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.

Published by Elsevier B.V.

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-

trols and patients with schizophrenia). The evidence for integration will be pursued with two approaches: by direct comparing the relevant ERPs (parieto-occipital components (P100, N170, P270) and fronto–central components (N100, P200, N250 and P400) to unimodal and to multimodal stimuli, and by following accepted methods of examining integration effects consisting of comparing multimodal ERPs to the sum of ERPs recorded to unimodal stimuli.

In addition, we predicted reduced N170 to face stimuli given previous reports of abnormal face processing in schizophrenia [35], and normal N100/P200 ERP response to neutral, non-semantic sounds, given reports from our laboratory suggesting more abnormality to emotional than to neutral prosody in schizophrenia [6,7].

2. Methods

2.1. Subjects

Twenty subjects with chronic schizophrenia (1 female) and 22HC (2 female) matched for age, handedness and parental socioeconomic status, with normal hearing as assessed by audiometry and normal/corrected vision as assessed by Snellen criteria (25/20 as normal vision), participated in the experiment (see Table 1 for socio-demographic information). Comparison subjects were recruited via advertisements in local newspapers and websites. The inclusion criteria for all subjects were: age 18–55, English as first language; right handedness [37]; no history of neurological illness; no history of DSM-IV diagnosis of drug or alcohol abuse [38]; verbal intelligence quotient (IQ) above 75 [39]; no hearing (as tested with an audiometer), vision or upper body impairment. For HC, an additional exclusion criterion was a history of psychiatric disorder in oneself or in first-degree relatives. Patients were diagnosed (screened for HC) using the Structured Clinical Interview for DSM-IV for Axis I [40] and Axis II [41] disorders.

Before participation in the study, all participants signed an informed consent form to confirm their willingness to participate in the study (following Harvard Medical School and Veterans Affairs Boston Healthcare System guidelines).

Five of the participants (2 patients, 3HC) were excluded due to excessive artifacts in EEG (more than 30% of trials rejected), leaving 18 patients and 19 control subjects for the final data analyses (Table 1).

2.2. Stimuli

Stimuli were presented either unimodally or bimodally. Non-semantic sounds were chosen to avoid the confounding effects of linguistic (e.g., semantic) information on auditory and audiovisual processing.

Visual stimuli (VIS) were 36 human and 3 primate neutral colored static faces (see Fig. 1A and Ref. [9]). The human faces included 18 male and 18 female adult models displaying neutral emotion and were selected from the Radbound Faces Database (www.rafd.nl—see Ref. [9]). Auditory stimuli (AUD) were 36 neutral human sounds (neutral 'mmm' sound) and 3 primate neutral digitized voices (see Fig. 1A and Ref. [9]). In the AUDVIS condition, the image files were presented simultaneously with the sound files, with the onset of sound files locked to the onset of the image files, resulting in 36 human unique face–voice stimuli. The sounds were adjusted to 72 dB intensity and stayed on for the duration of face stimulus presentation.

The full description of both face and voice stimuli is included in [9]. Briefly, the 'mmm' sound was produced by 9 male and 9 female individuals. Each individual was asked to produce 4 samples of the sound which were then assessed by three judges. The two sounds that were consensually rated by the three judges as representing neutral emotion best were used in the study. All human voices were adjusted to 72 dB in sound intensity. Mean fundamental frequency (F0) for the 'mmm' sound was 168.35 Hz (SD = 49.3). In addition, the same sounds were presented to 10 (3 females; mean age = 31.9 ± 9.5) subjects who did not participate in the ERP experiment. These subjects were asked to assess the valence and arousal of each sound, by using a 9-point scale as in the Self-Assessment Manikin [see Ref. [9]]. Mean valence ratings for neutral sounds were 4.96 (SD = 0.48) and mean arousal ratings for neutral sounds were 2.91 (SD = 1.51).

Auditory stimuli were presented binaurally through Sennheiser HD 380 PRO Headphones, at a comfortable sound level for each participant (about 72 dB intensity). In the VIS block, each monkey face was repeated three times, as was each monkey voice in the AUD block. Thus, there were 36 unique human and 9 monkey stimuli in each AUDVIS, VIS and AUD block, which resulted in 45 stimuli per block/condition. The trials' structure was the same across the three experimental sessions (Fig. 1A).

Audiovisual stimuli (AUDVIS) consisted of the simultaneously presented face and voice and included 36 neutral human and 9 primate audiovisual species-specific and gender appropriate combinations: a male face was paired with a male voice, a female face was paired with a female voice, and a primate face was paired with a species specific primate voice, respectively.

We would like to note that in spite of the fact that these were non-dynamic faces paired with sounds, their pairing with the 'mmm' sounds, which are produced with lips closed and facial features not changing, gave them a high ecological validity: humans do produce such sounds, for example to 'buy' time when thinking of a suitable response. At the same time, the use of non-dynamic face allowed for very precise time-locking between face and voice stimuli.

2.3. Subject' Task

Each session was preceded by a practice session involving presenting 8 stimuli compatible with a session type that was to follow: auditory, visual, or audiovisual to ensure that subjects understood the instructions and were comfortable with a required response.

In order to avoid contaminating ERPs by preparatory neural responses and to maintain attention to the task, participants responded to stimuli which were not the focus of subsequent analyses. In the auditory condition, they were instructed to press a response button to a monkey voice. In the visual condition, they were instructed to press a response button to a monkey face, and in the audiovisual condition, they were instructed to press a response button to monkey face–voice (as explained above, monkey vocalizations were species specific). The subjects used their index finger to make a response with the use of the left and right index finger counterbalanced across subjects.

3. Procedure

All participants were tested on three occasions within a week to 10 days (minimum: 3 days, mean: 5.5 days, maximum: 10 days). The visual, auditory and audiovisual stimuli were presented in separate blocks, with the blocks presented in random order. In addition, stimuli within blocks were presented randomly using SuperLab 4.2 (Cedrus Corporation, San Pedro, California, USA). During the experimental session, subjects were seated at a 100 cm distance from a CRT computer screen and instructed to respond to target stimuli as described above.

Table 1
Socio-demographic data in healthy controls and schizophrenia patients, and clinical characterization of patients.

Variable	Healthy Controls	Patients	Statistical Test[a]	*P* value
Age (years)	44.6 ± 7.94	46.9 ± 9.66	−.762	.452
Gender (M;F)	17;1	16;1	.000	1.000
Education (years)	14.9 ± 2.09	12.9 ± 2.25	2.653	.012*
Verbal IQ	102.7 ± 12.69	97.5 ± 13.37	1.190	.242
Full Scale IQ	102.0 ± 14.40	92.1 ± 11.15	2.261	.030*
Subject's SES	2.1 ± 0.68	3.5 ± 1.38	−3.168	.002*
Parental SES	2.4 ± 0.85	2.9 ± 1.25	−1.523	.143
Onset age (years)	NA	26.2 ± 9.25	NA	NA
Duration (years)	NA	16.7 ± 9.19	NA	NA
Chlorpromazine EQ (mg)[b]	NA	614.1 ± 517.73	NA	NA
PANSS Total Score	NA	80.5 ± 30.79	NA	NA

NA—non-applicable; IQ—intelligence quotient; Chlorpromazine EQ—chlorpromazine equivalent (calculated based on Woods [58] and Stoll [59]); PANSS—Positive and Negative Syndrome Scale (Kay et al. [60]).

* $p < 0.05$.

[a] Independent-samples T test was used to test group differences in Age, Education and Verbal IQ; chi-square test was used to test for Gender differences and Mann–Whitney *U* test was used to test for group differences in parental and subject's SES.

[b] All patients were taking neuroleptic medication at the time of study: with 14 patients taking atypical antipsychotic medications, and 4 a combination of typical and atypical antipsychotic medications.

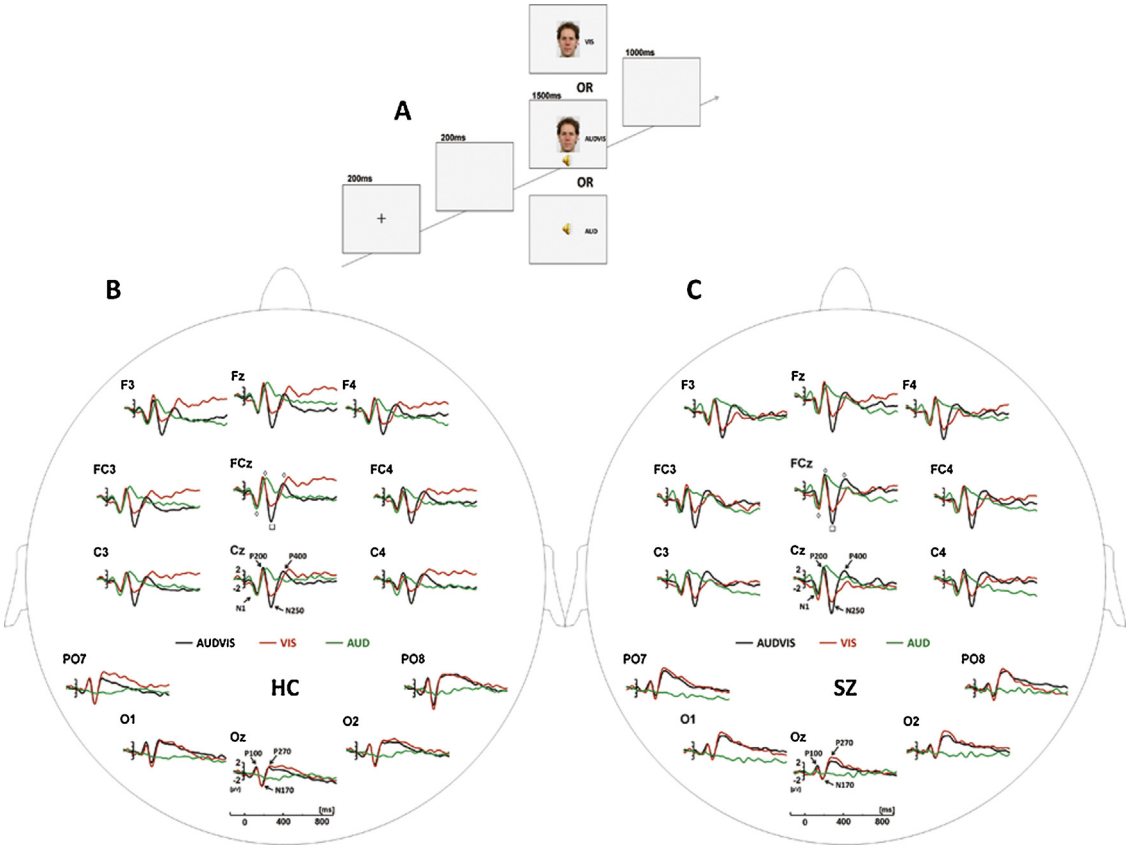

Fig 1. A: Schematic illustration of an experimental trial; B: Grand average waveforms to each condition in NC; C: Grand average waveforms to each condition in SZ. In the fronto–central region, N100 and P200 were observed across all conditions: AUD, AUDVIS and VIS, while N2Z50 and P400 were observed in AUDVIS and VIS but not in the AUD conditions. In the parieto-occipital region, P100, N170 and P270 were observed to AUDVIS and VIS but not to AUD conditions.
Symbol—significant condition latency differences.
Symbol—significant condition amplitude differences.

3.1. EEG recording and analysis

While participants performed the task, the EEG was recorded with 64 electrodes mounted on a custom-made cap (Electro-cap International), according to the expanded 10–20 system [42] using Biosemi Active 2 system. EEG was acquired in a continuous mode at a digitization rate of 512 Hz, with a bandpass of 0.1 to 100 Hz. Data were re-referenced offline to the mathematical average of the left and right mastoids. Horizontal and vertical ocular movements were recorded via electrodes placed on the left and right temples and one below the left eye.

EEG data were processed offline using BrainVision Analyzer package software (Brain Products GmbH, Munich, Germany). They were time-locked to the stimulus onset and segmented into 1000 ms epochs with a 100 ms baseline. Eye blink and movement artifacts were corrected using Gratton et al. [43] algorithm (BrainVision Analyzer package). Artifact-free trials (not exceeding ±100 microvolts at any electrode) were averaged separately for each con-

dition. Individual averages were digitally smoothed with a zero phase shift low pass filter at 12 Hz (24 dB/octave). For the subjects included in the analyses, there were no group differences in the number of trials per condition, $F(1,35) = 1.470$, $p > 0.05$ (29.4 ± 4.63 for HC vs. 29.1 ± 4.52 for SZ in the AUDVIS condition; 31.1 ± 3.71 for HC vs. 29.6 ± 4.70 for SZ in the VIS condition; 30.1 ± 4.94 for HC vs. 28.2 ± 5.34 for SZ in the AUD condition). ERP components' amplitude and latency were measured as the most positive (for positive components) or the most negative (for negative components) data point in a specified latency see below).

Based on visual inspection of ERP waveforms, two regions were selected for analysis: 1. the parieto-occipital region, in which P100 (85–180 ms latency window), N170 (130–230 ms latency window) and P270 (210–350 ms latency window) were observed in response to AUDVIS and VIS conditions, but not to the AUD condition and 2. the fronto–central region, in which N100 (75–170 ms latency window), P200 (150–240 ms latency window) were identified across all three conditions, while N250 (210–330 ms latency window) and P400 (310–480 ms latency window) were identified in both AUDVIS and VIS condition, but not in the AUD condition (see Fig. 1B and C).

3.2. Statistical analyses

3.2.1. Behavioral data analyses

Accuracy rates to target monkey faces/voices/face–voice/stimuli were tested with independent sample *t*-tests for AUD, VIS, and AUDVIS conditions.

3.2.2. ERP components analyses and multimodal integration

Several recent studies used traditional ERP analyses in which amplitude and/or latency differences between multimodal and unimodal stimuli were treated as evidence for multimodal integration [e.g.,16,17]. However, it has been also argued that a more stringent test of differences between the processing of multimodal and unimodal stimuli, in case of the ERP signal, involves the use of running time-point by time-point *t*-tests in which the ERP waveforms to audiovisual stimuli are compared to the sum of the auditory and visual conditions [e.g.,44] (see below for further explanation of this approach). We here present both types of analyses and discuss the limits of the use of the time-point by time-point analyses and the benefits of deriving complementary information from both approaches.

The amplitude and latency of parieto-occipital P100, N170 and P270 were separately submitted to multivariate analysis of variance (MANOVA) with group (HC, SZ) as between-subjects factor, and condition (AUDVIS, VIS) and electrode (PO7/PO8, O1/Oz/O2, PO9/PO10) as within-subjects factors. The amplitude and latency of fronto–central N100 and P200 were separately submitted to MANOVA with group (HC, SZ) as between-subjects factor, and condition (AUDVIS, VIS, AUD), region (Frontal-Fz, F1/2/3/4/5/6; Fronto–central-FCz, FC1/2/3/4/5/6; Central-Cz, C1/2/3/4/5/6) as within-subjects factors. For N250 and P400, the same statistical model was used but the conditions included were (AUDVIS and VIS) given the fact that these two components were absent in the AUD condition. Main effects and interactions were followed with planned comparisons with Bonferroni correction. The amplitude and latency values for all components are presented in Table 2. Only significant main effects and interactions are reported for ERP data and all effects are presented in Table 3. All significance levels in text are two-tailed with the preset alpha level for significance of $p < 0.05$.

3.2.3. The influence of antipsychotic medication

We have tested the relationships between all ERP components' amplitude and latency and chlorpromazine levels to examine whether antipsychotic medication impacted ERP responses.

3.2.4. Multisensory integration using timepoint by timepoint analysis

In imaging studies, it has been proposed that multisensory integration will manifest as differential activation recorded from multi-sensory brain regions relative to uni-sensory brain regions or, in case of EEG/MEG, differential activation observed to multisensory relative to the sum of activation observed to uni-sensory stimuli [19,45,46]. As mentioned in the Introduction, if the activation to the audiovisual stimuli is greater than the sum of activation to auditory and visual stimuli it is an evidence for superadditivity, and if it is less, it is treated as evidence for subadditivity. In the ERP research, the test of this assumption is accomplished by comparing a waveform from audiovisual condition to the waveform constructed by adding waveforms from the auditory and visual conditions. As noted in [47], the working assumption is that neural activation induced by audiovisual stimulus will equal to the sum of neural activities induced separately by auditory and visual stimulus plus the neural activation induced uniquely by multimodal interactions: ERP(AUDVIS) = ERP(AUD) + ERP(VIS) + ERP(AUD × VIS integration). This assumption is valid only if the EEG activity is not contaminated by response selection or motor activities. Under these conditions, it is valid for all configurations of generators based on the law of superimposition of electrical fields [47].

To test the effect of multisensory integration, the AUDVIS ('multimodal') ERPs were compared with the sum of AUD and VIS ('sum') ERPs in SZ and NC at Fz, Cz, Pz, POz and Oz, and the difference ERPs waveforms (AUDVIS-'sum') were calculated and compared for SZ and NC, using the formula [32,47]: AUDVIS—[AUD + VIS].

The amplitudes of the 'multimodal' and 'sum' ERPs at each electrode from 0 to 250 ms post-stimuli were compared using running timepoint-by-timepoint paired-samples *T* test (two-tailed) for each group. An AUDVIS integration was defined as at least 10 consecutive data points meeting a 0.05 alpha criterion [44]. This criterion has been established to be a stringent test of reliable effects when a large number of *t*-tests is conducted and is an accepted alternative to a Bonferroni correction for multiple comparisons [44,48].

To assess the differences in multisensory integration between groups, first, the difference ERP waveforms were obtained where the 'sum' ERPs were subtracted from the 'mutlimodal' ERPs for each group. Then, between-group comparisons of the difference ERP waves were conducted with timepoint-by-timepoint independent samples *t*-tests, and the same significance criteria as those listed above were applied.

This approach allows making inferences about the strength of the electro-physiological signal but not directly about ERP components involved. The identification of ERP components associated with differences between uni-modal vs. multimodal conditions is only possible based on the inspection of latency windows within which significant differences were found between conditions or groups.

Furthermore, we believe that this approach can be used only for these parts of the waveforms that are characterized by clear ERP components. Accordingly, our choice of 0–250 interval was dictated by the structure of ERP waveforms recorded in the three conditions: voice alone, face alone and face–voice. Waveforms in the voice condition showed two components: N100 and P200 whose latencies did not extend beyond 250 msec. Thus, calculations that involved the sum of face alone and voice alone ERP and that would go beyond 250 msec latency range would include amplitude values that did not reflect brain processes as expressed by components of the waveform. Therefore, the processes that went beyond this

Table 2
Amplitude and latency values for each component.

Components			AUDVIS (±S)		VIS		AUD	
			HC	SZ	HC	SZ	HC	SZ
Parietal-occipital	P100	Amplitude (μv)	2.21 ± 1.2	1.86 ± 1.6	2.17 ± 2.7	2.10 ± 1.8	N/A	N/A
		Latency (ms)	125.3 ± 9.9	129.7 ± 10.1	126.4 ± 9.6	131 ± 12.1	N/A	N/A
	N170	Amplitude (μv)	−4.16 ± 2.3	−2.34 ± 3.2	−4.77 ± 3.4	−2.73 ± 2.3	N/A	N/A
		Latency (ms)	180.7 ± 13.0	183.1 ± 11.2	179.5 ± 12.0	181.3 ± 13.9	N/A	N/A
	P270	Amplitude (μv)	2.81 ± 2.2	3.84 ± 3.7	3.32 ± 2.6	4.96 ± 4.9	N/A	N/A
		Latency (ms)	266.6 ± 16.7	284.0 ± 15.9	279.4 ± 20.4	277.9 ± 20.4	N/A	N/A
Frontal-central	N100	Amplitude (μv)	−3.48 ± 2.6	−3.68 ± 3.7	−3.57 ± 2.1	−4.27 ± 2.6	−4.03 ± 1.7	−4.04 ± 2.0
		Latency (ms)	126.9 ± 11.3	133.4 ± 11.0	125.9 ± 13.1	135.2 ± 9.8	122.9 ± 14.3	122.1 ± 14.3
	P200	Amplitude (μv)	2.67 ± 2.1	2.73 ± 2.2	2.57 ± 3.0	2.48 ± 2.6	2.37 ± 2.4	2.80 ± 4.0
		Latency (ms)	183.8 ± 14.2	190.6 ± 8.5	182.6 ± 9.9	188.3 ± 11.3	223.7 ± 15.5	226.6 ± 23.2
	N250	Amplitude (μ v)	−6.52 ± 2.4	−7.16 ± 3.2	−4.18 ± 2.0	−5.15 ± 2.8	N/A	N/A
		Latency (ms)	265.8 ± 9.9	267.6 ± 9.0	272.1 ± 21.3	270.8 ± 17.1	N/A	N/A
	P400	Amplitude (μv)	1.61 ± 2.4	2.40 ± 3.1	2.41 ± 2.8	0.45 ± 2.4	N/A	N/A
		Latency (ms)	386.3 ± 24.3	387.2 ± 32.4	427.9 ± 23.4	402.6 ± 25.7	N/A	N/A

HC—Healthy controls; SZ—Patients with schizophrenia.

Table 3
Significant MANOVA main effects and interactions.

	Group	Condition	Group x condition	Region	Group x region
Parietal–occipital components[a]					
P100	$F(1,33) = 0.172$, $p = 0.681$	$F(1,33) = 0.07$, $p = 0.793$	$F(1,33) = 0.127$, $p = 0.723$	N/A	N/A
	$F(1,33) = 2.762$, $p = 0.106$	$F(1,33) = 0.742$, $p = 0.395$	$F(1,33) = 0.082$, $p = 0.776$	N/A	N/A
N170	$F(1,33) = 5.38$, $p = 0.027$	$F(1,33) = 1.024$, $p = 0.319$	$F(1,33) = 0.050$, $p = 0.825$	N/A	N/A
	$F(1,33) = 0.296$, $p = 0.590$	$F(1,33) = 0.711$, $p = 0.405$	$F(1,33) = 0.035$, $p = 0.852$	N/A	N/A
P270	$F(1,33) = 1.5$, $p = 0.229$	$F(1,33) = 3.493$, $p = 0.071$	$F(1,33) = 0.474$, $p = 0.496$	N/A	N/A
	$F(1,33) = 2.603$, $p = 0.116$	$F(1,33) = 0.755$, $p = 0.391$	$F(1,33) = 6.219$, $p = 0.018$	N/A	N/A
Fronto–central components[b]					
N100	$F(1,33) = 0.262$, $p = 0.612$	$F(2,32) = 0.321$, $p = 0.728$	$F(2,32) = 0.307$, $p = 0.738$	$F(2,32) = 4.592$, $p = 0.018$	$F(2,32) = 5.258$, $p = 0.011$
	$F(1,33) = 2.722$, $p = 0.108$	$F(2,32) = 3.751$, $p = 0.034$	$F(2,32) = 1.532$, $p = 0.231$	$F(2,32) = 1.039$, $p = 0.365$	$F(2,32) = 1.547$, $p = 0.228$
P200	$F(1,33) = 0.049$, $p = 0.826$	$F(2,32) = 0.077$, $p = 0.926$	$F(2,32) = 0.065$, $p = 0.937$	$F(2,32) = 40.804$, $p < 0.001$	$F(2,32) = 2.2121$, $p = 0.126$
	$F(1,33) = 1.982$, $p = 0.168$	$F(2,32) = 78.467$, $p<0.001$	$F(2,32) = 0.153$, $p = 0.859$	$F(2,32) = 0.09$, $p = 0.914$	$F(2,32) = 0.08$, $p = 0.924$
N250	$F(1,33) = 1.082$, $p = 0.306$	$F(1,33) = 22.929$, $p < 0.001$	$F(1,33) = 0.127$, $p = 0.724$	$F(2,32) = 2.22$, $p = 0.125$	$F(2,32) = 0.151$, $p = 0.861$
	$F(1,33) = 0.005$, $p = 0.946$	$F(1,33) = 2.066$, $p = 0.160$	$F(1,33) = 0.221$, $p = 0.649$	$F(2,32) = 2.689$, $p = 0.083$	$F(2,32) = 0.748$, $p = 0.482$
P400	$F(1,33) = 0.809$, $p = 0.375$	$F(1,33) = 0.839$, $p = 0.366$	$F(1,33) = 4.798$, $p = 0.036$	$F(2,32) = 13.799$, $p<0.001$	$F(2,32) = 0.722$, $p = 0.493$
	$F(1,33) = 2.739$, $p = 0.107$	$F(1,33) = 31.071$, $p < 0.001$	$F(1,33) = 6.525$, $p = 0.015$	$F(2,32) = 0.104$, $p = 0.902$	$F(2,32) = 0.879$, $p = 0.425$

Notes: A—amplitude; L—latency; n.s.—non significant.

[a] Please note that parietal-occipital components, P100, N170 and P270, were recorded and measured in the audivisual (face–voice) and visual (face only) conditions but not in the auditory (voice only) condition.

[b] Please note that fronto–central components, N100 and P200 were recorded and measured in all three conditions, while the N250 and P400 were recorded and measured in the audiovisual (face–voice) and visual (face only) conditions but not in the auditory (voice only) condition.

latency range were assessed with traditional ERP analyses of N250 and P400 as indices of multimodal stimulus processing,

4. Results

4.1. Behavioral data

No group differences were found in response accuracy ($p > .05$ for all conditions) (AUDVIS: HC vs. SZ: 96.91% vs. 95.13%), (VIS: 97.86% vs. 97.56%) and (AUD: 80.95% vs. 79.76%, $p > 0.05$).

4.2. ERP data

4.2.1. Parieto-occipital components

P100: No significant main effect or interactions involving *condition* or *group* factors were found for P100 amplitude or latency.

N170: A main effect of *group* was observed for N170 amplitude: N170 was less negative in the schizophrenia group relative to HC (see Fig. 2 & Table 3).

P270: No significant effects were found for P270 amplitude.

A significant *group by condition* interaction was observed for P270 latency. Post-hoc analyses indicated that P270 peaked later in schizophrenia group relative to HC in the AUDVIS condition, but not in the VIS condition. Furthermore, P270 peaked earlier in the AUDVIS relative to the VIS condition in HC, with no condition differences observed in schizophrenia (see Fig. 2 & Table 3).

4.2.2. Fronto–central components

N100: A main effect of *region* and a *group by region* interaction were observed for N100 amplitude. Subsequent analyses indicated more negative N100 in frontal than fronto–central or central regions in schizophrenia, with no difference between regions found in HC (see Table 3).

The MANOVA on N100 latency revealed a main effect of *condition*: N100 peaked earlier in the AUD relative to AUDVIS and VIS conditions across groups (see Fig. 1B and C & Table 3).

P200: The MANOVA on P200 latency revealed a significant main effect of *condition*: P200 peaked later in the AUD relative to AUDVIS or VIS conditions (see Fig. 1B and 1C & Table 3).

N250: A main effect of *condition* was observed for N250 amplitude: N250 was more negative in the AUDVIS relative to the VIS condition (see Fig. 1B and C & Table 3). No significant main effect or interactions were found for N250 latency.

P400: A significant *group by condition* interaction was found: P400 amplitude for the VIS condition only was reduced in schizophrenia relative to HC (see Fig. 2 & Table 3).

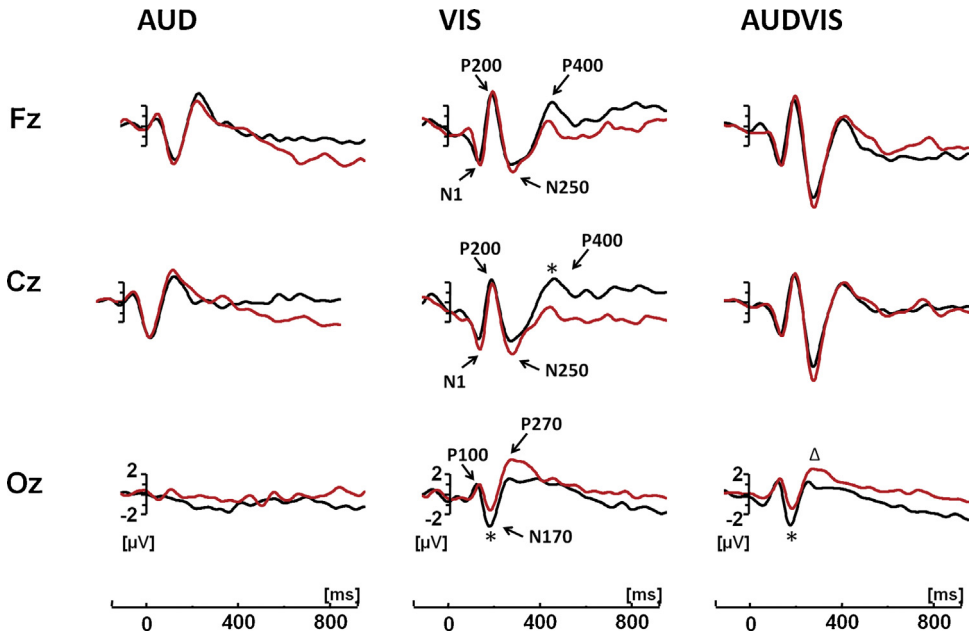

Fig. 2. Grand average waveforms in 18 SZ (red color waveform) and 19HC (black color waveform) at Fz, Cz, and Oz illustrating group comparisons for each condition. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).
*Symbol—significant amplitude group differences.
ΔSymbol—significant latency group differences.

A main effect of *condition* was observed for P400 latency: P400 peaked earlier in the AUDVIS relative to the VIS condition. Results also revealed a significant *group by condition* interaction: P400 peaked earlier in schizophrenia relative to HC in the VIS condition, but not in the AUDVIS condition (see Fig. 1B and C, Fig. 2 & Table 3).

4.2.3. The effects of antipsychotic medication

No correlations were found between chlorpromazine levels and all of the ERP components examined in this investigation (the *p* values ranged from 0.131 to 0.936).

4.2.4. Multisensory Integration

Significant effects of integration were observed in both groups across all electrodes tested: Fz, Cz, Pz, POz and Oz. At Fz, for AUDVIS vs 'sum' (AUD + VIS), over the 0–250 msec epoch, two segments of the waveform showed significant integration effects in HC (82-151; 186-250 msec) and two segments (104–143; 233–250 msec) in SZ. At Cz, two segments were found in HC (78–157; 229-250msec) and two segments in SZ (84-151; 233-250msec). At Pz, one segment was found in HC (80–157 msec) and one segment (84–149 msec) in SZ. At POz, two segments (82–133; 190–220 msec) were found in HC and one segment (94–143 msec) in SZ. At Oz, one segment (186–235 msec) was found in HC only (see Fig. 3).

No group effect was found on difference waveforms over the 0–250 msec epoch (see Fig. 3).

5. Discussion

To the best of our knowledge, this is the first study to provide electrophysiological evidence on the multimodal integration of socially relevant information from face and voice in schizophrenia. We examined the electrophysiological evidence for face–voice processing integration by comparing ERP responses to face alone, voice alone, and face–voice presented simultaneously as well as by timepoint by timepoint *t-test* approach.

Of note is that for both groups the morphology of ERP components observed in the multimodal AUDVIS (face–voice) condition was highly similar to the morphology of ERP components recorded to face only (VIS), Several studies suggested that populations of cells exist that are sensitive to both multimodal face–voice information in addition to populations of cells which are modality specific but lie close to multimodal cells, localized to superior temporal sulcus (STS) and ventral lateral prefrontal cortex [13,49]. The results from this study cannot be interpreted within the context of brain regions involved. However, they suggest, at the very least, similar processes associated with processing face alone and when co-occurring with voice information, at least for non-semantic, neutral sounds presented in this study.

In contrast, ERPs recorded to neutral voices in the voice-only condition had a different morphology from that observed in the face-only (VIS) and face–voice (AUDVIS) conditions in both groups and were similar to those reported in the previous ERP prosody studies [6,7,9].

In both traditional and timepoint by timepoint analyses, the differences between unimodal vs. multimodal stimuli were observed across several ERP components in both groups. The results suggested the presence of multisensory integration as indicated by the early components effects (N100 and P200 ERP and 0–250 msec interval in timepoint by timepoint analyses) but also by the relatively late components effects (N250 and P400). Notably, not all components were equally differentially sensitive to unimodal vs. multimodal social information.

While the evidence of multimodal integration was clear for both groups, the group differences in multimodal integration were less clear, especially in the timepoint by timepoint analyses. Below, we will discuss condition and group effects as they relate to unimodal vs multimodal processing as well as the results of time-point by time-point analyses which relate specifically to multimodal integration effects within and between groups. Finally, the complexities of both similarities and differences across conditions and groups speak to the complexities involved in the processing of socially relevant face and voice signals.

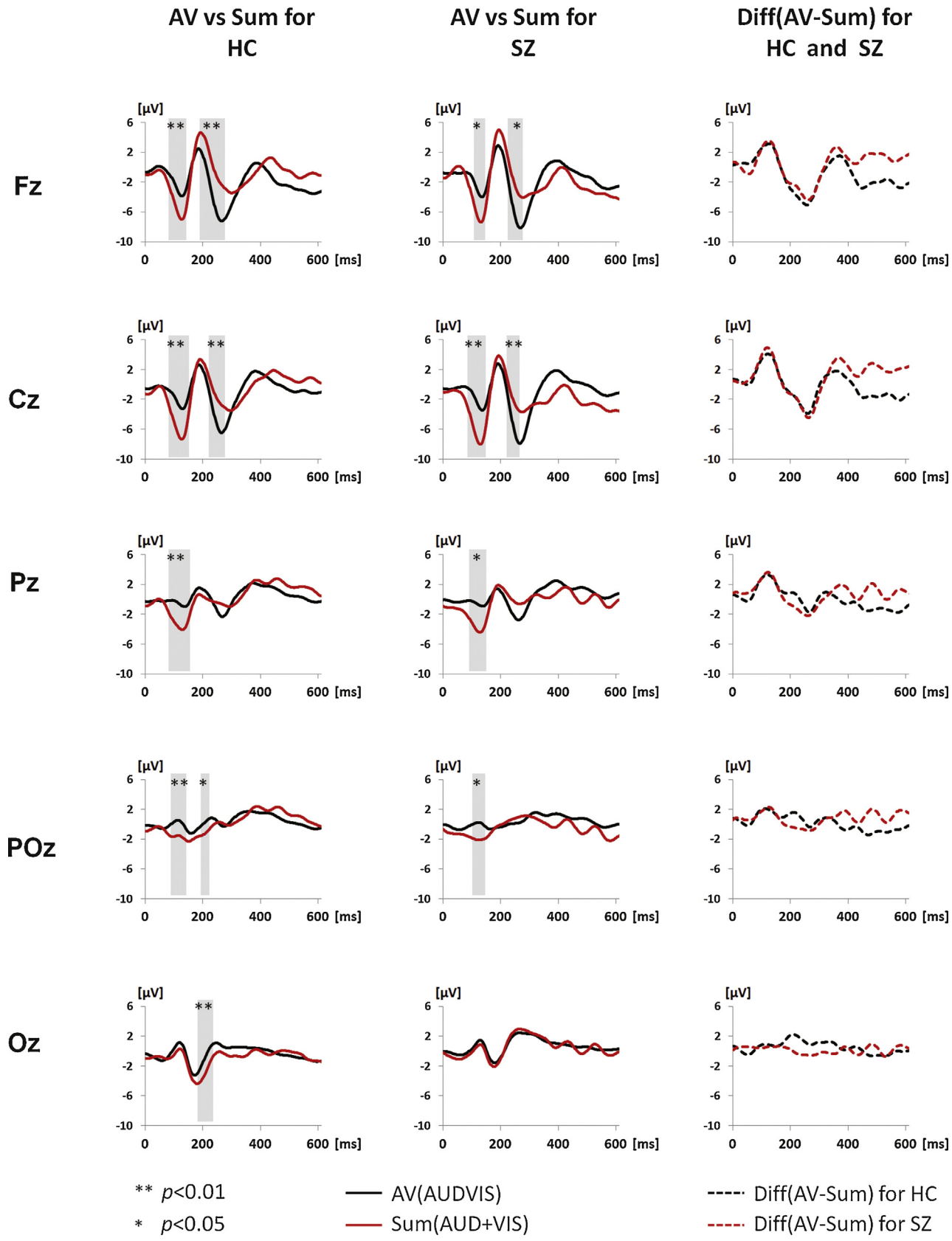

Fig. 3. Grand average waveforms in SZ and NC at Fz, Cz, Pz, POz, and Oz illustrating multisensory integration comparisons.

5.1. Significance of ERP based analyses for the understanding of unimodal and multimodal processes in healthy and schizophrenia individuals

In the parieto-occipital region, the N170 reflecting structural aspects of face processing was not sensitive to the distinction between face–voice (AUDVIS) and face-only (VIS) stimuli as mentioned above. However, the well documented deficit in the face processing in schizophrenia [50,51] was observed in this study. The N170 amplitude was reduced to both multimodal face–voice (AUDVIS) and unimodal face (VIS) stimuli in the schizophrenia group demonstrating for the first time that the face processing deficit extends to multimodal presentation of faces and non-semantic voices. The implications of a similar N170 recorded to face only and face–voice stimuli in both groups have been discussed above. The similar reduction of the N170 to both face only and face–voice stimuli in individuals experiencing schizophrenia relative to normal comparison subjects, in the context of a normal, frontally distributed N100 and P200 components whose latencies more or less coincide with the N170 may suggest that the impairment in the processing of structural face features drives the abnormality at this stage of processing. It also suggests that the impairment in the processing of structural face features is enduring and more severe than impairment in prosody processing. This may suggest that rehabilitation programs aimed at improving social cognition should target basic processes associated with structural face analysis before they move on to more complex skills.

On the other hand, the P270 indexing memory processes associated with featural face representations was sensitive to unimodal-multimodal distinctions. The P270 latency pointed to group differences: in HC, the P270 peaked earlier in the multimodal AUDVIS relative to unimodal VIS condition, while no condition differences were observed in schizophrenia suggesting that while HC benefited from the audiovisual cues, the patients did not.

Thus, at parietal sites, in traditional ERP analyses, the first potential that was sensitive to multimodal effects was P270 and the type of sensitivity was group specific as described above. The observed group differences provide valuable clues in terms of how to structure rehabilitative efforts. As described above, the first step should be training patients on structural properties of faces, and then moving on to classifying faces into groups based on emotion and facial expression. In addition, these results suggest that while the presence of voice does not hamper structural face processing further, it does so at the level of memory- based classifying processes.

In the fronto–central region, the N100, reflecting the processing of physical features, P200, reflecting the categorization of physical features, N250, reflecting emotional processing, and P400 reflecting the processing of salience, were identified. In both groups, the differential impact of unimodal relative to multimodal stimuli was observed on N100, with later N100 peak latency for face–voice (AUDVIS) relative to voice (AUD) stimuli; on P200, with earlier P200 peak latency for face–voice (AUDVIS) relative to voice (AUD) stimuli; on N250, with more negative N250 amplitude for face–voice (AUDVIS) relative to face (VIS) stimuli; and on P400 with earlier P400 peak latency for face–voice (AUDVIS) relative to face (VIS) stimuli.

Thus, in traditional ERP analyses, sensory voice features were processed faster than face-voice features, i.e., it took less time to process auditory relative to audiovisual or visual information. In contrast, for processes that entailed more complex processes such as preliminary categorization (P200) or extracting emotional salience (P400), the presence of voice made the processing of facial information faster, that is audiovisual information was processed faster than voice alone or face alone. We would also like to note that, given methodological considerations explained above (see 2.5.3), the results of traditional ERP analyses of the N250 and P400 components are the only evidence that can be offered regarding differences in the processing of uni-modal vs. multimodal stimuli (i.e., faces only vs. face–voices) at these later stages. The results suggest that the effects of processing unimodal vs multimodal stimuli are process-specific: extracting emotional meaning was more effortful as indexed by a larger N250 to face–voice stimuli relative to face only stimuli, while the presence of voice in the face–voice stimuli speeded up assigning emotional salience as indexed by a shorter latency of P400 to face–voice stimuli.

The groups differed in the P400 amplitude which was reduced in schizophrenia to faces (VIS), but not to face–voice stimuli (AUDVIS). In addition, the P400 peaked earlier in schizophrenia relative to HC to face (VIS), but not to face–voice stimuli (AUDVIS). The P400 has been related to attentional and categorization processes [52–54]. Thus, the presentation of visual (face) information alone was associated with a deficit in the patients, while the presence of congruent face–voice information normalized the categorization process. Somewhat similar effect has been reported by [32], where less impairment was found in the audiovisual relative to visual condition. As noted above, due to methodological considerations we were not able to test the effects of multimodal integration with point-by-point method beyond 250 msec post-stimulus. However, if we keep in mind the fact that the processing of emotionally neutral voice in schizophrenia was normal in this study (as it was in the earlier study [6]), the normalized P400 to the face–voice stimuli would suggest there exists a process whereby auditory and visual information interact, also at the level of attentional analyses related to salience attribution. In this specific circumstance, the normal processing of neutral voice helped assigning salience to neutral face resulting in the normalization of this process. Since most prosody studies point to abnormalities in the processing of emotional prosody, it would be interesting to examine if abnormalities in emotional voice processing further exacerbate emotional face processing.

Despite many reports of lower-level auditory deficits in schizophrenia [55–57], we did not observe group differences in N100 and P200 recorded to voice (AUD) and face–voice stimuli (AUDVIS). We note that our previous study of prosody processing using both semantic and non-semantic (distorted) sentences suggested group differences for prosody embedded in semantic but not in neutral non-semantic speech sounds [6]. This finding points to a normal processing of non-semantic *neutral* vocalizations in schizophrenia patients, both when presented alone or in combination with a face. In conjunction with our previous study [6], this finding lends support to the hypothesis that neutral prosody is processed normally by patients with schizophrenia. Furthermore, in terms of the rehabilitation efforts, this result suggests that emphasis should be put on exposing patients to stimuli with emotional content. At this point, it is not clear what are the origins of a more profound impairment in the processing of emotional stimuli: is this impairment related to specific physical features of such stimuli, or is this impairment impacted by inputs from brain emotional networks such as amygdala or insula. Future studies explicitly designed to test these different scenarios will help resolve these intriguing possibilities.

Additionally, it is plausible that the implicit task instructions influenced the sensory processing of auditory information (alone and combined with visual information) in this study.

5.2. Evidence for multisensory integration derived from timepoint by timepoint analysis

As noted in the Introduction, in imaging studies, the differential brain activity to multisensory relative to unisensory stimulus is taken as evidence for multisensory integration [19,20]. In the

ERP studies, this is tested by contrasting responses to multimodal stimuli with the sum of the responses to the unimodal stimuli.

Using this approach, in timepoint by timepoint analyses, at parieto-occipital regions, both groups showed modest effects of integration with no evidence of group differences. Robust evidence for multisensory integration was observed at the fronto–central region in both groups at latencies overlapping with N100 and P200. These early effects were consistent with subadditivity suggesting that less effort/resources were devoted to the processing of multimodal stimuli. Thus, as in several previous studies on multimodal processing using other stimuli than human faces and voices, simultaneous presentation of congruent face–voice information made the process of this social information analysis easier for early cognitive operations. These effects were observed in both groups.

As mentioned in the Method section, even though we constructed the sum of visual and auditory modality waveforms spanning 600 msec, post-stimulus, we did not think that testing for multisensory integration effects and group differences beyond the P200 latency range (250 msec) would result in legitimate condition or group effects. As noted above, waveforms to auditory stimuli consisted of N100 and P200 components, with the rest of the epoch more or less returning to baseline. Thus, the sum of auditory and visual modalities beyond 250 msec consisted of data points reflecting N250 and P400 in the visual modality and baseline activity in the waveforms to auditory stimuli. Effectively, one would compare amplitude values from visual modality to amplitude values in the audiovisual modality rather than the sum of auditory and visual modality amplitude values to amplitudes in the audiovisual modality.

No group differences were found in 0–250 msec interval. We note that in within group analyses, the two groups showed significant differences between the sum and the multimodal waveforms in non-identical latency windows.

6. Overall Conclusions

In both the traditional ERP and time-point by time-point analyses, we have demonstrated the impact of multisensory, socially relevant stimuli (faces and voices) on several processes indexed with distinct ERP components in both subject groups. These effects were focused on early processes in the fronto–central scalp regions. There were no group differences in the multisensory effects as tested with the timepoint analysis. However, these differences emerged in the traditional ERP analyses: while the N250 was more negative in the audiovisual relative to the visual condition in both groups, the P400 was reduced to faces but not to face–voice information in the schizophrenia group. As mentioned above, a somewhat similar effect reported in the Stone et al. [32] study was interpreted as reflecting a greater ease of processing congruent multisensory information in the patient group.

Interestingly, at parieto-occipital locations, the N170 was identical to both face alone and face–voice presentations and thus was not impacted by multimodal facilitation effects. In group comparisons, the face processing difficulty existed both to face alone and face–voice stimuli and extended to a lack of facilitatory effects of face–voice stimuli (AUDVIS) on P270 latency in the schizophrenia group.

Acknowledgments

This work was supported by: (1) the National Institute of Mental Health (R21 to M.N.) (2) National Natural Science Foundation of China (NSFC/31200844) awarded to LT. (3) *Fundação para a Ciência e a Tecnologia-Portugal* and FEDER (Fundo Europeu de Desenvolvimento Regional) through the European programs QREN (Quadro de Referência Estrateígico Nacional) and COMPETE (Programa Operacional Factores de Competitividade). (PTDC/PSI-PCL/116626/2010; PTDC/MHC-PCN/3606/2012; Post-Doctoral Grant no. SFRH/BD/35882/2007; and Grant IF/00334/2012 to A.P.P.).

References

[1] M.F. Green, D.L. Penn, R. Bentall, W.T. Carpenter, W. Gaebel, R.C. Gur, et al., Social cognition in schizophrenia: an NIMH workshop on definitions, assessment, and research opportunities, Schizophr. Bull. 34 (6) (2008) 1211–1220.
[2] J. Hall, J.M. Harris, R. Sprengelmeyer, A. Sprengelmeyer, A.W. Young, I.M. Santos, et al., Social cognition and face processing in schizophrenia, Br. J. Psychiatry 185 (2004) 169–170.
[3] J.F. Whittaker, J.F. Deakin, B. Tomenson, Face processing in schizophrenia: defining the deficit, Psychol Med 31 (3) (2001) 499–507.
[4] D.I. Leitman, J.J. Foxe, P.D. Butler, A. Saperstein, N. Revheim, D.C. Javitt, Sensory contributions to impaired prosodic processing in schizophrenia, Biol. Psychiatry 58 (1) (2005) 56–61.
[5] D.I. Leitman, P. Laukka, P.N. Juslin, E. Saccente, P. Butler, D.C. Javitt, Getting the cue: sensory contributions to auditory emotion recognition impairments in schizophrenia, Schizophr. Bull. 36 (3) (2010) 545–556.
[6] A.P. Pinheiro, E. Del Re, J. Mezin, P.G. Nestor, A. Rauber, R.W. McCarley, et al., Sensory-based and higher-order operations contribute to abnormal emotional prosody processing in schizophrenia: an electrophysiological investigation, Psychol. Med. 43 (3) (2013) 603–618.
[7] A.P. Pinheiro, N. Rezaii, A. Rauber, T. Liu, P.G. Nestor, R.W. McCarley, et al., Abnormalities in the processing of emotional prosody from single words in schizophrenia, Schizophr. Res. 152 (2014) 235–241.
[8] S. Campanella, P. Belin, Integrating face and voice in person perception, Trends Cogn. Sci. 11 (12) (2007) 535–543.
[9] T. Liu, A. Pinheiro, Z. Zhao, P.G. Nestor, R.W. McCarley, M.A. Niznikiewicz, Emotional cues during simultaneous face and voice processing: electrophysiological insights, PLoS One 7 (2) (2012) e31001, http://dx.doi.org/10.1371/journal.pone.0031001.
[10] B. Kreifelts, T. Ethofer, W. Grodd, M. Erb, D. Wildgruber, Audiovisual integration of emotional signals in voice and face: an event-related fMRI study, Neuroimage 37 (4) (2007) 1445–1456.
[11] G. Pourtois, B. de Gelder, A. Bol, M. Crommelinck, Perception of facial expressions and voices and of their combination in the human brain, Cortex 41 (1) (2005) 49–59.
[12] H. Holle, J. Obleser, S.A. Rueschemeyer, T.C. Gunter, Integration of iconic gestures and speech in left superior temporal areas boosts speech comprehension under adverse listening conditions, Neuroimage 49 (1) (2010) 875–884.
[13] B. Kreifelts, T. Ethofer, T. Shiozawa, W. Grodd, D. Wildgruber, Cerebral representation of non-verbal emotional perception: fMRI reveals audiovisual integration area between voice- and face-sensitive regions in the superior temporal sulcus, Neuropsychologia 47 (14) (2009) 3059–3066.
[14] F. Joassin, M. Pesenti, P. Maurage, E. Verreckt, R. Bruyer, S. Campanella, Cross-modal interactions between human faces and voices involved in person recognition, Cortex 47 (3) (2011) 367–376.
[15] S. Molholm, P. Sehatpour, A.D. Mehta, M. Shpaner, M. Gomez-Ramirez, S. Ortigue, et al., Audio–visual multisensory integration in superior parietal lobule revealed by human intracranial recordings, J. Neurophysiol. 96 (2) (2006) 721–729.
[16] S. Jessen, S.A. Kotz, The temporal dynamics of processing emotions from vocal, facial, and bodily expressions, Neuroimage 58 (2) (2011) 665–674.
[17] M. Latinus, R. VanRullen, M.J. Taylor, Top–down and bottom–up modulation in processing bimodal face/voice stimuli, BMC Neurosci. 11 (2010) 36.
[18] G.A. Calvert, Crossmodal processing in the human brain: insights from functional neuroimaging studies, Cereb. Cortex 11 (12) (2001) 1110–1123.
[19] P.J. Laurienti, T.J. Perrault, T.R. Stanford, M.T. Wallace, B.E. Stein, On the use of superadditivity as a metric for characterizing multisensory integration in functional neuroimaging studies, Exp. Brain Res. 166 (3–4) (2005) 289–297.
[20] T.J. Perrault Jr., J.W. Vaughan, B.E. Stein, M.T. Wallace, Superior colliculus neurons use distinct operational modes in the integration of multisensory stimuli, J. Neurophysiol. 93 (5) (2005) 2575–2586.
[21] O. Collignon, S. Girard, F. Gosselin, S. Roy, D. Saint-Amour, M. Lassonde, et al., Audio–visual integration of emotion expression, Brain Res. 1242 (2008) 126–135.
[22] B. de Gelder, J. Vroomen, L. Annen, E. Masthof, P. Hodiamont, Audio–visual integration in schizophrenia, Schizophr. Res. 59 (2-3) (2003) 211–218.
[23] B. de Gelder, J. Vroomen, S.J. de Jong, E.D. Masthoff, F.J. Trompenaars, P.P. Hodiamont, Multisensory integration of emotional faces and voices in schizophrenics, Schizophr. Res. 72 (2) (2005) 195–203.
[24] J.J. de Jong, P.P. Hodiamont, J. Van den Stock, B. de Gelder, Audiovisual emotion recognition in schizophrenia: reduced integration of facial and vocal affect, Schizophr. Res. 107 (2–3) (2009) 286–293.
[25] J.J. de Jong, P.P. Hodiamont, B. de Gelder, Modality-specific attention and multisensory integration of emotions in schizophrenia: reduced regulatory effects, Schizophr. Res. 122 (1-3) (2010) 136–143.

[26] L.E. Williams, G.A. Light, D.L. Braff, V.S. Ramachandran, Reduced multisensory integration in patients with schizophrenia on a target detection task, Neuropsychologia 48 (10) (2010) 3128–3136.
[27] L.A. Ross, D. Saint-Amour, V.M. Leavitt, S. Molholm, D.C. Javitt, J.J. Foxe, Impaired multisensory processing in schizophrenia: deficits in the visual enhancement of speech comprehension under noisy environmental conditions, Schizophr. Res. 97 (1-3) (2007) 173–183.
[28] J. Seubert, J. Loughead, T. Kellermann, F. Boers, C.M. Brensinger, U. Habel, Multisensory integration of emotionally valenced olfactory-visual information in patients with schizophrenia and healthy controls, J. Psychiatry Neurosci. 35 (3) (2010) 185–194.
[29] J. Van den Stock, S.J. de Jong, P.P. Hodiamont, B. de Gelder, Perceiving emotions from bodily expressions and multisensory integration of emotion cues in schizophrenia, Soc. Neurosci. 6 (5–6) (2011) 537–547.
[30] S.A. Surguladze, G.A. Calvert, M.J. Brammer, R. Campbell, E.T. Bullmore, V. Giampietro, et al., Audio-visual speech perception in schizophrenia: an fMRI study, Psychiatry Res. 106 (1) (2001) 1–14.
[31] G.R. Szycik, T.F. Munte, W. Dillo, B. Mohammadi, A. Samii, H.M. Emrich, et al., Audiovisual integration of speech is disturbed in schizophrenia: an fMRI study, Schizophr Res 110 (1-3) (2009) 111–118.
[32] D.B. Stone, L.J. Urrea, C.J. Aine, J.R. Bustillo, V.P. Clark, J.M. Stephen, Unisensory processing and multisensory integration in schizophrenia: a high-density electrical mapping study, Neuropsychologia 49 (12) (2011) 3178–3187.
[33] J.J. Stekelenburg, J.P. Maes, A.R. Van Gool, M. Sitskoorn, J. Vroomen, Deficient multisensory integration in schizophrenia: an event-related potential study, Schizophr. Res. 147 (2–3) (2013) 253–261.
[34] K. Tiippana, What is the McGurk effect? Front. Psychol. 5 (2014) 725.
[35] M.J. Herrmann, H. Ellgring, A.J. Fallgatter, Early-stage face processing dysfunction in patients with schizophrenia, Am. J. Psychiatry 161 (5) (2004) 915–917.
[36] G. Pourtois, B. de Gelder, J. Vroomen, B. Rossion, M. Crommelinck, The time-course of intermodal binding between seeing and hearing affective information, Neuroreport 11 (6) (2000) 1329–1333.
[37] R.C. Oldfield, The assessment and analysis of handedness: the Edinburgh inventory, Neuropsychologia 9 (1971) 97–113.
[38] American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders, 4th ed., Washington, DC: American Psychiatric Association Press, 2002.
[39] D. Wechsler, Wechsler Adult Intelligence Scale: Administration and Scoring Manual, 3rd ed., San Antonio, TX: The Psychological Corporation, 1997.
[40] First MB, Spitzer RL, Gibbon M, & Williams JBW. (2002). *Structured Clinical Interview for DSM-IV Axis I Diagnosis-Patient Edition (SCID-I/P, version 2.0)*. New York, NY: Biometric Research Department, New York State Psychiatric Institute.
[41] First MB, Spitzer RL, Gibbon M, & Williams JBM. (1995). *Structured Clinical Interview for DSM-IV Axis II Personality Disorders (SCID-II, version 2.0)*. New York, NY: Biometrics Research Department New York State Psychiatric Institute.
[42] F. Sharbrough, G. Chatrian, R. Lesser, H. Luders, M. Nuwer, T. Picton, American electroencephalographic society guidelines for standard electrode position nomenclature, J. Clin. Neurophysiol. 8 (2) (1991) 200–202.
[43] G. Gratton, M.G. Coles, E. Donchin, A new method for off-line removal of ocular artifact, Electroencephalogr. Clin. Neurophysiol. 55 (1983) 468–484.
[44] S. Molholm, W. Ritter, M.M. Murray, D.C. Javitt, C.E. Schroeder, J.J. Foxe, Multisensory auditory–visual interactions during early sensory processing in humans: a high-density electrical mapping study, Brain Res. Cogn. Brain Res. 14 (2002) 115–128.
[45] T. Ethofer, G. Pourtois, D. Wildgruber, Investigating audiovisual integration of emotional signals in the human brain, Prog. Brain Res. 156 (2006) 345–361.
[46] G.A. Calvert, T. Thesen, Multisensory integration: methodological approaches and emerging principles in the human brain, J. Physiol. Paris 98 (2004) 191–205.
[47] M.H. Giard, F. Peronnet, Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study? J. Cogn. Neurosci. 11 (5) (1999) 473–490.
[48] G. Guthrie, J.S. Buchwald, Significance testing of difference potentials? Psychophysiology 28 (2) (1991) 240–244.
[49] L.M. Romanski, Representation and integration of auditory and visual stimuli in the primate ventral lateral prefrontal cortex, Cereb. Cortex 17 (Suppl. 1) (2007) i61–69.
[50] T. Onitsuka, M.A. Niznikiewicz, K.M. Spencer, M. Frumin, N. Kuroki, L.C. Lucia, et al., Functional and structural deficits in brain regions subserving face perception in schizophrenia, Am. J. Psychiatry 163 (3) (2006) 455–462.
[51] B.I. Turetsky, C.G. Kohler, T. Indersmitten, M.T. Bhati, D. Charbonnier, R.C. Gur, Facial emotion recognition in schizophrenia: when and why does it go awry? Schizophr. Res. 94 (1–3) (2007) 253–263.
[52] S. Nasr, Differential impact of attention on the early and late categorization related human brain potentials, J. Vis. 10 (11) (2010) 18.
[53] A. Puce, J.A. Epling, J.C. Thompson, O.K. Carrick, Neural responses elicited to face motion and vocalization pairings, Neuropsychologia 45 (1) (2007) 93–106, http://dx.doi.org/10.1016/j.neuropsychologia.2006.04.017.
[54] A. Puce, M.E. McNeely, M.E. Berrebi, J.C. Thompson, J. Hardee, J. Brefczynski-Lewis, Multiple faces elicit augmented neural activity, Front. Hum. Neurosci. 7 (2013) 282, http://dx.doi.org/10.3389/fnhum.2013.00282.
[55] D.C. Javitt, A.M. Shelley, G. Silipo, J.A. Lieberman, Deficits in auditory and visual context-dependent processing in schizophrenia: defining the pattern, Arch. Gen. Psychiatry 57 (12) (2000) 1131–1137.
[56] P.T. Michie, What has MMN revealed about the auditory system in schizophrenia? Int. J. Psychophysiol 42 (2) (2001) 177–194.
[57] T. Rosburg, N.N. Boutros, J.M. Ford, Reduced auditory evoked potential component N100 in schizophrenia—a critical review, Psychiatry Res. 161 (3) (2008) 259–274.
[58] S.W. Woods, Chlorpromazine equivalent doses for the newer atypical antipsychotics, J. Clin. Psychiatry 64 (6) (2003) 663–667.
[59] Stoll, A.L. The Psychopharmacology Reference Card. 1989–2001.
[60] S.R. Kay, L.A. Opler, A. Fizbein, Positive and Negative Syndrome Scale (PANSS) Manual, North Tonawanda, New York, M ulti Health Systems, Inc, 1986.