Rosburg et al., 2008) and delayed (Boutros et al., 2004) in the patient group. Our results are in accordance with previous reports using behavioral measures where impairments in cross-modal gain of stimulus detection and discrimination were found (Ross et al., 2007; Williams et al., 2010), as well as a reduced McGurk-effect.
in schizophrenia (de Gelder et al., 2003; Pearl et al., 2009), and with electrophysiological studies that used simplified audiovisual stimuli that also reported abnormalities in audiovisual integration in schizophrenia (Stone et al., 2011). As in our study, Stone et al. (2011) also did not find any correlations between audiovisual integration scores and psychopathology.

![Fig. 3. The scalp topographies of peak N1 and P2 for auditory-only (A) and the audiovisual minus visual-only difference wave for congruent (AVC − V) and incongruent (AVI − V) presentations, per stimulus category (speech and non-speech) and per group (healthy controls and schizophrenia). The range of the voltage maps in μV is displayed below each map. The two lower rows display the Student’s t-maps at the same latency as for the voltage maps. The t-maps reflect the distribution of the t-values across the scalp as a result of testing between A-only and the AVC − V difference wave, and between AVC and AVI conditions.](image)

![Fig. 4. Time-course of AV interactions pooled across speech and non-speech stimuli using point-wise t-tests at every electrode in a 1–300 ms post-stimulus window. In the left panel the AVC − (A + V) difference wave were tested between groups. In the right panel the AVC − AVI difference wave was tested between groups.](image)
Fig. 5. LAURA distributed source estimations for the auditory N1, the audiovisual interactions (AVC − [A + V]) at N1 latency and the congruency effect (AVC − AVI) at P2 latency, separately for speech and non-speech stimuli (scale between left and right hemispheres in μA/mm³).
What neural mechanisms might underlie deficient audiovisual integration in schizophrenia? Before we can answer this question we need to consider the hypothesized functional significance of the N1-suppression. In earlier studies we found that the auditory N1-suppression is obtained in audiovisual stimuli when visual information precedes and reliably predicts sound onset (Stekelenburg and Vroomen, 2007; Vroomen and Stekelenburg, 2010). The N1-reduction likely reflects a reduction of temporal uncertainty of the auditory stimulus by the leading visual signal, thus possibly reducing the saliency of the sound. This idea fits nicely with results from the early 1970s where in motor-sensory research it was found that the auditory N1 is dampened by self-generated tones (Schafer and Marcus, 1973; McCarthy and Donchin, 1976; Martikainen et al., 2005) if compared to tones replayed to the participant. This effect has been attributed to reduced temporal uncertainty induced by a forward model that predicts and inhibits the sensory consequences of one’s own actions (Schafer and Marcus, 1973). Interestingly, patients with schizophrenia have deficits in their ability to predict the sensory consequences of their own actions (Ford et al., 2007; Synofzik et al., 2010). In line with this, it has been reported that the dampening of the auditory N1 observed in control subjects to self-initiated sounds was not evident in patients with schizophrenia (Ford et al., 2007). This is taken as evidence for reduced communication between frontal and temporal areas in schizophrenia (Ford et al., 2002, 2007, 2008). It has further been hypothesized that disrupted connectivity between frontal and temporal lobes may contribute to misperception of thoughts as voices (Ford et al., 2002).

As audiovisual integration implies intact connections between visual and auditory cortices, the results of our study showing impaired audiovisual interactions may therefore be compatible with deficits in the neural networks subserving audiovisual integration in schizophrenia. Source localization in the current study has indeed identified abnormalities in the neural routes that are involved in audiovisual integration. First, for the control group, integration of congruent audiovisual stimuli in an 80–130 ms window was associated with activity in pSTG for non-speech stimuli and pMTG for speech stimuli. For the schizophrenia group such activations were either absent (non-speech stimuli) or found at a different location (speech stimuli). Second, we demonstrated in healthy controls that the inferior frontal gyrus is involved in the integration of the phonetic information. This location is in close vicinity with those found in fMRI studies reporting activation in Broca’s area when incongruent speech was contrasted with congruent speech (Ojanen et al., 2005; Szyic et al., 2009). For non-speech stimuli the mismatch contrast was associated with activity in the pSTG. In accordance with other studies (Arnal et al., 2009, 2011; Blank and von Kriegstein, 2013) the error between visual prediction and auditory input is likely represented in pSTG in the healthy controls. In the schizophrenia patients violations of visual prediction were not associated with activity in either frontal or temporal regions. Our data thus show that for healthy controls a widespread network, including frontal and temporal regions, is involved in audiovisual integration, whereas for the schizophrenia patients this integrative network is impaired, which then may point to disrupted functional connectivity between occipital, temporal en frontal cortices.

The current data thus fit with a theory that schizophrenia patients suffer from functional disconnections between distant brain areas and that dysfunctional interactions between specialized brain regions might underlie cognitive impairments in schizophrenia (Friston and Frith, 1995). Available evidence further suggests that both in the sensorimotor and sensory domain, schizophrenia patients show abnormalities in predictive mechanisms and that in both domains this may be due to faulty connections between frontal, occipital and temporal lobes. As sensory prediction may serve an important role in maintaining a ‘sense of coherence’ (Stephan et al., 2009), consequently when sensory input is processed in relative isolation, the world may be experienced as fragmented and incoherent, and anxiety provoked. Degraded multisensory integration might thereby lower the threshold for breakdown of reality testing in the form of delusions and hallucinations.

A limitation of the current study is that the patients were medicated with unknown effects on audiovisual integration. No correlation was found between dosage and neural markers of audiovisual integration. Future studies might examine whether medication affects N1 and N1-reduction by testing unmedicated patient groups.

To conclude, we found clear evidence that patients with schizophrenia show degraded audiovisual integration at the neural level. From a functional perspective, patients with schizophrenia may have difficulties in predicting sound onset and content from vision. As the data did not allow us to directly trace neural pathways, the exact underlying neural substrates remain elusive, although a functional disconnection between visual, auditory and motor cortex can be hypothesized. All patients in the current study were on antipsychotic medication and had a relatively long illness duration. It remains for future study to examine whether these results generalize to unmedicated patients and first episode psychosis patients. Testing unaffected first degree relatives of patients with schizophrenia might provide insight into the genetic underpinnings of the found phenomenon.

Role of funding source
Funding played no role in the realization of this study.

Contributors
Jeroen J. Stekelenburg and Jean Vroomen designed the experiment. Jeroen J. Stekelenburg recruited the healthy controls, collected the data, performed the data analysis and prepared the manuscript. Jean Vroomen assisted in data analysis and manuscript preparation, Jan Pieter Maes and Arthur R. Van Goor were responsible for recruitment of the patients and clinical assessments and reviewed and edited the manuscript. Margriet Sitskoorn reviewed and edited the manuscript. All authors have approved the final manuscript.

Conflict of interest
There are no conflicts of interest.

Acknowledgments
We would like to express our gratitude to all participants involved in this study. The Cartool software (brainmapping.unige.ch/cartool) has been programmed by Denis Brunet, from the Functional Brain Mapping Laboratory, Geneva, Switzerland, and is supported by the Center for Biomedical Imaging (CIBM) of Geneva and Lausanne.

References