Schizophrenia Research 110 (2009) 111–118

Contents lists available at ScienceDirect

Schizophrenia Research

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

Audiovisual integration of speech is disturbed in schizophrenia: An fMRI study

G.R. Szycik [a,b,*], T.F. Münte [b,c,d], W. Dillo [a], B. Mohammadi [b,c], A. Samii [c], H.M. Emrich [a], D.E. Dietrich [a]

[a] Clinic for Psychiatry, Social Psychiatry and Psychotherapy, Medical School Hannover, Germany
[b] Department of Neuropsychology, University of Magdeburg, Germany
[c] International Neuroscience Institute, Hannover, Germany
[d] Center for Behavioral Brain Science, Magdeburg, Germany

ARTICLE INFO

Article history:
Received 22 June 2008
Received in revised form 22 February 2009
Accepted 3 March 2009
Available online 19 March 2009

Keywords:
Speech
Audiovisual integration
fMRI
Multimodal processing
Schizophrenia
Right hemisphere

ABSTRACT

Speech perception is an essential part of social interaction. Visual information (lip movements, facial expression) may supplement auditory information in particular under inadvertent listening situations. Schizophrenia patients have been shown to have a deficit in integrating articulatory motions with the auditory speech input. The goal of this study was to investigate the neural basis of this deficit in audiovisual speech processing in schizophrenia patients by using fMRI. Disyllabic nouns were presented in congruent (audio matches visual information) and incongruent conditions in a slow event related fMRI design. Schizophrenia patients ($n = 15$) were compared to age and gender matched control participants. The statistical examination was conducted by analysis of variance with main factors: audiovisual congruency and group membership. The patients' brain activity differed from the control group as evidenced by congruency by group interaction effects. The pertinent brain sites were located predominantly in the right hemisphere and comprised the pars opercularis, middle frontal sulcus, and superior temporal gyrus. In addition, we observed interactions bilaterally in the fusiform gyrus and the nucleus accumbens. We suggest that schizophrenia patients' deficits in audiovisual integration during speech perception are due to a dysfunction of the speech motor system in the right hemisphere. Furthermore the results can be also seen as a reflection of reduced lateralization of language functions to the left hemisphere in schizophrenia.

1. Introduction

To integrate information from different modalities into a coherent representation of the environment is an essential task which is not easy given the different processing times in the different modalities and the widely differing anatomical pathways and networks subserving the processing of unimodal information. A special case of a multisensory integration process is the perception of speech. Unlike our introspective intuitions as normal hearing listeners we are constantly evaluating the characteristic lip movements associated with each phoneme. Indeed, seeing a speaker's articulatory movements facilitates the recognition of spoken words in noisy environments substantially (Erber, 1969; Grant and Seitz, 2000; Munhall et al., 2004; O'Neill, 1954; Sumby and Pollack, 1954). This gain in intelligibility from viewing visual articulations is maximal at intermediate signal-to-noise ratios (Ross et al., 2007). Furthermore, the presentation of audiovisual (AV) incongruent (auditory stream does not match the articulatory movements) speech may lead to new hybrid percepts occurring as a result of the fused information from both channels. The most prominent example is the McGurk effect (Dekle et al., 1992; McGurk and MacDonald, 1976). Thus, the comprehension of language results from the integration of auditory and visual modality information that is used and combined in a flexible, context dependent manner. Recently, a number of brain regions including the superior temporal sulcus (Calvert et al., 2000; Sekiyama et al., 2003; Szycik et al., 2008a; Wright et al., 2003), Broca's area

* Corresponding author. Clinic for Psychiatry, Social Psychiatry and Psychotherapy Medical School Hannover, Carl-Neuberg-Straße 1, 30625 Hannover, Germany. Tel.: +49 511 5325246; fax: +49 511 5323187.
E-mail address: szycik.gregor@mh-hannover.de (G.R. Szycik).

0920-9964/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.schres.2009.03.003

(Ojanen et al., 2005; Pekkola et al., 2006) as well areas on the dorsal surface of the temporal lobe (Callan et al., 2001; Calvert et al., 1999; Möttönen et al., 2004) have been suggested to be involved in audiovisual integration of speech stimuli.

Schizophrenia patients show many problems of interaction and communication in social situations. There is a good evidence for speech processing deficits in schizophrenia (Covington et al., 2005; DeLisi, 2001), but little is known on possible impairments of the integration of audiovisual information during speech perception. Behavioural studies comparing schizophrenia patients with healthy controls examined different aspects of AV speech integration. In the ventriloquism effect auditory speech information is perceived as if coming from the puppet's mouth (Slutsky and Recanzone, 2001). Schizophrenia patients have been shown to have a similar bias as controls towards locating the auditory information nearer to the visual stimulus (De Gelder et al., 2003). However, schizophrenia patients are much less susceptible to illusory phoneme fusions (i.e. the McGurk illusion) than healthy participants showing deficits in their ability to integrate articulatory motion with the auditory speech input (De Gelder et al., 2003; Ross et al., 2007). At the same time, no deficits in unisensory speech processing were observed, be it lip reading in the visual modality (Myslobodsky et al., 1992; Ross et al., 2007; Schonauer et al., 1998) or purely auditory speech perception (De Gelder et al., 2003; Ross et al., 2007). Thus, the current behavioural data hints at a specific deficit in multisensory speech processing in schizophrenia with intact unimodal speech processing. To our knowledge only one prior study has addressed the neural correlates of these deficits in schizophrenia patients (Surguladze et al., 2001). In this earlier study, schizophrenia patients and control participants either listened to auditory speech, performed silent lip reading, or were exposed to meaningless lip movements (visual non-speech condition). The main result of the study was that schizophrenia patients with psychotic symptoms showed activation of polysensory structures in the visual non-speech condition which was interpreted as a possible risk factor with regard to paranoid and hallucinatory symptoms.

Our goal in this study was to investigate the neuronal correlates underlying the deficit of AV integration during speech perception in schizophrenia patients. In contrast to the fMRI study of Surguladze et al. (2001) which presented unimodal speech stimuli, we therefore focused our experiment on the bimodal processing of speech.

2. Methods

2.1. Participants

All procedures have been approved by the local ethics committee. 15 schizophrenia patients (DSM-IV criteria; APA, 1994; 8 women, mean age 38.2 ± 10.6 years) and 15 normal controls (8 women, mean age was 36.5 ± 9.4) gave informed consent and participated for a small monetary compensation. Two additional patients were lost, because of claustrophobia in the scanner. All subjects were (self-reported) right-handed and native speakers of German.

All patients were under antipsychotic treatment (chlorpromazine equivalent in mg/day was 380 ± 222) at the time of the experiment. Three patients got typical and the remaining twelve atypical antipsychotic medications. Twelve patients were in remission after an acute episode and three showed residual state. The mean illness duration was 9.9 ± 9.4 years. The mean Brief Psychiatric Rating Scale Score was 24.1 ± 8.2. Control participants were free of previous or current psychiatric or neurological disorders.

2.2. Stimuli and design

The paradigm is illustrated in Fig. 1 and corresponds to a previous study by our group in healthy individuals (Szycik et al., 2008a). Stimuli were derived from the German part of

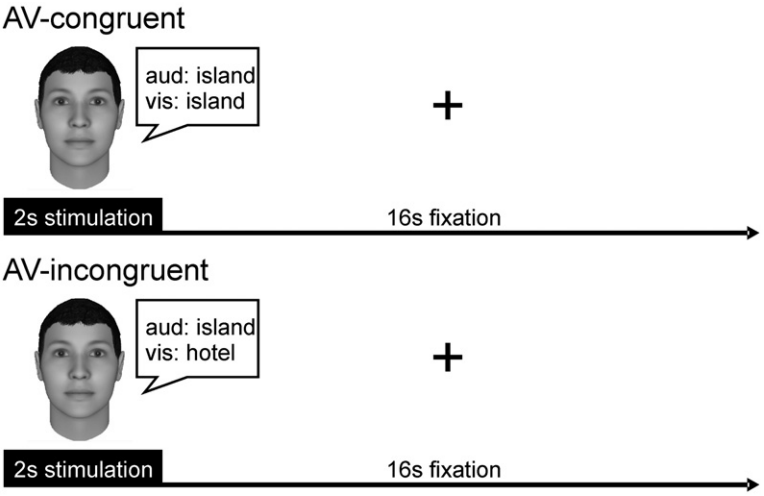

Fig. 1. Illustration of the paradigm: each trial comprised a video of a face pronouncing a 2 syllable German noun. In the AV-congruent condition (upper row) the auditory and the visual information matched, in the AV-incongruent condition (lower row) the visual information corresponded to the pronunciation of a different word. Each video segment was 2 s in duration and was followed by a period of 16 s during which only a fixation cross was present. The participants had to react to rare occurrences of a certain semantic category (animals) by a button press.