Schizophrenia Research 147 (2013) 253–261

Contents lists available at SciVerse ScienceDirect

Schizophrenia Research

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

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

ARTICLE INFO

Article history:
Received 6 December 2012
Received in revised form 10 April 2013
Accepted 27 April 2013
Available online 23 May 2013

Keywords:
Schizophrenia
Event-related brain potentials
Auditory N1
P2
Multisensory perception
Audiovisual integration
Temporal prediction

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.

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., 2005; de Jong et al., 2009), sensitivity for detection of audiovisual temporal order (Foucher et al., 2007), and intersensory facilitation of reaction times in which bimodal targets, with cues from two sensory modalities, are detected faster than unimodal targets (Williams et al., 2010).

Here we examined the neural correlates that may underlie these deficits of multisensory integration, using stimuli that have produced robust and consistent AV interaction effects in healthy participants across a number of studies (Klucharev et al., 2003; Besle et al., 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

* Corresponding author. Tel.: +31 13 466 8171; fax: +31 13 466 2067.
E-mail address: J.J.Stekelenburg@uvt.nl (J.J. Stekelenburg).

0920-9964/$ – see front matter © 2013 Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.schres.2013.04.038

associated audiovisual interactions (AV = A + V + [A × V interactions]) (Giard and Peronnet, 1999). If the unimodal signals are processed independently, then the bimodal response equals the sum of unisensory 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 (Giard 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 unisensory 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) or 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) stimulus 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 the

Table 1
Demographic and clinical characteristics of schizophrenia patients and healthy controls.

	Schizophrenia patients	Healthy controls
N	18	18
Gender	17 male	17 male
Age (years)	39.1 (8.2)	38.0 (9.0)
Handedness	15 R, 3 L	12 R, 6 L
Education level		
Elementary	1	0
Middle	13	13
Higher	4	5
Illness duration (years)	16.2 (5.6)	–
BPRS total score	41.4 (11.5)	–
Antipsychotic medication	N	Mean daily dosage in mg (range)
Clozapine	7	546.4 (125–1000)
Olanzapine	6	20.4 (10–40)
Risperidone	1	4
Quetiapine	1	600
Zuclopenthixol	1	20
Bromperidol	1	2
Haloperidol	1	10