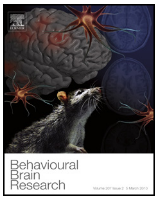
Behavioural Brain Research 305 (2016) 76–86

Contents lists available at ScienceDirect

Behavioural Brain Research

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



Research report

Simultaneous face and voice processing in schizophrenia

Taosheng Liu [a,b], Ana P. Pinheiro [c,d], Zhongxin Zhao [b], Paul G. Nestor [c,e], Robert W. McCarley [c], Margaret Niznikiewicz [c,*]

[a] Department of Psychology, Second Military Medical University (SMMU), Shanghai, China
[b] Department of Neurology, Changzheng Hospital, SMMU, Shanghai, China
[c] Clinical Neuroscience Division, Laboratory of Neuroscience, Department of Psychiatry, Boston VA Healthcare System, Brockton Division and Harvard Medical School Boston, MA, United States
[d] Neuropsychophysiology Laboratory, CIPsi, School of Psychology, University of Minho, Braga, Portugal
[e] University of Massachusetts, Boston, MA, United States

HIGHLIGHTS

- Electrophysiological indices of face–voice processing in schizophrenia are studied.
- ERP indices of face alone, voice alone and face–voice processing are examined.
- Abnormal processing of face alone and face–voice in schizophrenia.
- Evidence of multimodal integration in early ERPs found in both HC and schizophrenia.
- Late ERP potentials show group differences in unimodal vs multimodal processes.

ARTICLE INFO

Article history:
Received 6 January 2015
Received in revised form 6 January 2016
Accepted 17 January 2016
Available online 22 January 2016

Keywords:
Multimodal processing
Face
Voice
Schizophrenia
N170
P270
P400

ABSTRACT

While several studies have consistently demonstrated abnormalities in the unisensory processing of face and voice in schizophrenia (SZ), the extent of abnormalities in the simultaneous processing of both types of information remains unclear. To address this issue, we used event-related potentials (ERP) methodology to probe the multisensory integration of face and non-semantic sounds in schizophrenia.

EEG was recorded from 18 schizophrenia patients and 19 healthy control (HC) subjects in three conditions: neutral faces (visual condition-VIS); neutral non-semantic sounds (auditory condition-AUD); neutral faces presented simultaneously with neutral non-semantic sounds (audiovisual condition-AUDVIS).

When compared with HC, the schizophrenia group showed less negative N170 to both face and face–voice stimuli; later P270 peak latency in the multimodal condition of face–voice relative to unimodal condition of face (the reverse was true in HC); reduced P400 amplitude and earlier P400 peak latency in the face but not in the voice–face condition.

Thus, the analysis of ERP components suggests that deficits in the encoding of facial information extend to multimodal face–voice stimuli and that delays exist in feature extraction from multimodal face–voice stimuli in schizophrenia. In contrast, categorization processes seem to benefit from the presentation of simultaneous face–voice information. Timepoint by timepoint tests of multimodal integration did not suggest impairment in the initial stages of processing in schizophrenia.

1. Introduction

Abnormal social cognition is increasingly recognized as an important component of schizophrenia (SZ) pathology [1]. The two richest sources of social information are facial expressions and tone of voice, or prosody. Abnormalities in the processing of both facial expressions [2,3] and prosody [4–7] have been reported

---

* Corresponding author at: Boston VA Medical Center and Harvard Medical School Psychiatry 116A, 940 Belmont St, Brockton, MA 02301, United States. Fax: +1 508 583 5900.
E-mail addresses: margaret_niznikiewicz@hms.harvard.edu, margaretniznikiewicz@gmail.com (M. Niznikiewicz).

http://dx.doi.org/10.1016/j.bbr.2016.01.039
0166-4328/Published by Elsevier B.V.

in schizophrenia. However, socially relevant information rarely comes from one channel: in a typical social situation, both facial expressions and a tone of voice convey crucial social information. Thus, successful navigation in social environments is predicated on the ability to integrate multisensory information [8,9].

Therefore, this study focuses on face and voice processing as sources of *simultaneous,* socially relevant information, in contrast to the studies that focus on fusion effects related to speech perception, such as those observed in the McGurk or McDonald effects [34]. There are several reasons why integration of face–voice should be of interest to the study of schizophrenia. For example, deficits in the processing of social cognition have been associated with poor functional outcomes in schizophrenia. Schizophrenia has been consistently associated with impairment in face processing in a range of designs [35]. Prosody processing abnormalities have been also recently reported in schizophrenia; of note, our recently published studies on prosody processing in schizophrenia suggest that a level of impairment may depend on emotional content of prosodic stimuli [6,7]. Specifically, we found more impairment in the emotional prosody processing relative to neutral prosody processing in the patient group. At the same time, face processing alone was found impaired in schizophrenia for both neutral and emotional faces [50]. Thus, starting an investigation from examining neuro-cognitive indices of neutral emotion processing seemed a reasonable first step in the investigation of the poorly understood phenomenon of simultaneous face–voice processing.

Importantly, studies on simultaneous processing of face and voice information in schizophrenia are currently non-existent. Therefore, it is not clear at all if the processing of such stimuli further impairs the processing of emotional information or, conversely, whether the simultaneous presentation of congruent face–voice expressions aids the processing of emotional content over and above how such co-occurring presentations can aid processing in healthy control subjects. In fact, most conclusions regarding the processing of socially relevant information in schizophrenia have been reached based on studies using unimodal (i.e., either face or voice) stimuli. And yet, as mentioned already above, most social interactions involve the use of both face and voice information.

There is a growing body of evidence on brain regions involved in multisensory integration of face and voice information: they are localized to heteromodal, temporo-parietal regions and include posterior superior temporal gyrus (STG) [10,11], posterior superior temporal sulcus [12,13], and superior parietal lobule [14,15]. However, fMRI studies cannot examine the temporal dynamics of integrating face and voice information; event related potential (ERP) methodology is an excellent tool to examine temporal dynamics and yet very few ERP studies have been devoted to the issue of integrating information from face and voice [16,17]. These studies suggested early sensitivity to multimodal stimuli but also indicated that the specificity of neural responses might depend on stimuli characteristics and cognitive demands of an experimental task. For example, the Jessen and Kotz study [16] using both neutral and emotional face–voice pairings found that both N100 and P200 amplitudes differed in the audiovisual relative to unimodal conditions. The Latinus et al., [17] study used face–voice pairings to examine the impact of unattended stimulus (either voice or face) on the processing of a target stimulus (either face or voice), or in a gender discrimination task. The processing of an unattended social cue impacted the processing of an attended social cue at later (180–230 msec, post-stimulus onset) latencies, while the processing of gender face–voice compatibility impacted the processing of early potentials (30–100 msec, post-stimulus).

Across different brain imaging designs, the evidence for multisensory integration is considered to be either a differential brain activity to multimodal relative to unimodal stimuli [18,19] or additivity where the response to audiovisual stimulation is greater than the sum of the unisensory responses (superadditivity) or, in some designs, less than the sum of the unisensory responses (subadditivity) [19,20]. In behavioral studies, greater accuracy in target (e.g., emotion) recognition for congruent audiovisual stimuli (i.e., where both voice and face express the same emotion) relative to unimodal stimuli is also taken as evidence for multisensory integration [21].

In schizophrenia, most studies that examined multisensory integration have used behavioral measures [22–29]. Abnormalities in sensory integration were found in most of these studies. However, the type of abnormalities differed across studies: most found reduced benefits from processing multisensory stimuli in schizophrenia patients [22,24,26], but some studies found stronger effects of integration in schizophrenia relative to healthy control (HC) subjects [23,25].

Two fMRI studies examined different aspects of audiovisual speech perception in schizophrenia and found abnormalities in a network of regions important for multimodal integration. In the speech lip-reading task, patients had reduced activation in the same regions as HC (posterior inferior temporal cortex, occipital cortical sites (BA19), temporal regions (BA 21,22,42) and inferior frontal gyrus), while in the non-speech lip reading task, actively psychotic patients showed stronger activations in the insula and striatal regions relative to HC [30]. Similarly, Szycik et al. [31] reported lower activation in the patient group relative to HC in hetero-modal brain regions including the inferior frontal gyrus and STG in response to incongruent visual-auditory speech stimuli.

The two published ERP studies on multimodal integration in schizophrenia have produced somewhat contradictory results, which perhaps highlights the dependence of the results on the task and stimuli used. In the ERP study of a soccer ball and a sound simulating its movement, Stone et al. [32] found that in spite of impaired ERP responses to unimodal stimuli (a ball or a bouncing ball sound), there was an increased facilitation in the patient group to multimodal stimuli. In the time-point by time-point analysis of the sum of the auditory and visual ERPs contrasted with the audiovisual ERP, patients with schizophrenia had higher amplitudes within the interval of 70–94 msec, post-stimulus over the right occipital locations. In contrast, abnormal N100 and P200 (reduced amplitude and a lack of modulation due to multi-modal stimulus presentation) responses in a schizophrenia group were found to non-speech stimuli [33].

The current study is the first to investigate whether abnormalities exist in the processes of integrating information from human faces and voices in schizophrenia using ERP measures. This study will fill in important gaps in our understanding of the processing of socially relevant, not related to dynamic speech perception, multimodal information in this disorder. It will help identify the nature of multimodal integration of voice–face information and whether the processes involved are characterized by super-additivity or subadditvity. It will help identify stages of information processing that are abnormal in schizophrenia in the course of voice-face processing as characterized by specific ERP components. Thus, we believe, it will significantly contribute to constructing ecologically valid models of social information processing in schizophrenia.

We hypothesized that for both groups, integration effects would be observed both in the early (parieto-occipital: P100, N170; fronto–central: N100, P200) and late ERP (parieto-occipital: P270; fronto–central: N250, P400) components [9,32,36] listed above, given that face/voice stimuli used in this study are by design more complex and therefore presumably more ecologically valid than those tested in the previous studies. Our experimental design allows for the direct comparison of neurophysiological response of unimodal and multimodal processing of socially relevant visual and auditory stimuli in the form of faces and voices, respectively, both within each group and between the two groups (healthy con-