Sensory-based and higher-order operations contribute to abnormal emotional prosody processing in schizophrenia: an electrophysiological investigation

A. P. Pinheiro[1,2], E. del Re[2], J. Mezin[2], P. G. Nestor[2,3], A. Rauber[4], R. W. McCarley[2], Ó. F. Gonçalves[1] and M. A. Niznikiewicz[2]*

[1] *Neuropsychophysiology Laboratory, CiPsi, School of Psychology, University of Minho, Braga, Portugal*

[2] *Clinical Neuroscience Division, Laboratory of Neuroscience, Department of Psychiatry, Boston VA Healthcare System, Brockton Division and Harvard Medical School, Boston, MA, USA*

[3] *University of Massachusetts, Boston, MA, USA*

[4] *Phonetics Laboratory, Catholic University of Pelotas, Pelotas, Brazil*

Background. Schizophrenia is characterized by deficits in emotional prosody (EP) perception. However, it is not clear which stages of processing prosody are abnormal and whether the presence of semantic content contributes to the abnormality. This study aimed to examine event-related potential (ERP) correlates of EP processing in 15 chronic schizophrenia individuals and 15 healthy controls.

Method. A total of 114 sentences with neutral semantic content [sentences with semantic content (SSC) condition] were generated by a female speaker (38 with happy, 38 with angry, and 38 with neutral intonation). The same sentences were synthesized and presented in the 'pure prosody' sentences (PPS) condition where semantic content was unintelligible.

Results. Group differences were observed for N100 and P200 amplitude: patients were characterized by more negative N100 for SSC, and more positive P200 for angry and happy SSC and happy PPS. Correlations were found between delusions and P200 amplitude for happy SSC and PPS. Higher error rates in the recognition of EP were also observed in schizophrenia: higher error rates in neutral SSC were associated with reduced N100, and higher error rates in angry SSC were associated with reduced P200.

Conclusions. These results indicate that abnormalities in prosody processing occur at the three stages of EP processing, and are enhanced in SSC. Correlations between P200 amplitude for happy prosody and delusions suggest a role that abnormalities in the processing of emotionally salient acoustic cues may play in schizophrenia symptomatology. Correlations between ERP and behavioral data point to a relationship between early sensory abnormalities and prosody recognition in schizophrenia.

Introduction

Schizophrenia is a psychiatric disorder associated with several abnormalities at the cognitive, behavioral and brain levels (Shenton *et al.* 2001; Wible *et al.* 2009). Recently, social cognition impairments have been reported in schizophrenia, including deficits in emotional prosody perception as observed in both behavioral and functional magnetic resonance imaging investigations (Shaw *et al.* 1999; Edwards *et al.* 2001; Kucharska-Pietura *et al.* 2005; Bozikas *et al.* 2006; Hoekert *et al.* 2007; Shea *et al.* 2007; Leitman *et al.*

2011). An association between inefficient prosody processing in schizophrenia and clinical symptomatology that included both positive (Rossell & Boundy, 2005; Shea *et al.* 2007) and negative symptoms (e.g. Leitman *et al.* 2005) was noted in several studies, while other studies did not find this association (e.g. Kucharska-Pietura *et al.* 2005). In addition, recent studies have highlighted the contributions of sensory auditory abnormalities (e.g. pitch perception) to prosody dysfunction in schizophrenia (e.g. Leitman *et al.* 2010).

Emotional prosody represents the non-verbal vocal expression of emotion and its perception is a multi-stage process (Schirmer & Kotz, 2006; Wildgruber *et al.* 2006; Paulmann & Kotz, 2008*a, b*; Paulmann *et al.* 2009). According to the three-stage model of prosody processing developed by Schirmer & Kotz (2006), the

* Address for correspondence: M. Niznikiewicz, Ph.D., Department of Psychiatry-116A, Boston VA Healthcare System, 940 Belmont Street, Brockton, MA 02301, USA.

(Email: margaret_niznikiewicz@hms.harvard.edu)

first stage (around 100 ms) is related to the sensory processing of the acoustic signal. It is followed by the detection of emotionally salient acoustic cues (around 200 ms), and by the cognitive evaluation of the emotional significance of the vocal information (after 300 ms). The first stage of emotional prosody processing is mediated by the bilateral secondary auditory cortex (e.g. Rauschecker, 1998; Hart *et al.* 2003), and is indexed by the N100 component (Paulmann & Kotz, 2008*b*; Paulmann *et al.* 2009). In fact, a recent study demonstrated that the sensitivity to emotional salience can occur as early as 100 ms and is indexed by N100 amplitude modulated by the emotional valence of non-verbal vocalizations (Liu *et al.* 2012).

The second stage recruits temporal areas, including the superior temporal gyrus and the anterior temporal sulcus (Kotz *et al.* 2003; Mitchell *et al.* 2003; Grandjean *et al.* 2005). Electrophysiologically, the P200 component was found to index the detection of emotional salience from speech stimuli (Paulmann & Kotz, 2008; Paulmann *et al.* 2009). Finally, the last stage recruits frontal areas, including the inferior frontal gyrus and the orbito-frontal cortex (Buchanan *et al.* 2000; Gandour *et al.* 2003; Wildgruber *et al.* 2005). While the third stage cannot be probed with event-related potential (ERP) methodology, behavioral data can shed some light on these integrative processes (Paulmann & Kotz, 2008*b*; Paulmann *et al.* 2009). Importantly, these stages are reciprocally connected, in the sense that sensory stages make an impact on higher-order processes, and top-down mechanisms (e.g. attention) may modulate sensory processes, as demonstrated in both healthy subjects and schizophrenia patients (e.g. Ethofer *et al.* 2006; Leitman *et al.* 2010, 2011).

At perceptual and physical levels, emotional prosody is instantiated by intensity, pitch (fundamental frequency; f_0), speech rhythm (duration of syllables and pauses) and voice quality/timbre (Schirmer & Kotz, 2006; Wildgruber *et al.* 2006). Each emotion seems to have a particular acoustic profile (Banse & Scherer, 1996). There are only two published ERP studies of processing prosody using a 'naturalistic' design where sentences are delivered with either neutral or emotional intonation without introducing discrepancy between sentence fragments, or between message and the tone with which it was delivered (Paulmann & Kotz, 2008*b*; Paulmann *et al.* 2009). In both studies the ERP prosody effects were found within the P200 latency window, with no late latency components found sensitive to prosody changes.

In spite of consistently reported deficits in emotional prosody discrimination in schizophrenia, no ERP studies of prosody processing have been conducted in schizophrenia. In the present study we used

a 'naturalistic' paradigm, similar to the studies by Paulmann & Kotz (2008) and Paulmann *et al.* (2009), to investigate the temporal course of emotional prosody processing in schizophrenia. In line with previous studies in non-clinical subjects, we expected the prosody effects to be indexed by N100 and P200. Given the functional significance of these two components, we expected to probe the first two stages of prosody processing. Behavioral data were expected to provide an indirect probe of the third stage of prosody processing.

We used both sentences with semantic content (SSC) and 'pure prosody' sentences (PPS) where the semantic content was 'unintelligible'. Prosodic information that is carried in a speech signal by the dynamic combination of different acoustic parameters normally co-exists with semantic information. However, it is not fully understood how lexical and supra-segmental features of the speech signal may interact to convey emotion and how such processes differ from the processes of extracting emotional information from a speech signal that carries supra-segmental information alone, in healthy controls and in schizophrenia.

Our primary hypothesis was that schizophrenia patients would show deficits at all three stages of prosody processing. Our secondary hypothesis was that the first two stages of prosody processing indexed by N100 and P200 would be more impaired in the patient group in the SSC. Based on consistent reports of disrupted sensory processing of prosodic acoustic cues in schizophrenia (Leitman *et al.* 2005, 2011) and of abnormalities in processing language-specific cues (e.g. Niznikiewicz *et al.* 2010), we expected that prosodic abnormalities would be more pronounced for more complex speech stimuli containing lexical–semantic cues than for stimuli relying on purely prosodic cues.

Specifically, given previous reports of difficulties in prosody discrimination in schizophrenia that have been related to sensory abnormalities (e.g. pitch perception), we hypothesized a lack of ERP differentiation between different prosody types, as indexed by similar P200 amplitude to the three types of emotional prosody (Paulmann & Kotz, 2008*b*). Additionally, we expected that difficulties in evaluating the emotional significance of speech stimuli would be reflected in reduced accuracy in schizophrenia when compared with healthy control individuals.

Finally, given previous studies suggesting an association between deficits in emotional prosody discrimination and positive symptomatology in general (Poole *et al.* 2000), as well as with auditory verbal hallucinations in particular (Rossell & Boundy, 2005; Shea *et al.* 2007) or negative symptomatology

(Leitman *et al.* 2005, 2010), we predicted that N100 and P200 amplitude would be associated with clinical scores.

Method

Participants

A total of 15 right-handed males diagnosed with chronic schizophrenia and 15 healthy male controls matched for age, handedness and parental socio-economic status participated in the experiment (Table 1). Comparison subjects were recruited from advertisements in local newspapers.

The inclusion criteria were: English as first language; right handedness (Oldfield, 1971); no history of neurological illness; no history of Diagnostic and Statistical Manual of Mental Disorders, fourth edition, text revision (DSM-IV) diagnosis of drug or alcohol abuse (APA, 2000); verbal intelligence quotient (IQ) above 75 (Wechsler, 1997); no hearing, vision or upper body impairment. For healthy controls, an additional exclusion criterion was a history of psychiatric disorder in oneself or in first-degree relatives. Patients were diagnosed (for healthy controls, screened) using the Structured Clinical Interview for DSM-IV for Axis I (First *et al.* 2002) and Axis II (First *et al.* 1995) disorders.

Before participation in the study, all participants had the procedures fully explained to them and read and 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).

Stimuli

Stimuli were 228 auditory sentences presented in a pseudo-randomized order: 114 sentences were spoken by a female speaker of American English with training in theatre techniques. The recordings were made in a quiet room with an Edirol R-09 recorder and a CS-15 cardioid-type stereo microphone (Eridol, USA), with a sampling rate of 22 kHz and 16-bit quantization.

All sentences had neutral semantic content, similar syntactic structure (subject + verb + object) and length (four words), and all started with a proper noun (see Appendix Fig. A1). A total of 38 sentences were spoken with happy, 38 with angry, and 38 with neutral intonation. Auditory stimuli were acoustically analysed using Praat (version 5.0.43; P. Boersma and D. Weenink, The Netherlands; http://www. fon. hum. uva.nl/praat/). Mean pitch, intensity and duration were compared across conditions (Fig. 1). There were no differences between emotional categories in mean intensity. However, significant differences across emotional categories were found for: minimum pitch ($p < 0.0001$), maximum pitch ($p < 0.0001$), mean pitch ($p < 0.0001$) and mean duration ($p < 0.0001$).

A total of 18 subjects (11 female) not participating in the ERP experiment assessed the emotional valence of the sentences. Angry sentences were rated as 'angry' by 94.08% (S.D. = 7.72), happy sentences were rated as 'happy' by 91.44% (S.D. = 7.54), and neutral sentences were rated as 'neutral' by 99.87% (S.D. = 0.85) of subjects. Only sentences appropriately rated by at least 90% of participants were included in the stimuli list (SSC condition). For the synthesis of sentences used in the PPS condition, we followed the procedure described in Pinheiro *et al.* (2011) (see Fig. 1). In order to ensure that the PPS condition was indeed devoid of semantic content, 10 volunteers listened to the PPS and none of them was able to identify semantic meaning in these sentences. These same volunteers (six female) rated the emotional valence of PPS. Angry PPS sentences were rated as 'angry' by 55% of subjects, happy PPS sentences were rated as 'happy' by 65% of subjects, and neutral PPS sentences were rated as 'neutral' by 89% of subjects.

Procedure

Each participant was seated comfortably at a distance of 100 cm from a computer monitor in a sound-attenuating chamber. The experimental session was divided into two blocks (block 1 – SSC; block 2 – PPS), each containing 114 pseudorandomized sentences (38 in each emotional prosody type). All sentences were presented binaurally through headphones at a sound level comfortable for each subject, and were not repeated during the experiment. Stimuli presentation, timing of events and recording of subjects' responses were controlled by the Superlab Pro software package (2008; http://www.superlab.com/). Before each experimental block, participants were given a brief training with feedback, to make sure that they understood the instructions and became familiarized with the task and with the response box.

Before each sentence onset, a fixation cross was presented centrally on the screen for 1000 ms and remained there during sentence presentation to minimize eye movements. At the end of the sentence, the cross was replaced by a blank screen and, 1500 ms later, a question mark appeared for 5 s (see Fig. 1 for illustration of the task structure). Participants were asked to decide if the sentence was spoken in a neutral, positive or negative intonation by pressing one of three keys (the order of the keys counterbalanced

Table 1. *Characteristics of healthy controls and schizophrenia participants*

	Participants		Group comparisons	
	Healthy controls ($n=15$)	Schizophrenia participants ($n=15$)	F	p
Sociodemographic data				
Age, years	44.73 (9.57)	49.73 (9.08)	2.155	0.153
Educational level, years	14.77 (1.43)	14.33 (1.80)	0.535	0.471
Parents' socio-economic status[a]	2.33 (1.11)	2.67 (1.40)	0.522	0.476
Subject socio-economic status	1.80 (0.86)	3.13 (0.99)	15.470	0.001*
Race			N.A.	
Caucasian	9	11		
African-American	5	4		
Asian-American	1	0		
Cognitive data (WAIS-III)[b]				
Verbal IQ	108.55 (17.46)	94.85 (28.21)	1.952	0.176
Full-scale IQ	110.30 (14.34)	92.23 (29.44)	3.163	0.090
Verbal Comprehension Index	97.18 (34.25)	97.92 (30.30)	0.003	0.956
Working Memory Index	93.10 (37.02)	82.00 (31.30)	0.607	0.445
Clinical data				
Duration of illness	N.A.	22.33 (10.54)	N.A.	
Medication, mg chlorpromazine equivalents	N.A.	331.05 (170.92)	N.A.	
First generation (zuclopenthixol, flupenthixol), n	N.A.	2	N.A.	
Second generation (risperidone, clozapine, quetiapine, olanzapine, amilsupiride), n	N.A.	12	N.A.	
Third generation (aripiprazole), n	N.A.	0	N.A.	
Other psychoactive drugs	N.A.		N.A.	
Benzodiazepines (clonazepam), n	N.A.	2	N.A.	
Anticonvulsivant mood-stabilizers (carbamazapine, lamotrigine, divalproex sodium, lithium), n	N.A.	5	N.A.	
Antidepressants (bupropion, sertraline), n	N.A.	2	N.A.	
PANSS data[c]	N.A.			
Total score	N.A.	71.47 (28.62)	N.A.	
General score	N.A.	33.80 (15.07)	N.A.	
Positive syndrome scale score	N.A.	19.87 (8.97)	N.A.	
Negative syndrome scale score	N.A.	17.80 (7.49)	N.A.	
Conceptual disorganization score	N.A.	2.87 (1.92)	N.A.	
Delusions score	N.A.	4.53 (2.03)	N.A.	
Hallucinations score	N.A.	4.13 (2.42)	N.A.	
Total SAPS	N.A.	9.36 (4.07)	N.A.	
Total SANS	N.A.	8.68 (5.36)	N.A.	

Data are given as mean (standard deviation) or number of participants.

N.A., Non-applicable; WAIS, Wechsler Adult Intelligence Scale; IQ, intelligence quotient; PANSS, Positive and Negative Syndrome Scale; SAPS, Scale for the Assessment of Positive Symptoms; SANS, Scale for the Assessment of Negative Symptoms.

[a] Hollingshead Two-Factor Index of Social Position (Hollingshead, 1965). [b] Wechsler (1997). [c] Kay *et al.* (1987).

* Significant.

across subjects). Each response key was marked with an emoticon to minimize working memory demands. A short pause was provided after 57 sentences. No feedback was provided. The experimental session lasted 45 min.

Data acquisition and analysis

Electroencephalogram (EEG) recording procedure

The EEG was recorded with 64 electrodes mounted on a custom-made cap (Electro-cap International, USA),
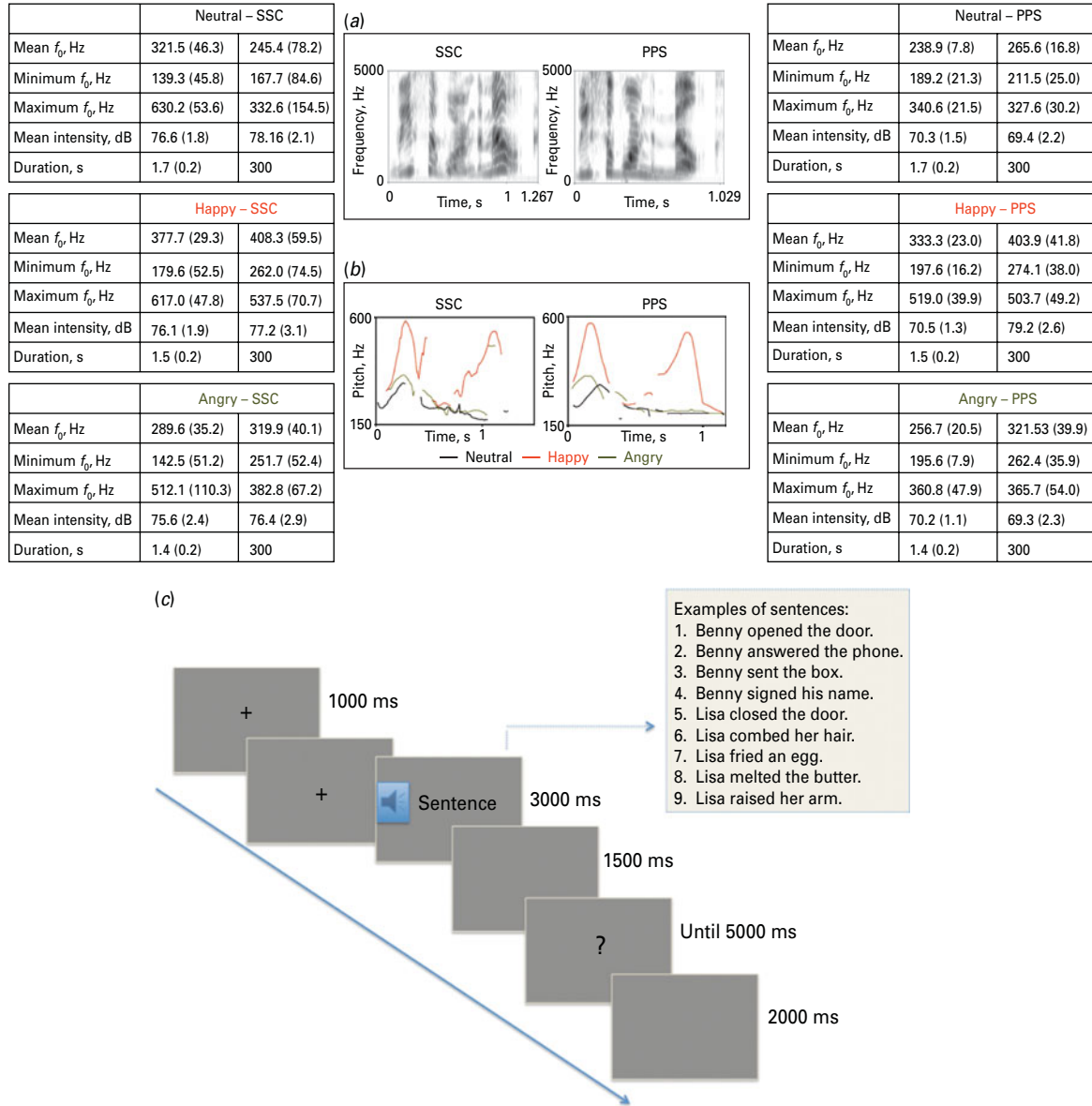
Neutral – SSC		
Mean f_0, Hz	321.5 (46.3)	245.4 (78.2)
Minimum f_0, Hz	139.3 (45.8)	167.7 (84.6)
Maximum f_0, Hz	630.2 (53.6)	332.6 (154.5)
Mean intensity, dB	76.6 (1.8)	78.16 (2.1)
Duration, s	1.7 (0.2)	300

Happy – SSC		
Mean f_0, Hz	377.7 (29.3)	408.3 (59.5)
Minimum f_0, Hz	179.6 (52.5)	262.0 (74.5)
Maximum f_0, Hz	617.0 (47.8)	537.5 (70.7)
Mean intensity, dB	76.1 (1.9)	77.2 (3.1)
Duration, s	1.5 (0.2)	300

Angry – SSC		
Mean f_0, Hz	289.6 (35.2)	319.9 (40.1)
Minimum f_0, Hz	142.5 (51.2)	251.7 (52.4)
Maximum f_0, Hz	512.1 (110.3)	382.8 (67.2)
Mean intensity, dB	75.6 (2.4)	76.4 (2.9)
Duration, s	1.4 (0.2)	300

Neutral – PPS		
Mean f_0, Hz	238.9 (7.8)	265.6 (16.8)
Minimum f_0, Hz	189.2 (21.3)	211.5 (25.0)
Maximum f_0, Hz	340.6 (21.5)	327.6 (30.2)
Mean intensity, dB	70.3 (1.5)	69.4 (2.2)
Duration, s	1.7 (0.2)	300

Happy – PPS		
Mean f_0, Hz	333.3 (23.0)	403.9 (41.8)
Minimum f_0, Hz	197.6 (16.2)	274.1 (38.0)
Maximum f_0, Hz	519.0 (39.9)	503.7 (49.2)
Mean intensity, dB	70.5 (1.3)	79.2 (2.6)
Duration, s	1.5 (0.2)	300

Angry – PPS		
Mean f_0, Hz	256.7 (20.5)	321.53 (39.9)
Minimum f_0, Hz	195.6 (7.9)	262.4 (35.9)
Maximum f_0, Hz	360.8 (47.9)	365.7 (54.0)
Mean intensity, dB	70.2 (1.1)	69.3 (2.3)
Duration, s	1.4 (0.2)	300

Examples of sentences:
1. Benny opened the door.
2. Benny answered the phone.
3. Benny sent the box.
4. Benny signed his name.
5. Lisa closed the door.
6. Lisa combed her hair.
7. Lisa fried an egg.
8. Lisa melted the butter.
9. Lisa raised her arm.

Fig. 1. (*a*) Example of a wide-band spectrogram of a speech signal for a happy prosody sentence ('Lisa warmed the milk'), before (SSC) and after (PPS) concatenative synthesis. SSC (sentences with semantic content condition) illustrates the frequency spectrum (0–5 kHz) of a sentence with semantic content. PPS ('pure prosody' sentences condition) illustrates the frequency spectrum of a transformed sentence ('pure prosody' condition). The spectral information is similar across conditions; sentences sounded as natural as possible, but no intelligible semantic information was present in the 'pure prosody' condition. (*b*) Pitch contour of speech signals before (SSC) and after (PPS) concatenative synthesis, for each of the prosody types (happy, angry and neutral). Acoustic properties of neutral, happy and angry SSC and PPS are also shown, first for the whole sentences of each valence (column 2), and then for the first 300 ms of the sentences (column 3). Data are given as mean (standard deviation). In the PPS condition, the phones of each sentence (from the list of 114 'natural' SSC sentences) were manually segmented in Praat (version 5.0.43; P. Boersma and D. Weenink, The Netherlands; http://www.fon.hum.uva.nl/praat/). Fundamental frequency (f_0) was automatically extracted in Praat at four points of each segment (20%, 40%, 60% and 80%). Occasional f_0 error measurements were manually corrected. Based on procedures of Ramus & Mehler (1999), duration and f_0 values were then transferred to MBROLA (Dutoit *et al.* 1996) for concatenative synthesis by using the American English (female) diphone database. All fricatives were replaced with the phoneme /s/, all stop consonants with /t/, all glides with /j/, all stressed vowels with /æ/ and all unstressed vowels with /ə/, assuring that the synthesis of new sentences preserved characteristics such as global intonation, syllabic rhythm and broad phonotactics (Ramus & Mehler, 1999). This technique, in comparison with the filtered speech approach, creates more natural sentences by eliminating intelligible lexical–semantic content while preserving emotional prosody. (*c*) Illustration of an experimental trial. All sentences had neutral semantic content, similar length (four words) and simple syntactic complexity (subject–verb–object), describing actions that can occur in daily life.
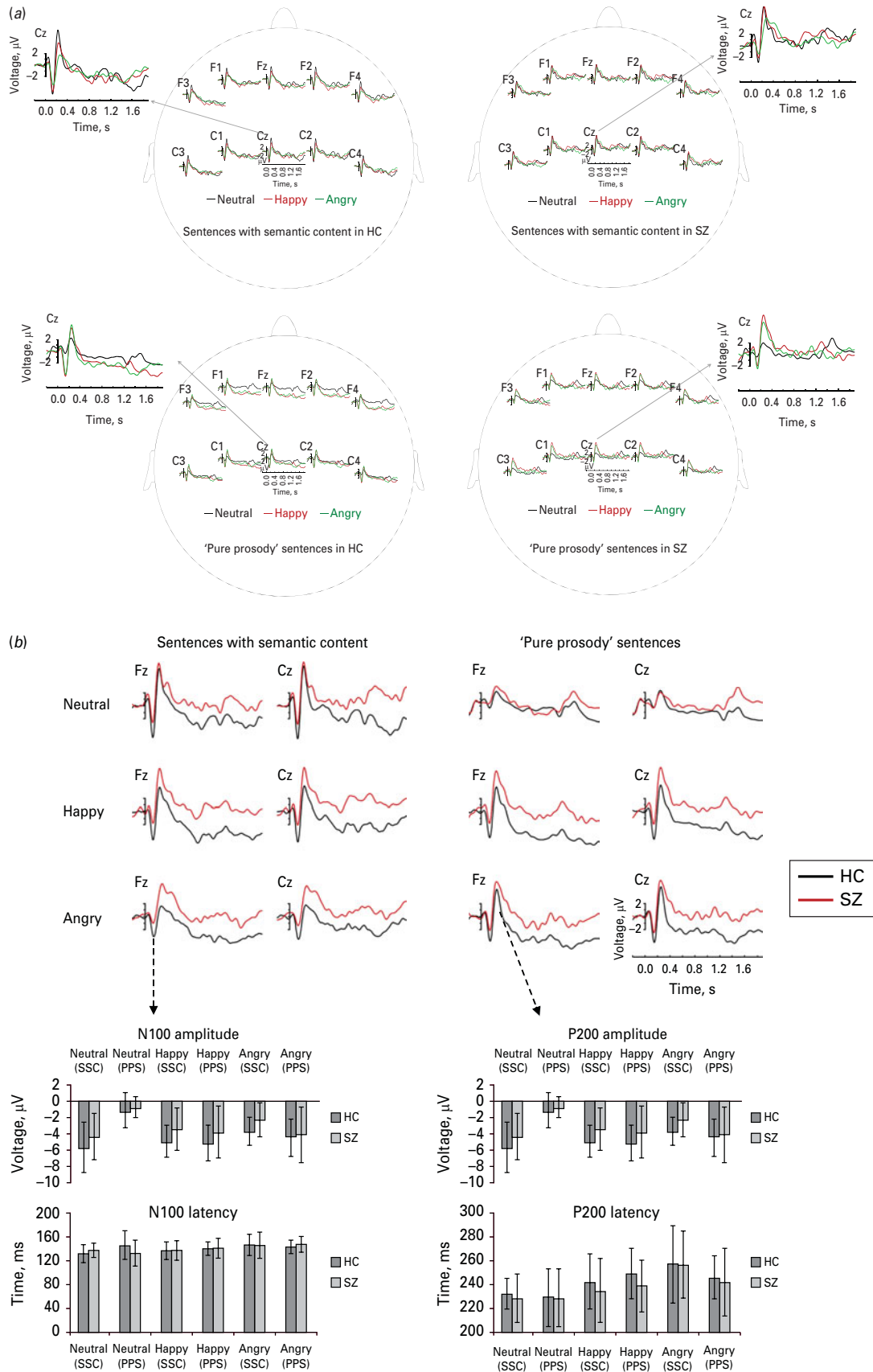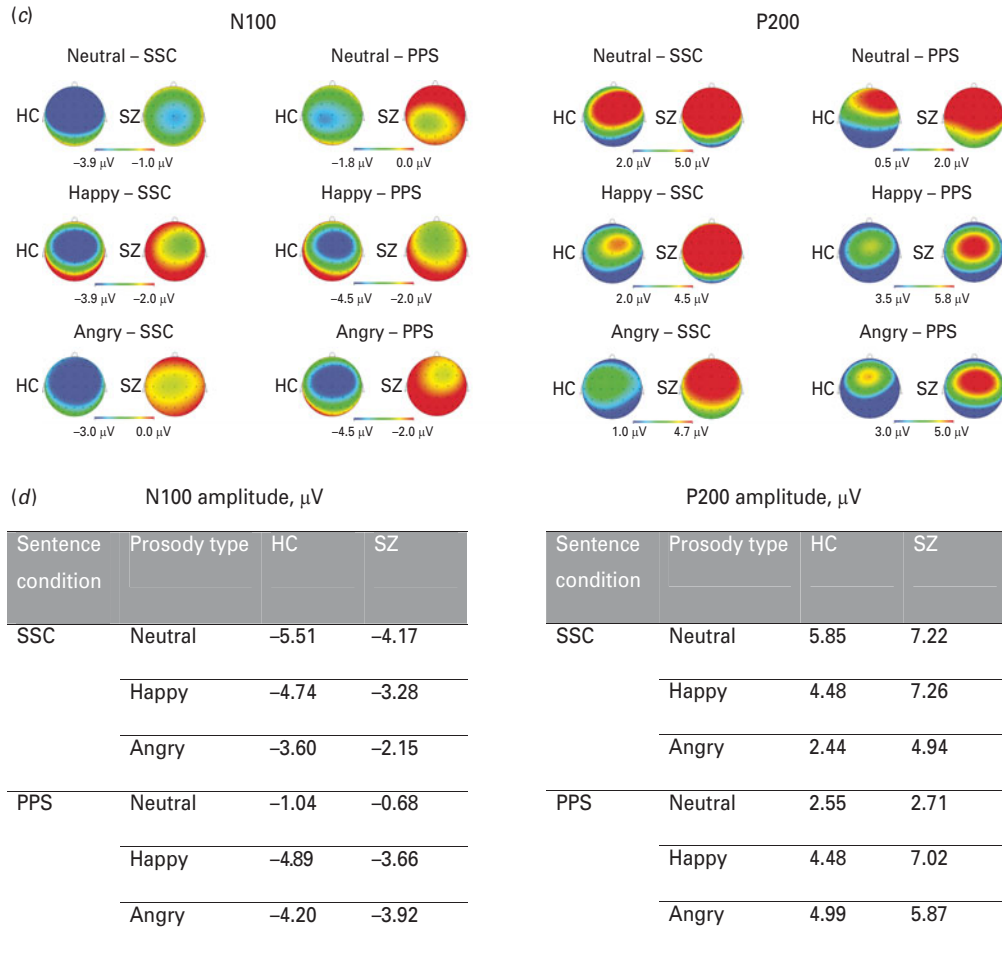
Fig. 2. For legend see opposite.

(c)

N100

Neutral – SSC	Neutral – PPS
HC ... SZ	HC ... SZ
−3.9 µV −1.0 µV	−1.8 µV 0.0 µV

Happy – SSC	Happy – PPS
HC ... SZ	HC ... SZ
−3.9 µV −2.0 µV	−4.5 µV −2.0 µV

Angry – SSC	Angry – PPS
HC ... SZ	HC ... SZ
−3.0 µV 0.0 µV	−4.5 µV −2.0 µV

P200

Neutral – SSC	Neutral – PPS
HC ... SZ	HC ... SZ
2.0 µV 5.0 µV	0.5 µV 2.0 µV

Happy – SSC	Happy – PPS
HC ... SZ	HC ... SZ
2.0 µV 4.5 µV	3.5 µV 5.8 µV

Angry – SSC	Angry – PPS
HC ... SZ	HC ... SZ
1.0 µV 4.7 µV	3.0 µV 5.0 µV

(d)

N100 amplitude, µV

Sentence condition	Prosody type	HC	SZ
SSC	Neutral	−5.51	−4.17
	Happy	−4.74	−3.28
	Angry	−3.60	−2.15
PPS	Neutral	−1.04	−0.68
	Happy	−4.89	−3.66
	Angry	−4.20	−3.92

P200 amplitude, µV

Sentence condition	Prosody type	HC	SZ
SSC	Neutral	5.85	7.22
	Happy	4.48	7.26
	Angry	2.44	4.94
PPS	Neutral	2.55	2.71
	Happy	4.48	7.02
	Angry	4.99	5.87

Fig. 2. (*a*) Grand averaged waveforms for neutral, happy and angry prosody in sentences with semantic content (SSC) and 'pure prosody' sentences (PPS), in both healthy controls (HC) and schizophrenia patients (SZ). (*b*) Group comparisons for neutral, happy and angry prosody in SSC and PPS. (*c*) Topographic maps showing the scalp distribution of N100 and P200 voltage for neutral, happy and angry prosody in SSC and PPS, in both groups. (*d*) Mean amplitude for N100 and P200 event-related potential components across sentence conditions (SSC and PPS) and prosody types (neutral, happy and angry) in both HC and SZ patients.

according to the modified expanded 10–20 system (American Electroencephalographic Society, 1991) using the Biosemi system (Active 2). The EEG was acquired in a continuous mode at a digitization rate of 512 Hz, with a bandpass of 0.01 to 100 Hz. Data were re-referenced offline to the mathematical average of the mastoids. Horizontal and vertical electro-oculograms were recorded for eye movement and blink detection and rejection, via electrodes placed on the left and right temples and one below the left eye. Electrode impedances were kept below 5 kΩ.

EEG data analysis

The EEG data were processed using Brain Analyzer software (Brain Products GmbH, Germany). EEG epochs containing eye blinks or movement artifacts exceeding $\pm 100\,\mu$V were not included in individual ERP averages. After artifact rejection, at least 75% of trials per condition per subject entered the analyses. Separate ERPs for each condition were created for each participant. Averages were computed using a 200-ms pre-stimulus baseline, time-locked to the sentence onset and spanning the length of a sentence (1800 ms). This approach was adopted following all existing ERP studies on emotional prosody processing using non-spliced sentences (Paulmann & Kotz, 2008*b*; Paulmann *et al*. 2009) to ensure the comparability of our results with previous studies. Additionally, as a careful examination of the f_0 (that carries the majority of the prosodic information – e.g. Scherer, 1995; Banse & Scherer, 1996) distribution

illustrates, the first 300 ms of the sentence duration included all major shifts in f_0 sufficiently contributing to prosody recognition.

After the inspection of grand average waveforms and following previous ERP studies of emotional prosody processing (Paulmann & Kotz, 2008*b*; Paulmann *et al.* 2009), the N100 and P200 were selected for analysis. The N100 was measured as the most negative data point between 100 and 200 ms post-stimulus, and P200 was measured as the most positive data point between 200 and 300 ms. Since maximal effects were observed at fronto-central sites (Fig. 2), consistent with previous reports (Paulmann & Kotz, 2008*b*), N100 and P200 were measured at frontal (Fz, F3/4) and central electrodes (Cz, C3/4) and all these electrodes were entered into analyses reported below.

Statistical analyses

ERP and behavioral data

For ERP data, multivariate analyses of variance (MANOVAs) were computed for the between-group comparisons of N100 and P200 peak amplitude, for SSC and PPS separately, with group as a between-subjects factor and sentence condition (SSC, PPS) and emotion (neutral, happy, angry) as within-subjects factors, using SPSS 20.0 (SPSS, Corp., USA). Within-group comparisons in both SSC and PPS served to characterize patterns of ERP differences in emotion processing in each group that were not captured in between-group comparisons. Finally, in order to test for hemispheric differences, we conducted an exploratory analysis of hemisphere (left: F3, C3; right: F4, C4) effects. The MANOVA model was identical to that described above with the hemisphere used as one of the within-subjects factors. Accuracy data were subjected to MANOVAs with sentence condition and emotion as within-subjects factors, and group as a between-subjects factor.

For all significant main effects and interactions, pairwise comparisons were run using the Sidak correction.

Correlational analyses

Spearman's ρ correlations were performed in an exploratory analysis of the relationship between N100 and P200 amplitudes (at frontal and central electrodes) in both SSC and PPS conditions and Positive and Negative Syndrome Scale subscales scores (Kay *et al.* 1987). In addition, we correlated N100 and P200 amplitude with mean equivalent chlorpromazine dosage, as well as with illness duration, to test for effects of medication and chronicity and with behavioral

data to test for the association between early sensory processes and behavioral indices of prosody recognition. All significance levels are two-tailed with the preset significance α level of $p < 0.05$.

Results

ERP results

N100 amplitude

The omnibus MANOVA yielded a main effect of sentence condition ($F_{1,28} = 8.451$, $p < 0.01$) and emotion ($F_{2,27} = 9.126$, $p < 0.01$). Planned comparisons showed that N100 was more negative in SSC than in PPS ($p < 0.01$), and for happy relative to neutral sentences ($p < 0.01$). No hemispheric differences were observed ($p > 0.05$). Importantly, a group × sentence condition × emotion interaction was found ($F_{2,27} = 0.628$, $p = 0.051$). We followed this interaction with MANOVAs for each sentence condition separately, and then for each emotion in each sentence condition separately.

In between-group comparisons for SCC, a main effect of group was found ($F_{1,28} = 5.101$, $p = 0.032$); N100 was less negative in patients than in healthy controls. In particular, groups differed for angry SCC ($F_{1,28} = 4.152$, $p = 0.051$) and tended to differ for happy SSC ($F_{1,28} = 3.325$, $p = 0.079$). No main effect of group or interaction involving the group factor was revealed in between-group comparisons for PPS.

In within-group comparisons for SSC, an emotion effect was observed in healthy controls ($F_{2,13} = 3.516$, $p = 0.024$) – N100 was more negative for neutral relative to angry sentences – but not in the schizophrenia group ($p > 0.05$). Within-group comparisons for PPS showed an emotion effect both in healthy controls ($F_{2,13} = 41.922$, $p < 0.001$) and in schizophrenia ($F_{2,13} = 8.615$, $p < 0.01$). In both groups, N100 was less negative for neutral relative to both happy and angry prosody.

P200 amplitude

The omnibus MANOVA revealed a main effect of sentence condition ($F_{1,28} = 5.026$, $p = 0.033$) and emotion ($F_{2,27} = 9.342$, $p < 0.01$). No hemispheric differences were observed ($p > 0.05$). Importantly, emotion interacted with group ($F_{2,27} = 4.775$, $p = 0.017$) and an effect of group approached significance ($F_{1,28} = 3.799$, $p = 0.061$). We followed this interaction with separate MANOVAs for each emotion type. A group × sentence condition interaction was observed for angry prosody ($F_{1,28} = 4.60$, $p = 0.041$); more positive P200 for angry SSC was found in the schizophrenia group relative to healthy controls ($p = 0.015$). Additionally, a group effect was observed for happy prosody ($F_{1,28} = 5.897$,
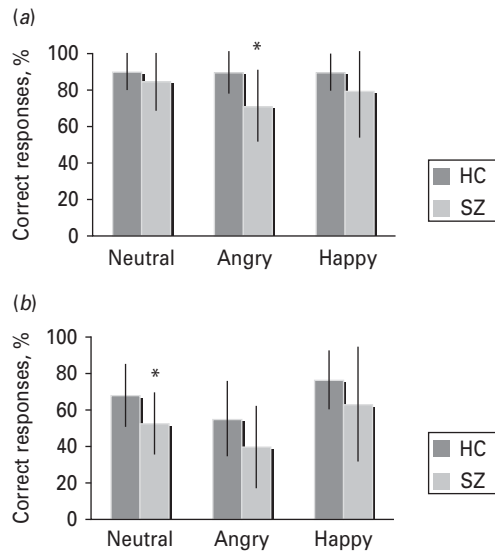
Fig. 3. (*a*) Sentences with semantic content (SSC). Data are percentage of mean correct responses in the behavioral task of emotional prosody discrimination in SSC, with standard deviations represented by vertical bars. (*b*) 'Pure prosody' sentences (PPS). Data are percentage of mean correct responses in the behavioral task of emotional prosody discrimination in PPS, with standard deviations represented by vertical bars. HC, Healthy controls; SZ, schizophrenia patients. * Mean value was significantly different from that of the HC group ($p < 0.05$).

$p = 0.022$); P200 was more positive in the schizophrenia group relative to healthy controls for both sentence conditions.

In within-group comparisons for SSC, an emotion effect was observed both in the healthy controls ($F_{2,13} = 27.406$, $p < 0.001$) and schizophrenia ($F_{2,13} = 5.452$, $p = 0.019$) groups. However, differences were observed in the way both groups extracted emotional salience from the acoustic signal. In healthy controls, P200 was more positive for neutral relative to both happy and angry sentences, and more positive for happy relative to angry sentences. In patients, P200 was more positive to both happy ($p = 0.022$) and neutral ($p = 0.033$) relative to angry prosody in SSC.

Within-group analyses for PPS showed an emotion effect in both groups (healthy controls: $F_{2,13} = 6.997$, $p < 0.01$; schizophrenia group: $F_{2,13} = 13.332$, $p < 0.01$). In spite of a larger P200 for happy PPS in schizophrenia, the profile of emotion effects was similar in both groups; P200 was less positive for neutral relative to both happy and angry prosody.

Behavioral results

The overall MANOVA showed an effect of sentence condition ($F_{1,28} = 137.439$, $p < 0.001$) and emotion

($F_{2,27} = 17.012$, $p < 0.001$); fewer correct responses were observed for PPS relative to SSC, and for angry relative to both happy ($p < 0.001$) and neutral ($p < 0.01$) sentences.

Groups differed in the number of correct responses ($F_{1,28} = 6.746$, $p = 0.015$), with higher error rates observed in the schizophrenia group (24.71, S.D. = 8.49) relative to healthy controls (29.7, S.D. = 5.49). When contrasting directly group performance for each emotion type separately in each sentence condition with one-way ANOVAs, results showed higher error rates in the patients' group for angry sentences in the SSC ($F_{1,28} = 7.515$, $p = 0.011$), and neutral sentences in the PPS condition ($F_{1,28} = 6.367$, $p = 0.018$), as well as a trend for reduced accuracy for angry sentences in the PPS condition ($F_{1,28} = 3.600$, $p = 0.068$) (Fig. 3).

Correlations between ERP and behavior results

Correlations were found between N100 and P200 amplitude and behavioral performance in schizophrenia and in healthy controls. In schizophrenic patients, less negative N100 amplitude in neutral SSC was associated with higher error rates in neutral SSC; less positive P200 amplitude in angry SSC was associated with higher error rates in angry SSC. In healthy controls, less negative N100 amplitude in angry PPS was associated with higher error rates in angry PPS.

Correlations between clinical scales and electrophysiological and behavioral results

Significant correlations were found between delusions and P200 amplitude for happy SSC and PPS. Higher scores on the delusion scale correlated with more positive P200 amplitude for happy prosody both in the SSC and PPS conditions. No correlations were found between ERP data and negative symptomatology, and between behavioral performance and clinical symptomatology ($p > 0.05$). In addition, N100 and P200 amplitudes were not modulated by medication dosage or illness duration ($p > 0.05$) (Fig. 4).

Discussion

Overall results

The results of this study provided insights both into the way prosody is processed in normal populations and into the differences between healthy and schizophrenia individuals. Both behavioral and electrophysiological results suggested that neutral prosody and emotional prosody are processed differently and that the presence of semantic information influences the way prosody is processed. Importantly, group differences spanned all three stages of prosody
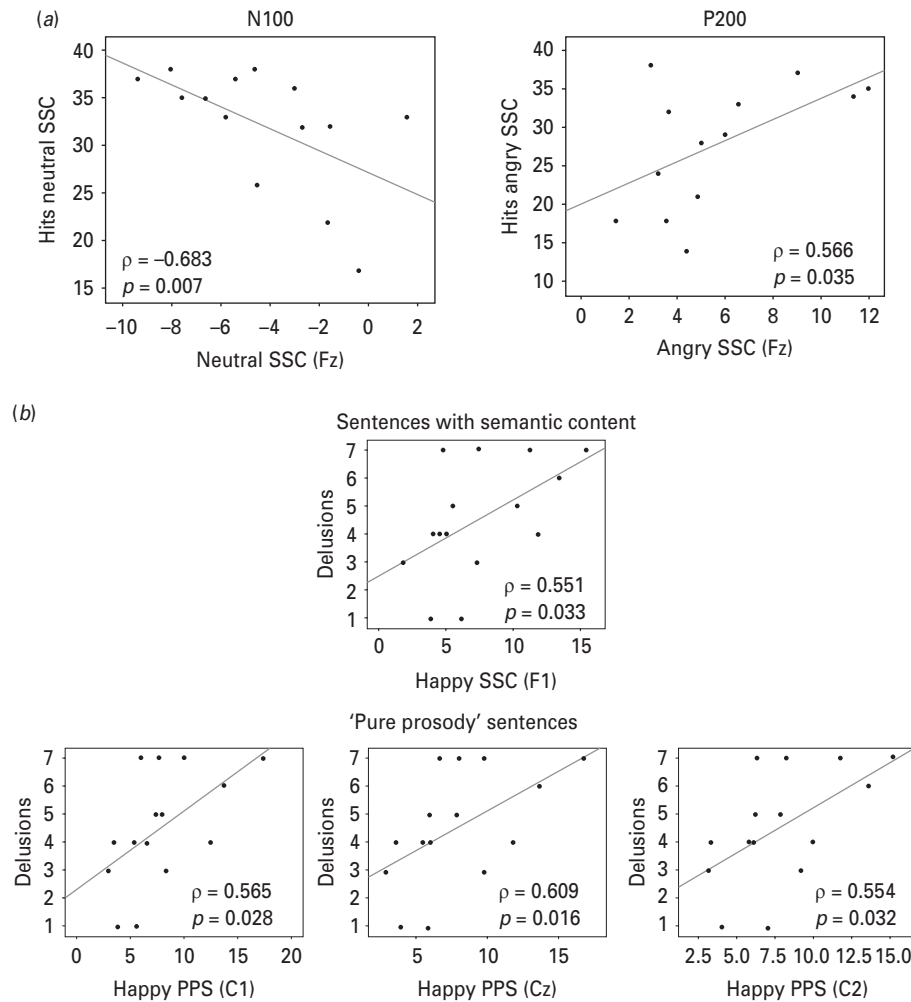
Fig. 4. (*a*) Significant correlations between N100 and P200 amplitude and behavioral results in schizophrenia. (*b*) Significant correlations between P200 amplitude and clinical symptoms (delusions). SSC, Sentences with semantic content; PPS, 'pure prosody' sentences.

processing (Schirmer & Kotz, 2006), and interacted with the semantic status of sentences.

N100: sensory processing of the acoustic signal

The first component found sensitive to both prosodic manipulations and group membership was the N100, suggesting that sensitivity to prosodically relevant features exists already at the N100 level (see also Liu *et al.* 2012). This is the first study to report this result in both normal and schizophrenia groups for prosody rather than emotional sounds processing since N100 in previous ERP studies of prosody processing, though present, was not formally analysed.

A reduced N100 amplitude in SSC was observed in schizophrenia. Reduced N100 reflecting compromised function and structure of the auditory cortex in schizophrenia has been reported in numerous studies (Forces *et al.* 2008; Rosburg *et al.* 2008; Turetsky *et al.* 2009). Here, we report, for the first time, reduced N100

to complex prosodic cues. Importantly, the N100 reduction was not present across the board for all stimuli but was dependent on sentence condition and emotion, with N100 reduced in patients relative to healthy controls specifically for angry SSC. This finding fits with previous evidence demonstrating reduced activation in the limbic system for negative emotional sounds in schizophrenia patients with chronic auditory hallucinations (Kang *et al.* 2009). The low reactivity to anger in normal speech may be related to the experience of hallucinations and/or delusions that often have an angry content and form (e.g. Garety *et al.* 2011). Even though no significant correlation was found between N100 amplitude for angry prosody and clinical symptoms, this relationship should be explored with a larger sample.

In contrast, N100 amplitude in PPS was not different from healthy controls. Note that, in PPS, the lexical–semantic information was absent due to the

systematical distortion of the acoustic signal (see Method section above). Thus, the specificities of this sentence condition might have contributed to the pattern of observed differences.

Additionally, in patients, N100 amplitude did not distinguish between emotional *versus* neutral prosody in SSC, even though more negative N100 to emotional relative to neutral prosody in PPS was observed in patients similarly to healthy controls. The finding of N100 group differences for normal speech supports previous studies indicating sensory contributions to impaired recognition of emotional prosody in schizophrenia, in particular deficits in pitch perception abilities (Leitman *et al.* 2010). However, the PPS findings qualify this conclusion by pointing to a dynamic interplay between bottom-up and top-down processes (e.g. memory predictions about an incoming stimulus) during sensory discrimination (Chandrasekaran *et al.* 2009; Diekhof *et al.* 2009; Krishnan *et al.* 2009; Schadow *et al.* 2009; Kumar *et al.* 2011; Marmel *et al.* 2011). Specifically, they point to the influence of top-down processes on the sensory processing of prosodic cues, suggesting that lexical–semantic information present in a non-distorted acoustic signal may interact with early sensory processes making it more difficult for schizophrenia patients to process normal speech relative to a 'pure prosody' speech signal.

Analysed separately, healthy controls showed more negative N100 amplitude to neutral relative to emotional (i.e. angry) prosody in SSC, and to emotional (both happy and angry) relative to neutral prosody in PPS [the finding of increased N100 for neutral relative to emotional prosody in SSC is similar to the result reported by Liu *et al.* (2012) for non-verbal emotional vocalizations]. Given the role of N100 as an index of sensory processing and evidence suggesting that its amplitude is modulated by stimulus physical properties, the differentiation observed at this early stage is probably related more to differences in the sensory signal characteristics than to valence processing *per se*. Previous studies have shown increased N100 amplitude for higher-intensity auditory stimuli relative to lower-intensity stimuli (Connolly, 1993; Gonsalvez *et al.* 2007). In addition, f_0 was found to modulate N100 amplitude and latency (Stufflebeam *et al.* 1998; Seither-Preisler *et al.* 2006).

P200: deriving emotional significance from acoustic cues

Larger P200 amplitude for angry SSC and for happy SSC and PPS was observed in schizophrenic patients when compared with healthy controls, suggesting abnormal detection of emotional salience in emotional auditory stimuli. The abnormalities found in happy

prosody processing both in SSC and PPS may indicate a specific difficulty in extracting emotional salience from happy stimuli, as proposed previously (deficit in positive emotion perception; Loughland *et al.* 2002).

Of note, a positive relationship was observed between abnormal processing of happy prosody in both SSC and PPS and delusions scores. The observed correlations are consistent with reports of impaired affect perception related to delusions (e.g. Rossell *et al.* 2010). Given that P200 amplitude is modulated by attention (decreased P200 amplitude seems to be related to increased attention; Crowley & Colrain, 2004), the increased P200 amplitude to positive social information in schizophrenia patients with higher delusions scores may represent a decreased capture of attention by happy prosody contrasted with increased attention to perceived threat. A complementary hypothesis is that the extraction of emotional salience from an acoustic signal associated with happy prosody is associated with increased difficulty when compared with angry prosody. This hypothesis is based on more recent studies suggesting an association between P200 amplitude and required cognitive effort (Lenz *et al.* 2007). The observed correlations thus support the hypothesis of a relationship between prosodic deficit and positive symptoms (Poole *et al.* 1997; Rossell & Boundy, 2005; Shea *et al.* 2007) and, in particular, with the role of misattribution of emotional salience in delusions (Holt *et al.* 2006).

Analysed separately, healthy controls showed more positive P200 for neutral relative to emotional prosody in SSC, similarly to Paulmann & Kotz (2008b). This supports previous studies suggesting that, at the P200 level, auditory information is segregated into neutral and emotional prosody, based on salience of acoustic features (Schirmer & Kotz, 2006; Liu *et al.* 2012). However, differently from Paulmann & Kotz (2008b), a valence-specific effect was found: P200 was more positive for happy relative to angry prosody. This result suggests that the processes associated with deriving emotional significance from acoustic cues indexed by P200 distinguished between specific kinds of emotion. Differences in stimuli (e.g. number of emotion types) and task instructions may have contributed to the discrepancy between our findings and those of Paulmann & Kotz (2008b). In the study by Paulmann & Kotz (2008b), participants did not explicitly judge the prosody type, but made decisions on a probe word that followed each prosodic sentence. The difference between the implicit and explicit emotional prosody recognition may have contributed to the different results.

In healthy controls, emotion effects were also observed in PPS: P200 was more positive for emotional relative to neutral prosody. Similar P200 effects were

reported by Sauter & Eimer (2010) and Liu *et al.* (2012) for non-semantic emotional vocalizations. Functionally, P200 is at the interface between sensory and cognitive processing and indexes initial categorization processes. The different profile of categorizing neutral *versus* emotional prosody observed in SSC and PPS corroborates the idea that sensory cues are used differently in both sentence conditions. Also, they support the idea that emotional salience is not derived from a single acoustic cue, but from a specific configuration of acoustic parameters that form an emotional auditory object or *gestalt* (Laukka, 2005; Schirmer & Kotz, 2006; Paulmann & Kotz, 2008*b*). The relative importance of each prosodic parameter is likely to change in the absence of intelligible semantic information.

Analysed separately, schizophrenia patients revealed again a pattern of prosody processing that was different from healthy controls, and depended on the presence of semantic information. In contrast to healthy controls, in SSC, similarly increased P200 amplitude was observed for neutral and happy sentences relative to angry sentences. Studies with healthy subjects suggest that P200 amplitude is modulated by both pitch (lower pitch associated with higher P200 amplitude) and attention (the higher the attention, the lower the P200 amplitude) (for a review, see Crowley & Colrain, 2004). Given the pattern of observed differences, it is likely that the differential engagement of attentional resources by each prosody type contributed to specific P200 effects in each group. Thus, healthy controls may have deployed more attentional resources to both emotional stimuli, resulting in the reduced P200. In contrast, in schizophrenia, reduction in P200 amplitude was selectively observed only for angry sentences which, as mentioned above, may be related to more attentional resources and/or less cognitive effort devoted to angry, relative to both happy and neutral affect in SSC. These data may indicate that the processing of neutral and happy prosody is more difficult in schizophrenia as suggested previously (Loughland *et al.* 2002; Holt *et al.* 2006; Seiferth *et al.* 2008), especially for speech signals carrying both prosodic and semantic information.

Behavioral response: cognitive evaluation of emotional significance

At the third stage of emotional prosody processing, the listener integrates the information provided by the physical properties of the stimuli with the meaning conveyed by linguistic (e.g. semantic) information so that cognitive judgements can be made (Schirmer & Kotz, 2006). As demonstrated by Paulmann & Kotz (2008*b*) and Paulmann *et al.* (2009), ERP data were insensitive to this stage of analysis. However, as suggested by Paulmann *et al.* (2009), behavioral results were informative in this respect. In all subjects, emotional recognition was better in SSC than in PPS. The absence of a memory representation for PPS may have increased task demands and made it more difficult to distinguish between different prosody types (Kotz *et al.* 2003). As suggested by previous studies (Kotz & Paulmann, 2007; Paulmann *et al.* 2009), the increased recognition rates for SSC over PPS may indicate that the availability of the semantic channel is related to an emotional processing advantage and that semantics cannot be ignored even when the task demands are focused on emotional prosody only. In between-group comparisons, schizophrenia individuals committed more errors irrespective of sentence condition or emotional category as reported in other studies (Bozikas *et al.* 2006; Shea *et al.* 2007). Thus, by the time a subject was ready to make a response, abnormalities were evident in processing all types of sentences and emotions.

In schizophrenia, an association was found between less negative N100 amplitude for neutral SSC and higher error rates in neutral prosody SSC recognition. This may suggest that a decrease in the level of attentiveness (indexed by less negative N100) to neutral prosody SSC may be related to higher error rates in neutral prosody (SSC) recognition. Additionally, less positive P200 for angry prosody in SSC processing correlated with higher error rates related to angry prosody recognition in SSC, an emotion for which patients showed higher error rates when compared with happy prosody discrimination. The finding that reduced P200 was associated with higher error rates for angry prosody recognition in SSC suggests that the ability to disengage attention from negative stimuli (that would be indexed by increased P200 amplitude) may, in fact, result in a better performance (Green *et al.* 2003).

These associations suggest that abnormalities at early processing stages are indeed related to later decision stages, thus supporting the contribution of sensory abnormalities to neutral and angry prosody recognition in schizophrenia.

Together, these findings point to group differences in the integration of information derived from early perceptual analyses (N100 and P200) with the conceptual knowledge of emotions that is important for cognitive evaluation of perceptual analyses. Thus, abnormalities in the initial processes of extracting acoustic information and assigning emotional salience to an utterance might set the stage for difficulties in assigning emotional meaning and integrating it with semantic content leading to errors in emotion recognition. Importantly, our data suggest reciprocal

relationships between higher-order processes and sensory-based operations in bringing about prosody dysfunction in schizophrenia.

Limitations and future directions

The limitations include a chronic, medicated schizophrenia sample, raising the issue of a potential medication effect. However, the correlational analyses between both medication dosage and illness duration, and ERP and behavioral responses were not significant. Also, deficits in the processing of emotional prosody were observed in first-episode patients (Haskins *et al.* 1995; Edwards *et al.* 2001), in children with schizophrenia (Baltaxe & Simmons, 1995), and tend to occur independent of medication (Kerr & Neale, 1993; Ross *et al.* 2001), arguing against the possibility that medication, duration of illness or institutionalization contributed to the results.

There is some evidence suggesting that deficits in emotional prosody perception tend to be dependent on schizophrenia subtype (Shea *et al.* 2007). Future studies should address emotional prosody processing in different schizophrenia subgroups.

Only males were studied in the current investigation and some studies suggest that emotional perception deficits are worse in male than in female patients (Bozikas *et al.* 2006; Scholten *et al.* 2008). Therefore, future studies should include women.

Conclusions

Our findings, for the first time, elucidate the interaction between sensory and higher-order processes in bringing about prosody processing abnormality in schizophrenia using ERP methodology. This is the first study to provide electrophysiological evidence for emotional prosody processing abnormalities.

These findings have important implications for understanding social interactions in schizophrenia, since difficulties in decoding emotions from prosodic cues may contribute to difficulties in social reciprocity (Brekke *et al.* 2005) and to poor outcome (Green *et al.* 2000; Leitman *et al.* 2011).
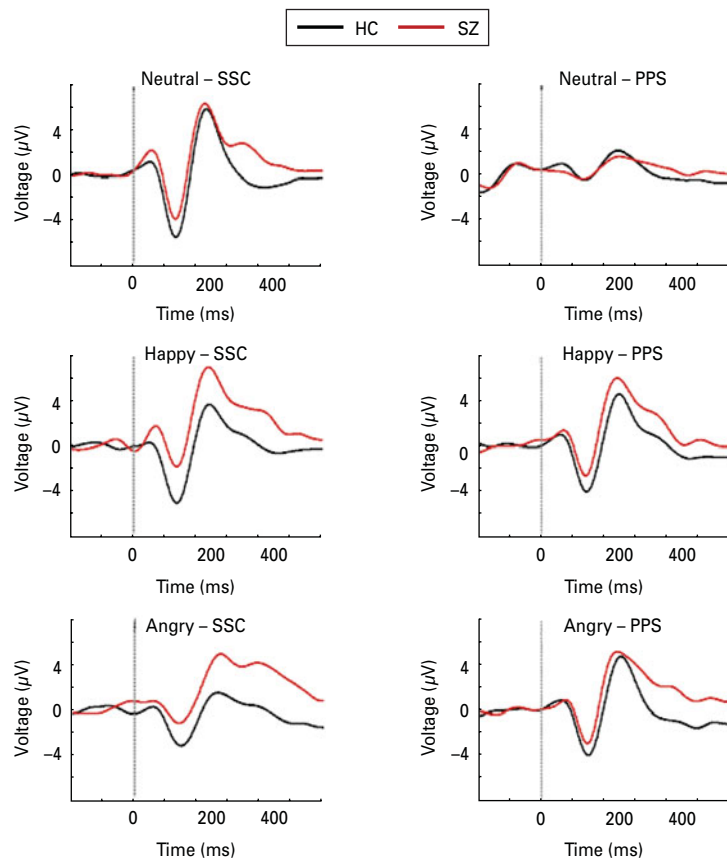
Appendix

Fig. A1. Grand average waveforms for each prosody type (neutral, happy and angry) in both sentence conditions [sentences with semantic content (SSC) and 'pure prosody' sentences PPS)] in healthy controls (HC) and schizophrenia (SZ) patients time-locked to the onset of the major prosodic shift (i.e. the onset of the sentence), with the epoch spanning 600 ms.

Acknowledgements

Declaration of Interest

None.

References

American Electroencephalographic Society (1991). Guidelines for standard electrode position nomenclature. *Journal of Clinical Neurophysiology* **8**, 200–202.

APA (2000). *Diagnostic and Statistical Manual of Mental Disorders*, 4th edn, revised. APA: Washington, DC.

Baltaxe CA, Simmons III JQ (1995). Speech and language disorders in children and adolescents with schizophrenia. *Schizophrenia Bulletin* **21**, 677–692.

Banse R, Scherer KR (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology* **70**, 614–636.

Bozikas V, Kosmidis M, Anezoulaki D, Giannakou M, Andreou C, Karavatos A (2006). Impaired perception of affective prosody in schizophrenia. *Journal of Neuropsychiatry and Clinical Neurosciences* **18**, 81–85.

Brekke J, Kay DD, Lee KS, Green MF (2005). Biosocial pathways to functional outcome in schizophrenia. *Schizophrenia Research* **80**, 213–225.

Buchanan TW, Lutz K, Mirzazade S, Specht K, Shah NJ, Zilles K, Jancke L (2000). Recognition of emotional prosody and verbal components of spoken language: an fMRI study. *Cognitive Brain Research* **9**, 227–238.

Chandrasekaran B, Krishnan A, Gandour JT (2009). Sensory processing of linguistic pitch as reflected by the mismatch negativity. *Ear and Hearing* **30**, 552–558.

Connolly JF (1993). The influence of stimulus intensity, contralateral masking and handedness on the temporal N1 and the T complex components of the auditory N1 wave. *Electroencephalography and Clinical Neurophysiology* **86**, 58–68.

Crowley KE, Colrain IM (2004). A review of the evidence for P2 being an independent component process: age, sleep and modality. *Clinical Neurophysiology* **115**, 732–744.

Diekhof EK, Biedermann F, Ruebsamen R, Gruber O (2009). Top-down and bottom-up modulation of brain structures involved in auditory discrimination. *Brain Research* **1297**, 118–123.

Dutoit T, Pagel V, Pierret N, Bataille F, Van Der Vreken O (1996). The MBROLA Project: towards a set of high-quality speech synthesizers free of use for non-commercial purposes. *Proceedings of ICSLP '96* **3**, 1393–1396.

Edwards J, Pattison PE, Jackson HJ, Wales RJ (2001). Facial affect and affective prosody recognition in first-episode schizophrenia. *Schizophrenia Research* **48**, 235–253.

Ethofer T, Anders S, Erb M, Herbert C, Wiethoff S, Kissler J, Grodd W, Wildgruber D (2006). Cerebral pathways in processing of affective prosody: a dynamic causal modeling study. *Neuroimage* **30**, 580–587.

First MB, Spitzer RL, Gibbon M, Williams JBW (1995). *Structured Clinical Interview for DSM-IV Axis II Personality Disorders* (*SCID-II, version 2.0*). Biometrics Research Department, New York State Psychiatric Institute: New York.

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*). Biometric Research Department, New York State Psychiatric Institute: New York.

Forces RB, Venables NC, Sponheim SR (2008). An auditory processing abnormality specific to liability for schizophrenia. *Schizophrenia Research* **103**, 298–310.

Gandour J, Wong D, Dzemidzic M, Lowe M, Tong Y, Li X (2003). A cross-linguistic fMRI study of perception of intonation and emotion in Chinese. *Human Brain Mapping* **18**, 149–157.

Garety PA, Kuipers E, Fowler D, Freeman D, Bebbington PE (2001). A cognitive model of the positive symptoms of psychosis. *Psychological Medicine* **31**, 189–195.

Gonsalvez CJ, Barry RJ, Rushby JA, Polich J (2007). Target-to-target interval, intensity, and P300 from an auditory single-stimulus task. *Psychophysiology* **44**, 245–250.

Grandjean D, Sander D, Pourtois G, Schwartz S, Seghier ML, Scherer KR, Vuilleumier P (2005). The voices of wrath: brain responses to angry prosody in meaningless speech. *Nature Neuroscience* **8**, 145–146.

Green MF, Kern RS, Braff DL, Mintz J (2000). Neurocognitive deficits and functional outcome in schizophrenia: are we measuring the "right stuff"? *Schizophrenia Bulletin* **26**, 119–136.

Green MJ, Williams LM, Davidson D (2003). Visual scanpaths to threat-related faces in deluded schizophrenia. *Psychiatry Research* **3**, 271–285.

Hart HC, Hall DA, Palmer AR (2003). The sound-level-dependent growth in the extent of fMRI activation in Heschl's gyrus is different for low- and high-frequency tones. *Hearing Research* **179**, 104–112.

Haskins B, Shutty MS, Kellogg E (1995). Affect processing in chronically psychotic patients: development of a reliable assessment tool. *Schizophrenia Research* **15**, 291–297.

Hillyard SA, Hink RF, Schwent VL, Picton TW (1973). Electrical signs of selective attention in the human brain. *Science* **182**, 177–180.

Hoekert M, Kahn RS, Pijnenborg M, Aleman A (2007). Impaired recognition and expression of emotional prosody

in schizophrenia: review and meta-analysis. *Schizophrenia Research* **96**, 135–145.

Hollingshead AB (1965). *Two-Factor Index of Social Position*. Yale Station: New Haven, CT.

Holt DJ, Titone D, Long LS, Goff DC, Cather C, Rauch SL, Judge A, Kuperberg GR (2006). The misattribution of salience in delusional patients with schizophrenia. *Schizophrenia Research* **83**, 247–256.

Kang JI, Kim J, Seok J, Chun JW, Lee S, Park H (2009). Abnormal brain response during the auditory emotional processing in schizophrenic patients with chronic auditory hallucinations. *Schizophrenia Research* **107**, 83–91.

Kay SR, Fiszbein A, Opler LA (1987). The Positive and Negative Syndrome Scale (PANSS). *Schizophrenia Bulletin* **13**, 261–276.

Kerr SL, Neale JM (1993). Emotion perception in schizophrenia: specific deficit or further evidence of generalized poor performance? *Journal of Abnormal Psychology* **102**, 312–318.

Kotz SA, Meyer M, Alter K, Besson M, von Cramon DY, Friederici AD (2003). On the lateralization of emotional prosody: an event-related functional MR investigation. *Brain and Language* **86**, 366–376.

Kotz SA, Paulmann S (2007). When emotional prosody and semantics dance cheek to cheek: ERP evidence. *Brain Research* **1151**, 107–118.

Krishnan RR, Keefe R, Kraus M (2009). Schizophrenia is a disorder of higher order hierarchical processing. *Medical Hypotheses* **6**, 740–744.

Kucharska-Pietura K, David A, Masiak K, Phillips M (2005). Perception of facial and vocal affect by people with schizophrenia in early and late stages of illness. *British Journal of Psychiatry* **187**, 523–528.

Kumar S, Sedley W, Nourski KV, Kawasaki H, Oya H, Patterson RD, Howard 3rd MA, Friston KJ, Griffiths TD (2011). Predictive coding and pitch processing in the auditory cortex. *Journal of Cognitive Neuroscience* **23**, 3084–3094.

Laukka P (2005). Categorical perception of vocal emotion expressions. *Emotion* **5**, 277–295.

Leitman DI, Foxe JJ, Butler PD, Saperstein A, Revheim N, Javitt DC (2005). Sensory contributions to impaired prosodic processing in schizophrenia. *Biological Psychiatry* **58**, 56–61.

Leitman DI, Laukka P, Juslin PN, Saccente E, Butler P, Javitt DC (2010). Getting the cue: sensory contributions to auditory emotion recognition impairments in schizophrenia. *Schizophrenia Bulletin* **36**, 545–556.

Leitman DI, Wolf DH, Laukka P, Ragland JD, Valdez JN, Turetsky BI, Gur RE, Gur RC (2011). Not pitch perfect: sensory contributions to affective communication impairment in schizophrenia. *Biological Psychiatry* **70**, 611–618.

Lenz D, Schadow J, Thaerig S, Busch NA, Herrmann CS (2007). What's that sound? Matches with auditory long-term memory induce gamma activity in human EEG. *International Journal of Psychophysiology* **64**, 31–38.

Liu T, Pinheiro AP, Guanghui D, Nestor PG, McCarley RW, Niznikiewicz M (2012). Electrophysiological insights into processing nonverbal emotional vocalizations. *Neuroreport* **23**, 108–112.

Loughland CM, Williams LM, Gordon E (2002). Visual scanpaths to positive and negative facial emotions in an outpatient schizophrenia sample. *Schizophrenia Research* **55**, 159–170.

Marmel F, Perrin F, Tillmann B (2011). Tonal expectations influence early pitch processing. *Journal of Cognitive Neuroscience* **23**, 3095–3104.

Mitchell RL, Elliott R, Barry M, Cruttenden A, Woodruff PW (2003). The neural response to emotional prosody, as revealed by functional magnetic resonance imaging. *Neuropsychologia* **41**, 1410–1421.

Niznikiewicz M, Mittal MS, Nestor PG, McCarley RW (2010). Abnormal inhibitory processes in semantic networks in schizophrenia. *International Journal of Psychophysiology* **75**, 133–140.

Oldfield RC (1971). The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia* **9**, 97–113.

Paulmann S, Kotz SA (2008*a*). An ERP investigation on the temporal dynamics of emotional prosody and emotional semantics in pseudo- and lexical-sentence context. *Brain and Language* **105**, 59–69.

Paulmann S, Kotz SA (2008*b*). Early emotional prosody perception based on different speaker voices. *Neuroreport* **19**, 209–213.

Paulmann S, Seifert S, Kotz SA (2009). Orbito-frontal lesions cause impairment during late but not early emotional prosodic processing. *Social Neuroscience* **5**, 59–75.

Pinheiro AP, Galdo-Alvarez S, Rauber A, Sampaio A, Niznikiewicz M, Goncalves OF (2011). Abnormal processing of emotional prosody in Williams syndrome: an event-related potentials study. *Research in Developmental Disabilities* **32**, 133–147.

Poole JH, Tobias FC, Vinogradov S (2000). The functional relevance of affect recognition errors in schizophrenia. *Journal of the International Neuropsychological Society* **6**, 649–658.

Ramus F, Mehler J (1999). Language identification with suprasegmental cues: a study based on speech resynthesis. *Journal of the Acoustical Society of America* **105**, 512–521.

Rauschecker JP (1997). Processing of complex sounds in the auditory cortex of cat, monkey, and man. *Acta Otolaryngological Supplementum* **532**, 34–38.

Rosburg T, Boutros NN, Ford JM (2008). Reduced auditory evoked potential component N100 in schizophrenia – a critical review. *Psychiatry Research* **161**, 259–274.

Ross ED, Orbelo DM, Cartwright J, Hansel S, Burgard M, Testa JA, Buck R (2001). Affective-prosodic deficits in schizophrenia: comparison to patients with brain damage and relation to schizophrenic symptoms. *Journal of Neurology, Neurosurgery and Psychiatry* **70**, 597–604.

Rossell SL, Batty RA, Hughes L (2010). Impaired semantic memory in the formation and maintenance of delusions post-traumatic brain injury: a new cognitive model. *European Archives of Psychiatry and Clinical Neuroscience* **260**, 571–581.

Rossell SL, Boundy CL (2005). Are auditory-verbal hallucinations associated with auditory affective processing deficits? *Schizophrenia Research* **78**, 95–106.

Sauter SA, Eimer M (2010). Rapid detection of emotion from human vocalizations. *Journal of Cognitive Neuroscience* **22**, 474–481.

Schadow J, Lenz D, Dettler N, Frund I, Hermann CS (2009). Early gamma-band responses reflect anticipatory top-down modulation in the auditory cortex. *Neuroimage* **47**, 651–658.

Scherer KR (1995). Expression of emotion in voice and music. *Journal of Voice* **9**, 235–248.

Schirmer A, Kotz SA (2006). Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences* **10**, 24–30.

Scholten MR, Aleman A, Kahn RS (2008). The processing of emotional prosody and semantics in schizophrenia: relationship to gender and IQ. *Psychological Medicine* **38**, 887–898.

Seiferth NY, Pauly K, Habel U, Kellermann T, Shah NJ, Ruhrmann S, Klosterkotter J, Schneider F, Kircher T (2008). Increased neural response related to neutral faces in individuals at risk for psychosis. *Neuroimage* **40**, 289–297.

Seither-Preisler A, Patterson R, Krumbholz K, Seither S, Lutkenhoner B (2006). Evidence of pitch processing in the N100m component of the auditory evoked field. *Hearing Research* **213**, 88–98.

Shaw RJ, Dong M, Lim KO, Faustman WO, Pouget ER, Alpert M (1999). The relationship between affect expression and affect recognition in schizophrenia. *Schizophrenia Research* **37**, 245–250.

Shea TL, Sergejew AA, Burnham D, Jones C, Rossell SL, Copolov DL, Egan GF (2007). Emotional prosodic processing in auditory hallucinations. *Schizophrenia Research* **90**, 214–220.

Shenton ME, Dickey CC, Frumin M, McCarley RW (2001). A review of MRI findings in schizophrenia. *Schizophrenia Research* **49**, 1–52.

Stufflebeam SM, Poeppel D, Rowley HA, Roberts TPL (1998). Peri-threshold encoding of stimulus frequency and intensity in the M100 latency. *NeuroReport* **9**, 91–94.

Turetsky BI, Bilker WB, Siegel SJ, Kohler CG, Gur RE (2009). Profile of auditory information-processing deficits in schizophrenia. *Psychiatry Research* **165**, 27–37.

Wechsler D (1997). *Wechsler Adult Intelligence Scale: Administration and Scoring Manual*, 3rd edn.. The Psychological Corporation: San Antonio, TX.

Wible CG, Preus AP, Hashimoto R (2009). A cognitive neuroscience view of schizophrenic symptoms: abnormal activation of a system for social perception and communication. *Brain Imaging and Behavior* **3**, 85–110.

Wildgruber D, Ackermann H, Kreifelts B, Ethofer T (2006). Cerebral processing of linguistic and emotional prosody: fMRI studies. *Progress in Brain Research* **156**, 249–268.

Wildgruber D, Riecker A, Hertrich I, Erb M, Grodd W, Ethofer T, Ackermann H (2005). Identification of emotional intonation evaluated by fMRI. *Neuroimage* **24**, 1233–1241.