

The integration of facial and vocal cues during emotional change perception: EEG markers

Xuhai Chen,^{1,2} Zhihui Pan,¹ Ping Wang,¹ Xiaohong Yang,³ Peng Liu,¹ Xuqun You,¹ and Jiajin Yuan⁴

¹Key Laboratory of Behavior and Cognitive Psychology in Shaanxi Province, School of Psychology, Shaanxi Normal University, Xi'an 710062, China, ²Key Laboratory of Modern Teaching Technology, Ministry of Education, Shaanxi Normal University, Xi'an 710062, China, ³Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China, and ⁴Key Laboratory of Cognition and Personality of Ministry of Education, School of Psychology, Southwest University, Chongqing 400715, China

Correspondence should be addressed to Jiajin Yuan, Faculty of Psychology, Southwest University, No.2 of Tiansheng Road, Beibei, Chongqing, 400715, China. E-mail: yuanjiajin168@126.com or yuanjiaj@swu.edu.cn.

Abstract

The ability to detect emotional changes is of primary importance for social living. Though emotional signals are often conveyed by multiple modalities, how emotional changes in vocal and facial modalities integrate into a unified percept has yet to be directly investigated. To address this issue, we asked participants to detect emotional changes delivered by facial, vocal and facial-vocal expressions while behavioral responses and electroencephalogram were recorded. Behavioral results showed that bimodal emotional changes were detected with higher accuracy and shorter response latencies compared with each unimodal condition. Moreover, the detection of emotional change, regardless of modalities, was associated with enhanced amplitudes in the N2 and P3 component, as well as greater theta synchronization. More importantly, the P3 amplitudes and theta synchronization were larger for the bimodal emotional change condition than for the sum of the two unimodal conditions. The superadditive responses in P3 amplitudes and theta synchronization were both positively correlated with the magnitude of the bimodal superadditivity in accuracy. These behavioral and electrophysiological data consistently illustrated an effect of audiovisual integration during the detection of emotional changes, which is most likely mediated by the P3 activity and theta oscillations in brain responses.

Key words: emotional change; bimodal integration; event related potential; event related spectral perturbation

Introduction

In a negotiation, it is often the case that the emotion expressions of the negotiators shift rapidly, from calm to frowns due to abrupt demands, or from frowns to smiles due to mutual concessions. This phenomenon tells that the ability to identify changes in emotion expressions is of primary importance for social living (Fujimura and Okanoya, 2013): not just for understanding interlocutors' mental states (Sabbagh et al., 2004), but also for avoiding interpersonal conflicts and constructing interpersonal harmony (Salovey and Mayer, 1989; Van Kleef, 2009). Therefore, how the brain processes emotional changes,

particularly those from facial and vocal expressions, has attracted great interest of researchers in affective neuroscience.

The processing of emotional changes in a single modality, either facial or vocal, has been widely investigated. Ample evidence suggests that inattentive detection of emotional changes in facial (Kimura et al., 2012; Stefanics et al., 2012; Xu et al., 2013) or vocal (Thonnessen et al., 2010) expressions is associated with mismatch negativity (MMN) in passive oddball paradigm, during which the deviant expressions elicit emotional changes by violating the contextual expressions (Winkler et al., 2009). Different from inattentive processing that uses a non-emotional, distracting task,

Received: 20 March 2015; Revised: 2 June 2015; Accepted: 24 June 2015

© The Author (2015). Published by Oxford University Press. For Permissions, please email: journals.permissions@oup.com

attentive processing requires subjects to detect emotional changes voluntarily (Campanella et al., 2002; Wambacq and Jerger, 2004). This line of studies revealed that the detection of deviant facial expression is associated with an occipito-frontal N2/P3a complex (Campanella et al., 2002, 2013). Similar findings were also observed in vocal emotional studies, which reported that attentive processing of prosodic emotional changes is associated with a P3a (Wambacq and Jerger, 2004). Furthermore, vocal emotional changes in the cross-splicing paradigm are associated with a positive deflection 350 ms post splicing point (Kotz and Paulmann, 2007; Chen et al., 2011; Paulmann et al., 2012) and theta synchronization (Chen et al., 2012, 2015), signaling a conscious mobilization of cerebral and somatic resources to evaluate and cope with emotional changes.

Nevertheless, the extant studies only examined the perception of emotional changes in single modality, leaving it unanswered whether and how bimodal emotional changes integrate into a unified emotional percept. However, effective recognition of emotional changes is highly important for social interaction and biological survival (Campanella et al., 2002; Kimura et al., 2012; Stefanics et al., 2012), and emotional cues in life settings are often delivered by multiple expressions (Campanella and Belin, 2007; Klasen et al., 2012). In this regard, it is necessary to examine cross-modal integration of emotional change perception and its brain mechanisms.

Ample evidences show that the processing efficiency for bimodally presented emotional signals is higher than that for unimodally presented emotional signals (Massaro and Egan, 1996; Pourtois et al., 2005; Ethofer et al., 2006; Kreifelts et al., 2007; Jessen, et al., 2012). Consistent with this behavioral facilitation effect, some studies reported reduced N1 (Jessen and Kotz, 2011; Jessen et al., 2012; Kokinous et al., 2014) and P2 (Paulmann et al., 2009; Kokinous et al., 2014) amplitudes during multimodal compared with unimodal emotions. More importantly, it has been found that the event related potential (ERP) components elicited by bimodal emotional stimulations, such as P1, N1 or P3b, are better than those elicited by unimodal in discerning affective disorders, like alexithymia (Delle-Vigne et al., 2014) and anxious/depressive comorbidity (Campanella et al., 2010). Additionally, time-frequency analyses yielded that multimodal relative to unimodal emotions induced a stronger alpha desynchronization associated with stronger attentional involvement (Chen et al., 2010; Jessen and Kotz, 2011), and beta desynchronization (Jessen et al., 2011, 2012) linked with deviant processing (Kim and Chung, 2008; Chen et al., 2015) or action planning (Tzagarakis et al., 2010).

On the basis of these evidences, we hypothesize that there may be a similar facilitation effect in bimodal emotional change perception. However, a facilitation effect alone is not sufficient to verify the occurrence of bimodal emotional integration. In fact, prior studies have indicated that the most effective way to quantify multisensory integration is to compare the bimodal responses with the summed unimodal responses, which results in either sub- or supra-additive response (Calvert, 2001; Stekelenburg and Vroomen, 2007; Hagan et al., 2009; Stein et al., 2009). Using this criterion, Hagan et al. (2009) observed a significant supra-additive gamma oscillation for congruent audiovisual emotional perception. Likewise, Jessen and colleagues (2011) found a stronger beta suppression for the multimodal than the summed unimodal conditions during emotion perception. Following these studies, the current study also used super-additive response as a key criterion to identify cross-modal integration of emotional change perception.

In summary, the current study addresses neural underpinnings of the audiovisual integration of emotional change

perception by focusing on the following two questions: (i) whether the classic electrophysiological markers of unimodal emotional change perception are identifiable in bimodal emotional change perception; (ii) whether there is a significant cross-modal integration during emotional change perception, for example, by observing superadditive effects in these markers. To address these questions, we asked participants to detect emotional changes embedded in concurrently presented facial and vocal expressions while behavioral responses and electroencephalogram (EEG) were recorded. We conducted ERP analysis and time-frequency analysis on brain activities to seek converging evidences. Compared with ERP analysis, which is adept in depicting the time course of cognitive processing but restricted to phase-locked averaging (Luck, 2005), time-frequency analysis is able to provide non-phase-locked indexes of cognition, such as event-related spectral perturbation (ERSP; Makeig et al., 2004). ERSP is a temporally sensitive index of the relative increase or decrease in mean EEG power from baseline associated with stimulus presentation or response execution, termed as event-related synchronization and desynchronization (Delorme and Makeig, 2004; Makeig et al., 2004).

Based on the findings that N2/P3 complex, theta synchronization and beta desynchronization were associated with unimodal emotional change (Campanella et al., 2002; Chen et al., 2011, 2015), we predict that similar electrophysiological correlates should be observed in bimodal emotional change. Additionally, given that bimodal cues lead to a facilitation effect and a cross-modal integration during emotion decoding (Hagan et al., 2009; Jessen and Kotz, 2011; Jessen et al., 2012; Klasen et al., 2012; Watson et al., 2014), we hypothesize that bimodal emotional changes should be detected with higher accuracy rates and shorter response latencies compared with unimodal changes, and the electrophysiological correlates associated with emotional changes should demonstrate a salient superadditive effect for bimodal emotional cues.

Materials and method

Participants

Twenty-two right-handed university students (13 women, aged 19–24, mean 21.56) were recruited to participate in the experiment. Four participants (2 women) were excluded from data analysis because of low accuracy rates. All participants reported normal auditory and normal or corrected-to-normal visual acuity, and were free of neurological or psychiatric problems. They were emotionally stable, as evidenced by the significant below-threshold (0) scoring in neuroticism assessment [$M = -25.83$; $SE = 5.65$; $t(17) = -5.65$, $P < .001$]. The experimental procedure adhered to the Declaration of Helsinki was approved by the Local Review Board for Human Participant Research. All participants gave informed consent prior to the study and were reimbursed ¥50 for their time.

Materials

The original facial stimuli were neutral and angry facial expressions of five male actors from the Ekman and Friesen Affect Series (Young et al., 2002). The original vocal stimuli were 50 sentences of neutral content produced by a trained male actor of Mandarin Chinese in neutral and angry prosodies, which have been validated in previous studies (Chen et al. 2011, 2012). Following the previous studies (Kotz and Paulmann 2007; Chen et al. 2011), sentences with vocal emotional change were obtained by cross-splicing the first part of a neutral prosody and

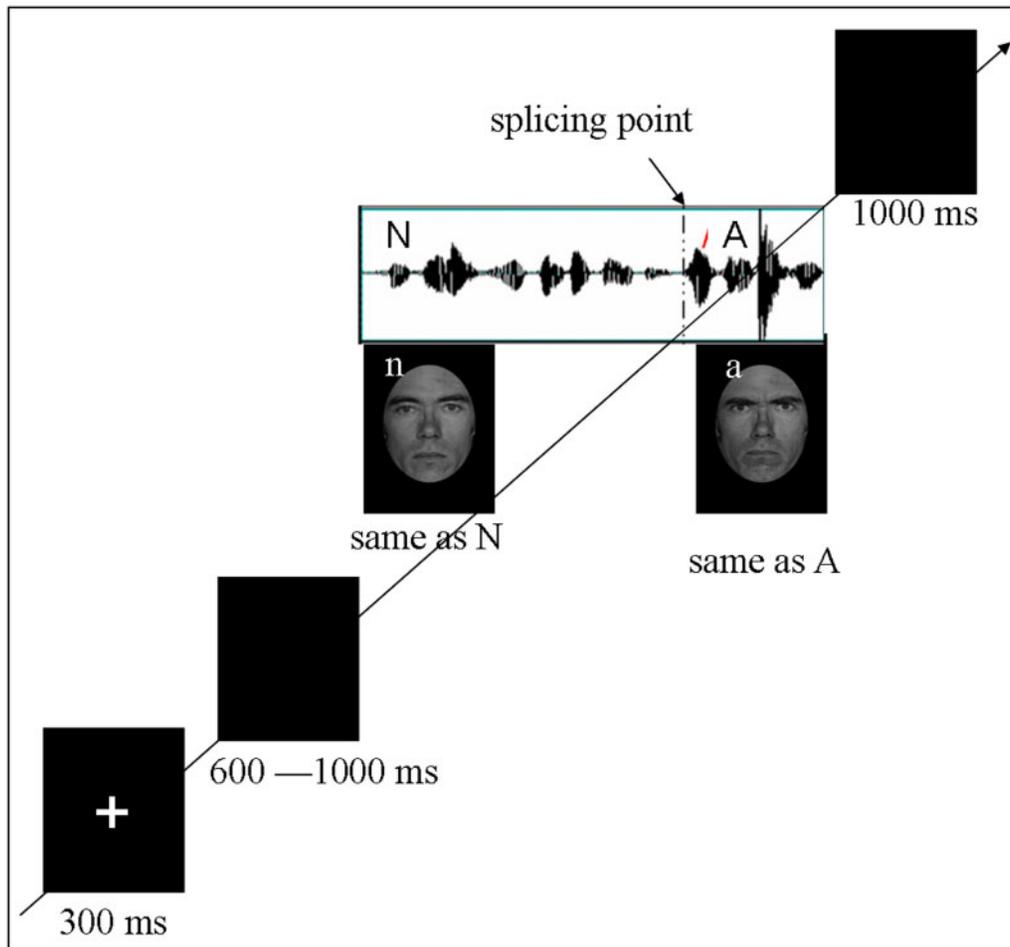


Fig. 1. Schematic representation of stimuli. Trials consisted of fixation, blank jitter, stimuli presentation and decision. All the auditory stimuli consisted with 12 syllables with each syllable lasts about 200 ms. The facial stimuli were presented with same duration as the auditory stimuli. 'N' and 'A' represents neural and angry vocal emotion respectively, while 'n' and 'a' represents neural and angry facial emotion respectively. By combinations of the two modalities, for types of stimuli (no-change, facial-change, vocal-change and bimodal-change) were established

the second part of an angry prosody (see Figure 1 for illustration). Then, each auditory sentence was paired with facial expressions in Latin square design, ensuring that each face of one actor paired with each sentence once. Consistent with the auditory sentences, the facial expressions varied at the splicing point from neutral to angry while the identity kept unchanged (see Figure 1 for illustration). Four types of facial-vocal stimuli, that is, no change, vocal-change, facial-change and bimodal-change, were established.

Procedure

Each participant was tested individually in a sound-attenuated room, seated comfortably at a distance of 100 cm from a computer monitor. Facial-vocal stimulus presentation was controlled using E-prime software. The Auditory stimuli were presented via loudspeakers placed at both sides of the monitor while the facial stimuli were presented on the monitor simultaneously at a central location against a black background (visual angle of 3.4° [width] \times 5.7° [height] from a viewing distance of 100 cm). The hair removed faces were presented in grayscale to minimize contrast differences between stimuli. All 200 key stimuli (each vocal sentence repeated 4 times while each face repeated 40 times) were pseudo-randomized and split up into 5 blocks containing 40

sentences each. In addition, 20 angry sentences without emotional change were included as filler in each block to balance the response type. After each block, participants were given a short self-paced break to rest. Each trial was initiated by a 300 ms presentation of a small white cross on the black computer screen. Then, a blank screen whose duration varied randomly between 600 and 1000 ms was presented and was then followed by the onset of facial-vocal stimuli. Participants were required to judge whether the emotion conveyed by the stimuli (regardless of vocal and facial modality) had changed or not by pressing the 'J' or 'F' button on the keyboard (as accurately and quickly as possible) with index fingers of both hands. Half of the participants pressed the 'J' button for judging emotion changed, and the 'F' button for judging emotion unchanged. The other half followed a reversed button assignment. Participants could respond upon the onset of the change and had to respond within the presentation of facial-vocal stimuli or 1000 ms after the stimulus elapsed (see Figure 1 for a schematic illustration). Pre-training with 30 practice trials was included in order to familiarize subjects with the procedure.

EEG recording

EEG was recorded at 64 scalp sites using tin electrodes mounted in an elastic cap (Brain Product, Munich, Germany)

according to the modified expanded 10–20 system, each referenced on-line to FCZ. Vertical electrooculogram (EOG) was recorded supra-orbitally and infra-orbitally from the right eye. The horizontal EOG was recorded as the left vs right orbital rim. The EEG and EOG were amplified using a 0.05–100 Hz bandpass and continuously digitized at 1000 Hz for offline analysis. The impedance of all electrodes was kept less than 5 k Ω .

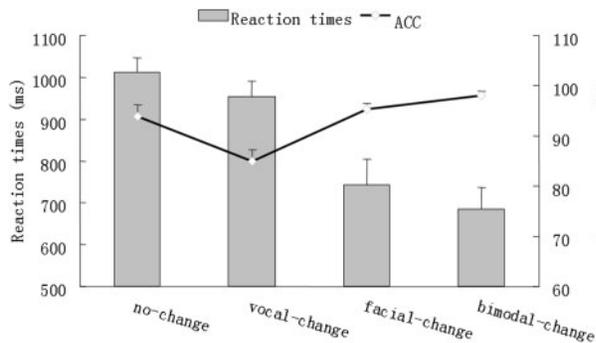


Fig. 2. A schematic illustration of behavioral performance. The bar chart reflects the reaction times while the line chart reflects the accuracy rates. Error bars indicate standard error

Data analysis

Preprocessing. Behavioral data. Accuracy (ACC) was computed by percentage, which we entered into as decimal for statistical convenience. Response times (RTs) beyond three standard deviations and with wrong response were excluded in RT calculation. The descriptive results of ACC and RTs were depicted in Figure 2.

ERP. EEG data were preprocessed using EEGLAB (Delorme and Makeig, 2004), an open source toolbox running under the MATLAB environment. First, the data were down sampled at 250 Hz, high pass filtered at 0.5 Hz and re-referenced offline to bilateral mastoid electrodes. Then the data were segmented in 3000 ms time window with pre-splicing-point 1000 ms. The epoched data were baseline corrected using the 1000–800 ms prior to the pre-splicing-point (200 ms before the sentence onset). EEG epochs with large artifacts (exceeding $\pm 100 \mu\text{V}$) and wrong response were removed and channels with poor signal quality were interpolated using EEGLAB toolbox. Trials contaminated by eye blinks and other artifacts were corrected using an independent component analysis algorithm (Delorme and Makeig, 2004). There were 46.83 ± 1.81 , 46.22 ± 2.26 , 42.28 ± 5.21 and 46.33 ± 1.75 artifact-free trials obtained for the no-change, facial-change, vocal-change and bimodal-change conditions respectively. The artifact free data

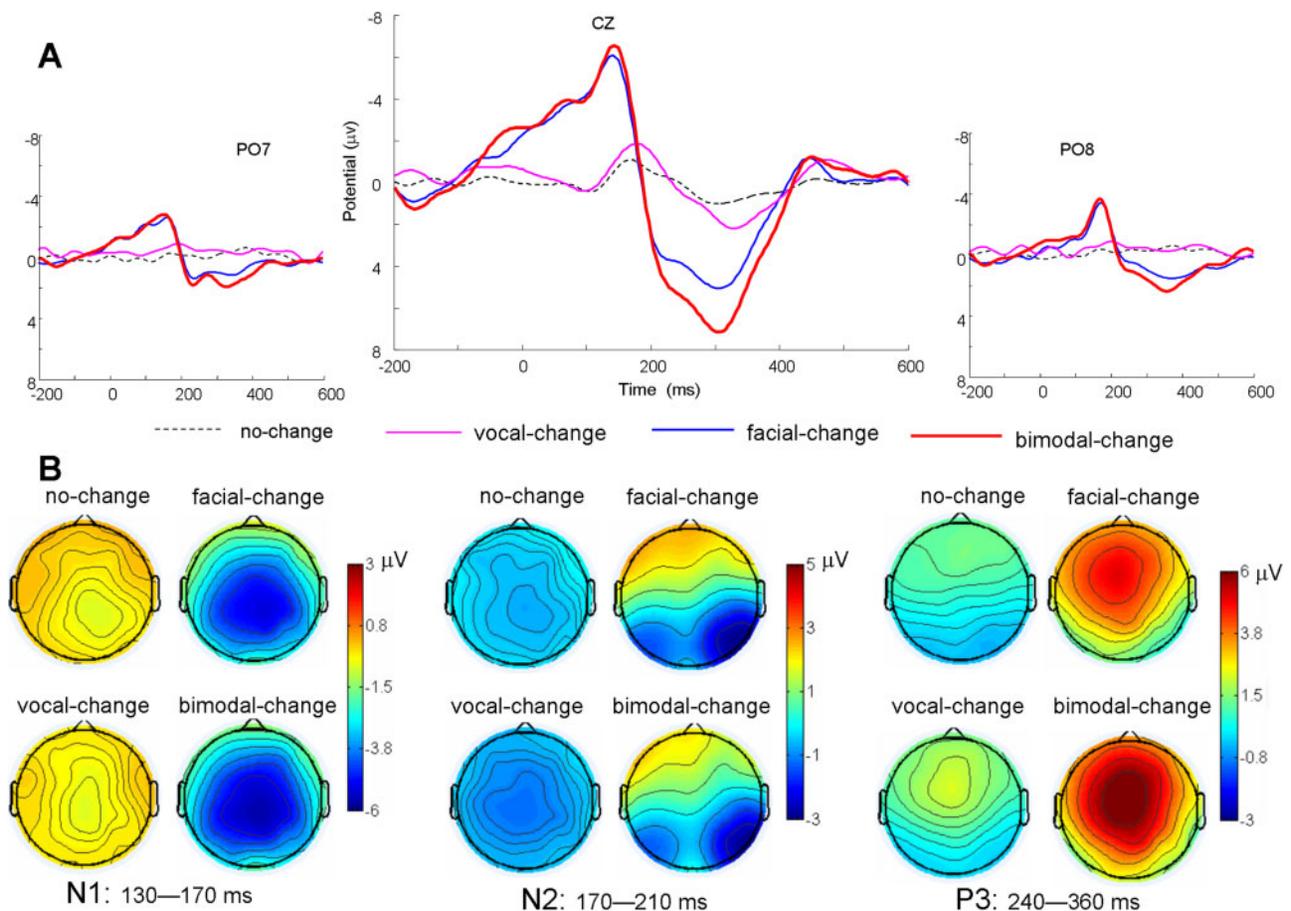


Fig. 3. Group-averaged ERP voltage waveforms, and scalp topography for N1, N2 and P3 as a function of condition. (A) Exemplar waveforms from CZ, PO7 and PO8 are displayed for nochange, vocal-change, facial-change and bimodal-change. (B) The ERP topography (top view shown) for N1, N2 and P3 components as a function of modality

were re-segmented to 800 ms epochs time-locked to the splicing point, starting 200 ms prior to splicing point. The data were then low-pass filtered of 30 Hz. The extracted average waveforms for each participant and condition were used to calculate grand-average waveforms (see Figure 3 for ERPs). To depict the distribution difference, nine scalp regions of interest (SROI) were defined¹ (for regional averaging, Dien and Santuzzi, 2004). Based on the previous studies addressing emotional change (Campanella *et al.* 2002; Chen *et al.* 2011) and permutation test implemented in the *statcond* function of EEGLAB, the amplitudes over 130–170, 170–210 and 240–360 ms after the splicing point were averaged for N1, N2 and P3 respectively for statistical analysis.

ERSP. The artifact free data were applied to wavelet decomposition using EEGLAB12.0.1.0b running under MATLAB. Changes in event-related spectral power response (in dB) were computed by the ERSP index (Delorme and Makeig, 2004) (1),

$$\text{ERSP}(f, t) = \frac{1}{n} \sum_{k=1}^n (F_k(f, t))^2 \quad (1)$$

where, for n trials, $F_k(f, t)$ is the spectral estimate of trial k at frequency f and time t . Power values were normalized with respect to a 200 ms pre-splicing point baseline and transformed into decibel scale ($10 \cdot \log_{10}$ of the signal). ERSPs were averaged over trials for each condition and transformed into time-frequency plots (see Figure 4 for ERSP). For conciseness, only the data of interest (4–30 Hz over –200 to 1000 ms) were presented. Based on the permutation test implemented in the *statcond* function of EEGLAB and previous studies addressing emotional change (Chen *et al.* 2015) and bimodal emotion integration (Jessen and Kotz, 2011), ERSP in theta band (4–8 Hz) during 50–500 ms, alpha band (8–12 Hz) during 500–900 ms and beta band (16–24 Hz) during 200–600 ms were averaged for statistical analysis.

Statistical analysis. To replicate previous emotional change effect, the ERP and ERSP data of interest were subjected to repeated measures ANOVAs with *condition* (no-change, facial-change, vocal-change and bimodal-change) and SROI as within-subject factors. Pair-wise comparisons were carried out between no-change and change conditions² (facial-change, vocal-change and bimodal-change) if significant effects involving *condition* were observed (listed in Table 1). Moreover, to examine whether bimodal emotional cues show facilitation effect and cross-modal integration, facial change effect (facial-change – no-change, F), vocal change effect (vocal-change – no-change, V), bimodal change effect (bimodal-change – no-change, FV) and summed unimodal change effect (F + V) on ACC, RT, ERP and ERSP were calculated and subjected to repeated measures ANOVAs with *change-type* (F, V, FV and F + V) as within subject factor. Pair-wise comparisons were carried out between FV and other change effect (F, V and F + V) if significant effects involving *change-type* were observed. The degrees of freedom of the F-ratio were

¹ Left anterior: F3, F5, F7, FC3, FC5 and FT7; middle anterior: F1, FZ, F2, FC1, FCZ and FC2; right anterior: F4, F6, F8, FC4, FC6 and FT8; left central: C3, C5, T7, CP3, CP5 and TP7; middle central: C1, CZ, C2, CP1, CPZ and CP2, right central: C4, C6, T8, CP4, CP6 and TP8; left posterior: P3, P5, P7, PO3, PO7 and O1; middle posterior: P1, PZ, P2, POZ and OZ; right posterior: P4, P6, P8, PO4, PO8 and O2.

² The comparisons without no-change condition in change analyses and those without FV in integration analyses were not reported since they are not directly relevant to the current argument.

corrected according to the Greenhouse–Geisser method and multiple comparisons were Bonferroni adjusted in all these analyses. The effect sizes were shown as partial eta squared (η_p^2).

Result

Behavioral performance

The repeated ANOVA of ACC revealed a significant main effect of *change-type* [$F(3,51) = 12.26, P < .001, \eta_p^2 = .46$], with the change effect significantly larger for FV (.04 ± .02) than for V (–.09 ± .03, $P < .001$) and F + V (–.07 ± .05, $P < .05$), and marginally larger for FV than for F (.01 ± .02, $P = .09$). The repeated measures ANOVA for RTs yielded a significant main effect of *change-type* [$F(3,51) = 26.49, P < .001, \eta_p^2 = .61$], with the change effect larger for FV (325 ± 44 ms) than for V (58 ± 26 ms, $P < .001$) and F (269 ± 52 ms, $P < .01$), but not for F + V (327 ± 71 ms).

ERP results

The analysis of N1 amplitudes showed a significant main effect of *condition* [$F(3,51) = 57.27, P < .001, \eta_p^2 = .77$], with the N1 amplitudes smaller for no-change than for facial-change and bimodal-change, but not for vocal-change (see Table 1 for details, the same hereinafter). The integration analysis yielded a significant main effect of *change-type* [$F(3,51) = 43.79, P < .001, \eta_p^2 = .72$], with the change effect larger for FV than for V and F, but not for F + V.

The repeated measures ANOVAs on N2 showed a significant interaction of *condition* × SROI [$F(24,408) = 15.23, P < .001, \eta_p^2 = .47$]. The follow-up analyses yielded that vocal-change elicited larger negative deflection compared with no-change over left-central, middle-central, left-posterior and middle-posterior areas. Meanwhile, facial-change and bimodal change elicited larger negative deflection relative to no-change over right-posterior area. The integration analysis yielded a significant interaction of *change-type* × SROI [$F(24,408) = 13.29, P < .001, \eta_p^2 = .44$]. Simple effect analysis found that the change effect was smaller for FV than for F + V over left-posterior area, and smaller for FV than for V over left-anterior area.

The repeated measures ANOVAs on P3 amplitudes showed a significant main effect of *condition* [$F(3,51) = 56.00, P < .001, \eta_p^2 = .77$] and a significant interaction of *condition* × SROI [$F(24,408) = 5.82, P < .001, \eta_p^2 = .26$]. Simple effect analysis showed that the P3 amplitudes were significantly smaller for no-change than for facial-change and bimodal-change, but not for vocal-change over all brain regions. The non-significant change effect of vocal-change might be due to its later latency and shorter duration. We thus carried out a repeated ANOVA on the peak amplitudes of P3 at 320 ms, which revealed a significant interaction between *condition* and SROI [$F(24,408) = 3.86, P < .01, \eta_p^2 = .19$]. The follow-up analyses yielded that the amplitudes for vocal-change were significantly larger than those for no-change over middle-anterior area. The integration analysis for mean P3 amplitudes yielded a significant main effect of *change-type* [$F(3,51) = 42.77, P < .001, \eta_p^2 = .72$], and a significant interaction of *change-type* × SROI [$F(24,408) = 4.34, P < .01, \eta_p^2 = .20$]. Further analysis yielded that the change effect was significant larger for FV than for F, V and F + V over all brain areas.

ERSP data

The repeated measures ANOVAs for theta power showed a significant main effect of *condition* [$F(3,51) = 39.21, P < .001, \eta_p^2 = .70$], and a significant interaction of *condition* × SROI [$F(24,408) = 2.51$,

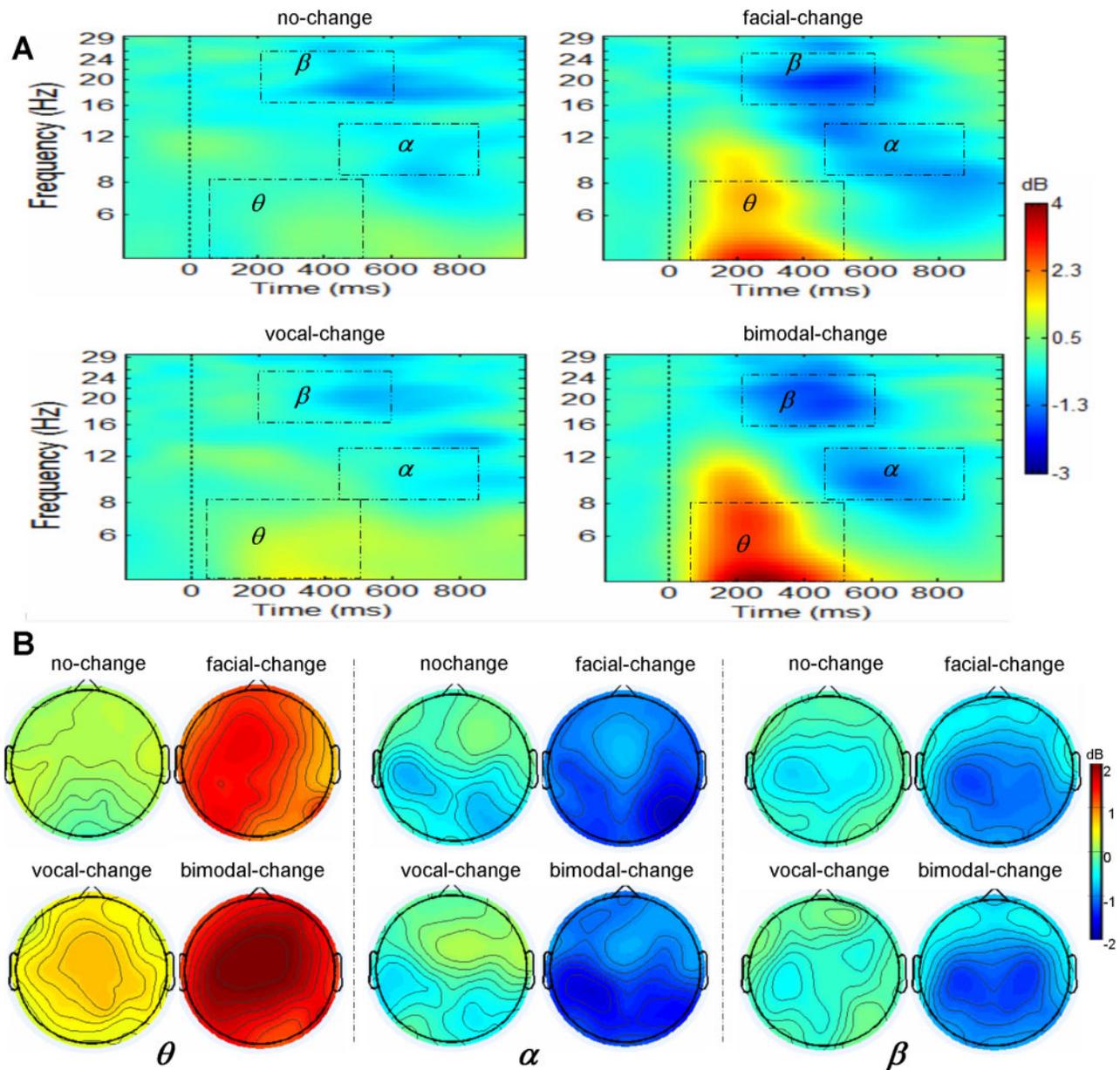


Fig. 4. Group-averaged ERS and scalp topography, for theta, alpha and beta as a function of modality. (A) Exemplar spectral maps at CZ are displayed for theta, alpha and beta as a function of modality. (B) The ERS topography (top view shown) for theta, alpha and beta as a function modality

$P < .05$, $\eta_p^2 = .13$). Further analysis showed that the theta synchronization was significantly smaller for no-change than for facial-change and bimodal-change over all brain areas, and smaller for no-change than for vocal-change over middle-anterior area. The integration analysis found a significant main effect of *change-type* [$F(3,51) = 15.05$, $P < .001$, $\eta_p^2 = .47$] and a significant interaction of *change-type* \times *SROI* [$F(24,408) = 2.48$, $P < .05$, $\eta_p^2 = .13$]. Further analysis showed the change effect was larger for FV than for V and F over all brain regions, and larger for FV than for F + V over right-anterior area.

The repeated measures ANOVAs on alpha activity only showed a significant main effect of *condition* [$F(3,51) = 7.83$, $P < .01$, $\eta_p^2 = .32$], with smaller alpha desynchronization for no-change than for facial-change and bimodal-change, but not for vocal-change. The integration analysis found a significant main effect of *change-type* [$F(3,51) = 6.52$, $P < .01$, $\eta_p^2 = .28$], with the change effect larger for FV than for V, but not for F and F + V.

The repeated measures ANOVAs on beta activity showed a significant main effect of *condition* [$F(3,51) = 15.53$, $P < .001$, $\eta_p^2 = .48$], and a significant interaction of *condition* \times *SROI* [$F(24,408) = 2.06$, $P < .05$, $\eta_p^2 = .11$]. Further analysis showed that the beta desynchronization for no-change was smaller than for facial-change and bimodal-change over middle-central, left-posterior, middle-posterior and right-posterior areas, but not than for vocal-change. The integration analysis only revealed a significant main effect of *change-type* [$F(3,51) = 11.99$, $P < .001$, $\eta_p^2 = .41$], with the change effect larger for FV than for V, but not for F and F + V.

Correlation analyses

To verify whether the superadditivity measured in P3 and theta activity is a neural index underlying the behavioral facilitation effect, we conducted a Spearman correlation analysis on the

Table 1. Results of follow-up analyses on change analyses and integration

		Change analyses				Integration analyses					
		SROI	No-change	Vocal change	Facial change	Bimodal change	SROI	FV	V	F	F + V
ERP (μ V)	N1	ALL	$-.20 \pm .10$ null	$-.55 \pm .13$ ns.	$-3.32 \pm .42$ $P < .001$	$-3.77 \pm .39$ $P < .001$	ALL	$-3.58 \pm .38$ null	$-.35 \pm .18$ $P < .001$	$-3.35 \pm .12$ $P < .05$	$-3.47 \pm .38$ ns.
	N2	LC/MCLP/MP	$-.39 \pm .12$ null	$-1.08 \pm .13$ $P < .001$	$-.41 \pm .65$ ns.	$-.43 \pm .66$ ns.	LA	$2.05 \pm .57$ null	$-.45 \pm .26$ $P < .001$	$2.22 \pm .49$ ns.	$1.76 \pm .56$ ns.
		RP	$-.41 \pm .12$ null	$-.86 \pm .17$ ns.	$-2.41 \pm .59$ $P < .001$	$-2.37 \pm .58$ $P < .001$	LP	$-.80 \pm .59$ null	$-.92 \pm .61$ ns.	$-.55 \pm .18$ ns.	$-1.46 \pm .62$ $P < .05$
	P3	ALL	$.50 \pm .16$ null	$.65 \pm .19$ ns.	$2.84 \pm .33$ $P < .001$	$6.20 \pm .82$ $P < .001$	ALL	$3.12 \pm .32$ null	$.16 \pm .17$ $P < .001$	$2.34 \pm .32$ $P < .01$	$2.50 \pm .36$ $P < .05$
ERSP (dB)	θ	ALL	$.08 \pm .08$ null	$.44 \pm .13$ $P < .05$	$1.06 \pm .20$ $P < .001$	$1.54 \pm .15$ $P < .001$	RA	$1.52 \pm .15$ null	$.18 \pm .16$ $P < .001$	$.68 \pm .20$ $P < .01$	$.87 \pm .34$ $P < .05$
	α	ALL	$-.39 \pm .07$ null	$-.29 \pm .15$ ns.	$-1.26 \pm .33$ $P < .05$	$-1.26 \pm .30$ $P < .05$	ALL	$-.66 \pm .28$ null	$-.04 \pm .13$ $P < .05$	$-.30 \pm .24$ ns.	$-.48 \pm .14$ ns.
	β	ALL	$-.37 \pm .10$ null	$-.29 \pm .09$ ns.	$-.86 \pm .14$ $P < .05$	$-.93 \pm .15$ $P < .01$	ALL	$-.56 \pm .16$ null	$.08 \pm .09$ $P < .001$	$-.48 \pm .14$ ns.	$-.40 \pm .22$ ns.

Note: ALL denotes all nine brain areas, whereas LC, MC, LP, MP, RP, MA and RA are short for left-central, middle-central, left-posterior, middle-posterior, right-posterior, middle-anterior and right-anterior area, respectively.

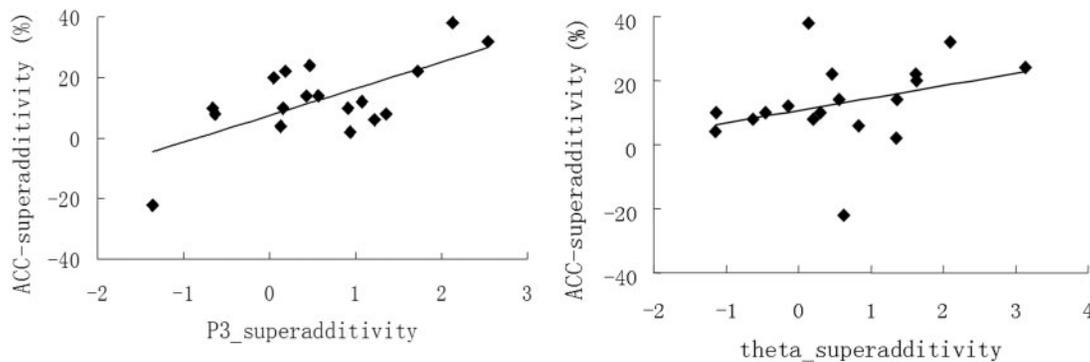


Fig. 5. (Left) The scatterplot for the correlation between superadditivity in ACC and the superadditivity in P3 amplitudes; (right) The scatterplot for the correlation between superadditivity in ACC and the superadditivity in theta synchronization

superadditive effect ($FV - [F+V]$) between the electrophysiological parameters and ACC. The results showed that the superadditive effect in ACC was positively correlated that both in P3 amplitude ($r=0.45$, $P=0.03$) and in theta synchronization ($r=0.41$, $P=0.04$; see Figure 5).

Discussion

This study observed that bimodal emotional changes were detected with higher accuracy and shorter response latency relative to unimodal emotional changes. Moreover, the detection of emotional changes, regardless of modalities, was associated with enhanced amplitudes in the N2/ P3 complex, as well as greater theta synchronization. More importantly, ACC, P3 amplitudes and theta synchronization showed superadditive responses between bimodal emotional cues.

An important aim of the current study is to test whether the electrophysiological responses associated with unimodal emotional changes can be replicated in bimodal emotional changes. As predicted, attentive processing of emotional changes elicited N2/P3 enhancement regardless of stimulus modality. Consistent with previous studies (Lovrich et al.,

1986; Breton et al., 1988), vocal changes elicited broadly distributed N2, while facial changes and bimodal changes elicited right-posterior distributed N2. Moreover, the P3 component for all the change conditions displayed a frontal-central distribution, comparable to the P3a in the previous studies (Campanella et al. 2002; Chen et al. 2011). P3a has been reported in response to novel stimuli (Berti and Schröger, 2001; Campanella et al., 2002) and is linked with attentive processing of both facial (Campanella et al., 2002, 2013) and vocal (Wambacq and Jerger, 2004; Chen et al., 2011; Paulmann et al., 2012) emotional changes. In the similar vein, this study created emotional changes by cross-splicing two parts of stimuli, in which the context preceding the cross-splicing point allowed participants to generate expectation, and the subsequent emotional change violated expectation, thus leading to prominent N2 and P3a activity.

In addition to the ERP analysis, we found that all types of emotional changes induced enhanced theta synchronization, consistent with previous studies (Chen et al. 2012, 2015). It has been indicated that theta synchronization is related to expectation violation (Tzur and Berger, 2007; Hsiao et al., 2009; Cavanagh et al., 2010) and is sensitive to the salience of the

violation (Tzur and Berger, 2007; Chen et al. 2012). This study created emotional changes via cross-splicing, and this method has been established to elicit expectation violation (Kotz and Paulmann, 2007; Chen et al., 2011). Thus, it is most likely that the expectation violation elicited by emotional changes resulted in significant theta synchronization.

Besides N2/P3a and theta synchronization, the facial and bimodal emotional change elicited significant N1 enhancement as well as larger alpha/beta desynchronization relative to the nochange condition. The N1 enhancement is consistent with the finding of Astikainen et al. (2013), which showed that visual stimulus changes elicited a posterior N1. Alpha desynchronization is associated with attention allocation (Jessen and Kotz, 2011; Klimesch, 2012) and deviant stimulus detection in the active oddball paradigm (Yordanova et al., 2001), whereas beta desynchronization is associated with deviant processing (Kim and Chung, 2008; Chen et al., 2015) and action planning (Tzagarakis et al., 2010). In line with these findings, unpredictable emotional changes should have elicited increased attention allocation, followed by change detection and action planning, for effective emotional decoding, which may have consequently induced alpha and beta desynchronization. However, it is worth noting that vocal change did not induce salient N1 enhancement as well as alpha and beta suppression. This might be due to the fact that vocal changes are not so salient when accompanied with unchanged facial expression. Meanwhile, our observation of similar brain responses for facial and bimodal emotional changes, is in line with previous studies reporting that facial information exerts a larger influence on bimodal emotion decoding relative to vocal information (Jessen and Kotz, 2011; Doi and Shinohara, 2014; Watson et al., 2014).

Taken together, the data in both time and frequency domains replicated prior findings from unimodal emotional change perception, and confirmed our hypothesis that the perception of bimodal emotional changes is associated with the same neurophysiologic underpinning as that of unimodal change perception. This suggests that the brain employs a similar neural mechanism for emotional change perception regardless of sensory modalities.

In line with previous findings that bimodal emotional signals are processed with higher processing efficiency than unimodal emotional signals (Pourtois et al., 2005; Ethofer et al., 2006; Kreifelts et al., 2007; Jessen et al., 2012), we observed that bimodal emotional change was perceived with higher accuracy and shorter response latency relative to unimodal conditions. Moreover, consistent with the behavioral data, bimodal emotional change evoked larger P3a and theta synchronization than each unimodal condition. Previous studies have reported that multimodal emotions elicited smaller N1 (Jessen and Kotz, 2011; Jessen et al., 2012; Kokinous et al., 2014) and P2 amplitudes (Paulmann et al., 2009; Kokinous et al., 2014) compared with unimodal conditions. Also, it has been reported that congruent multimodal emotions induced stronger suppression of alpha (Chen et al., 2010; Jessen and Kotz, 2011; Jessen et al., 2012) and beta activity (Jessen et al., 2012), and enhanced gamma activity (Hagan et al., 2009). Consistent with these evidences, this study observed that the behavioral and electrophysiological responses for bimodal cues were more pronounced than for unimodal conditions, providing direct support for cross-modal facilitation in emotional change recognition. However, the facilitation effect was not found in N1 amplitudes, alpha or beta desynchronization. This discrepancy might be due to our experimental paradigm which focused on emotional change detection rather

than emotional identification (Chen et al., 2010; Jessen and Kotz, 2011).

As stated before, a facilitation of bimodal cues relative to unimodal cues alone is insufficient to verify the occurrence of perceptual integration of audiovisual emotional changes (Stein et al., 2009; Jessen et al., 2012), while the superadditivity is an accepted, robust criterion for the detection of this integration (Hagan et al., 2009; Jessen and Kotz, 2011). We showed superadditive effects in ACC, P3 and theta synchronization. That is, all these measures were larger for the bimodal condition than for the sum of facial and vocal conditions. This is in line with the previous studies, which showed that congruent auditory and visual emotion demonstrated a significant superadditive effect in neural oscillatory indexes (Hagan et al., 2009; Jessen and Kotz, 2011). Moreover, the superadditivities in P3a amplitudes and theta synchronization were both positively correlated with superadditivity in ACC, suggesting that P3 activity and theta synchronization are important neural mechanisms subserving this behavioral bimodal integration effect during emotional change recognition. It is worth noting that we did not obtain superadditivity in gamma and beta activity (Hagan et al., 2009; Jessen and Kotz, 2011). This may be because our task required subjects to detect emotional change voluntarily, which determined that the detection of contextual violation was the dominant cognitive process. There were abundant evidences showing that the detection of contextual violation was linked with P3a activity (Campanella et al., 2002; Chen et al., 2011) and theta synchronization (Chen et al., 2012, 2015).

In short, this study adds to the literature by providing the first evidence for multisensory integration in emotional change perception. Previous studies have indicated that neurophysiological reactions to multimodal emotional stimuli are more effective than those to unimodal stimuli in differentiating subclinical alexithymia (Delle-Vigne et al., 2014), anxiety and depression (Campanella et al., 2010). Thus, the current findings, particularly P3 and theta oscillations as indexes of emotional change integration, may have clinical implication in differentiating affective disorder from healthy population. Moreover, the current study only focused on facial and vocal emotional change perception, leaving the role of body (Jessen and Kotz, 2011; Aviezer et al., 2012) and semantic cues (Paulmann et al., 2009) untouched. All these questions deserve investigation in future studies.

Conclusion

This study demonstrated that the detection of emotional changes, regardless of modalities, was associated with enhanced N2 and P3 components, as well as greater theta synchronization. More importantly, in addition to the behavioral facilitation effect for bimodal emotional change, the P3 amplitudes and theta synchronization for bimodal emotional change were not only larger than those for facial or vocal modality alone, but was also larger than the sum of the two unimodal conditions. The superadditive effects in P3 amplitudes and theta synchronization were both positively related to the magnitude of the bimodal superadditive effects observed in the accuracy of identifying emotional changes. These behavioral and electrophysiological data illustrated a robust effect of facial-vocal bimodal integration during the detection of emotional changes, which is most likely mediated by the P3 and theta oscillations in brain responses.

Funding

This work was supported by the National Natural Science Foundation of China under grant 31300835 and 31371042, the General Projects for Humanities and Social Science Research of Ministry of Education under grant 12XJC190003, Fundamental Research Funds for the Central Universities under grant 14SZYB07, and interdisciplinary Incubation project of learning science of Shaanxi Normal university.

Conflict of interest. None declared.

References

- Astikainen, P., Cong, F., Ristaniemi, T., Hietanen, J.K. (2013). Event-related potentials to unattended changes in facial expressions: detection of regularity violations or encoding of emotions? *Frontiers in Human Neuroscience*, 7. doi: 10.3389/fnhum.2013.00557.
- Aviezer, H., Trope, Y., Todorov, A. (2012). Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science*, 338, 1225–9.
- Berti, S., Schröger, E. (2001). A comparison of auditory and visual distraction effects: behavioral and event-related indices. *Cognitive Brain Research*, 10, 265–73.
- Breton, F., Ritter, W., Simson, R., Vaughan, H.G., Jr. (1988). The N2 component elicited by stimulus matches and multiple targets. *Biological Psychology*, 27, 23–44.
- Calvert, G.A. (2001). Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cerebral Cortex*, 11, 1110–23.
- Campanella, S., Belin, P. (2007). Integrating face and voice in person perception. *Trends in Cognitive Science*, 11, 535–43.
- Campanella, S., Gaspard, C., Debatisse, D., Bruyer, R., Crommelinck, M., Guerit, J.-M. (2002). Discrimination of emotional facial expressions in a visual oddball task: an ERP study. *Biological Psychology*, 59, 171–86.
- Campanella, S., Bruyer, R., Froidbise, S., et al. (2010). Is two better than one? A cross-modal oddball paradigm reveals greater sensitivity of the P300 to emotional face-voice associations. *Clinical Neurophysiology*, 121, 1855–62.
- Campanella, S., Bourguignon, M., Peigneux, P., et al. (2013). BOLD response to deviant face detection informed by P300 event-related potential parameters: a simultaneous ERP-fMRI study. *NeuroImage*, 71, 92–103.
- Cavanagh, J.F., Frank, M.J., Klein, T.J., Allen, J.J.B. (2010). Frontal theta links prediction errors to behavioral adaptation in reinforcement learning. *NeuroImage*, 49, 3198–209.
- Chen, Y.H., Edgar, J.C., Holroyd, T., et al. (2010). Neuromagnetic oscillations to emotional faces and prosody. *European Journal of Neuroscience*, 31, 1818–27.
- Chen, X., Zhao, L., Jiang, A., Yang, Y. (2011). Event-related potential correlates of the expectancy violation effect during emotional prosody processing. *Biological Psychology*, 86, 158–67.
- Chen, X., Yang, J., Gan, S., Yang, Y. (2012). The contribution of sound intensity in vocal emotion perception: behavioral and electrophysiological evidence. *PLoS One*, 7, e30278.
- Chen, X., Pan, Z., Wang, P., Zhang, L., Yuan, J. (2015). EEG oscillations reflect task effects for the change detection in vocal emotion. *Cognitive Neurodynamics*, 9, 351–8.
- Delorme, A., Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134, 9–21.
- Delle-Vigne, D., Kornreich, C., Verbanck, P., Campanella, S. (2014). Subclinical alexithymia modulates early audio-visual perceptive and attentional event-related potentials. *Frontier in Human Neuroscience*, 8.
- Dien, J., Santuzzi, A.M. (2004). Application of repeated measures ANOVA to high-density ERP datasets: A review and tutorial. In Handy T. (Ed.), *Event-related potentials: A methods handbook*, pp. 57–82, Cambridge, Mass: MIT Press.
- Doi, H., Shinohara, K. (2015). Unconscious presentation of fearful face modulates electrophysiological responses to emotional prosody. *Cerebral Cortex*, 25, 817–32.
- Ethofer, T., Anders, S., Erb, M., et al. (2006). Impact of voice on emotional judgment of faces: an event-related fMRI study. *Human Brain Mapping*, 27, 707–14.
- Fujimura, T., Okanoya, K. (2013). Event-related potentials elicited by pre-attentive emotional changes in temporal context. *PLoS ONE*, 8, e63703.
- Hagan, C.C., Woods, W., Johnson, S., Calder, A.J., Green, G.G., Young, A.W. (2009). MEG demonstrates a supra-additive response to facial and vocal emotion in the right superior temporal sulcus. *Proc Natl Acad Sci USA*, 106, 20010–5.
- Hsiao, F.J., Wu, Z.A., Ho, L.T., Lin, Y.Y. (2009). Theta oscillation during auditory change detection: an MEG study. *Biological Psychology*, 81, 58–66.
- Jessen, S., Kotz, S.A. (2011). The temporal dynamics of processing emotions from vocal, facial, and bodily expressions. *NeuroImage*, 58, 665–74.
- Jessen, S., Obleser, J., Kotz, S.A. (2012). How bodies and voices interact in early emotion perception. *PLoS One*, 7, e36070.
- Kim, J.S., Chung, C.K. (2008). Language lateralization using MEG beta frequency desynchronization during auditory oddball stimulation with one-syllable words. *NeuroImage*, 42, 1499–507.
- Kimura, M., Kondo, H., Ohira, H., Schröger, E. (2012). Unintentional temporal context-based prediction of emotional faces: an electrophysiological study. *Cerebral Cortex*, 22, 1774–85.
- Klasen, M., Chen, Y.H., Mathiak, K. (2012). Multisensory emotions: perception, combination and underlying neural processes. *Review in the Neurosciences*, 23, 381–92.
- Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Science*, 16, 606–17.
- Kokinous, J., Kotz, S.A., Tavano, A., Schroger, E. (2014). The role of emotion in dynamic audiovisual integration of faces and voices. *Social Cognitive and Affective Neuroscience*. doi: 10.1093/scan/nsu105
- Kotz, S.A., Paulmann, S. (2007). When emotional prosody and semantics dance cheek to cheek: ERP evidence. *Brain Research*, 1151, 107–18.
- Kreifelts, B., Ethofer, T., Grodd, W., Erb, M., Wildgruber, D. (2007). Audiovisual integration of emotional signals in voice and face: an event-related fMRI study. *NeuroImage*, 37, 1445–56.
- Lovrich, D., Simson, R., Vaughan, H.G., Jr., Ritter, W. (1986). Topography of visual event-related potentials during geometric and phonetic discriminations. *Electroencephalogr Clin Neurophysiol*, 65, 1–12.
- Luck, S. (2005). *An introduction to the event-related potential technique*. London: The MIT press.
- Makeig, S., Debener, S., Onton, J., Delorme, A. (2004). Mining event-related brain dynamics. *Trends in Cognitive Sciences*, 8, 204–10.
- Massaro, D.W., Egan, P.B. (1996). Perceiving affect from the voice and the face. *Psychonomic Bulletin Review*, 3, 215–21.

- Paulmann, S., Jessen, S., Kotz, S.A. (2009). Investigating the multimodal nature of human communication. *Journal of Psychophysiology*, **23**, 63–76.
- Paulmann, S., Jessen, S., Kotz, S.A. (2012). It's special the way you say it: an ERP investigation on the temporal dynamics of two types of prosody. *Neuropsychologia*, **50**, 1609–20.
- Pourtois, G., de Gelder, B., Bol, A., Crommelinck, M. (2005). Perception of facial expressions and voices and of their combination in the human brain. *Cortex*, **41**, 49–59.
- Sabbagh, M.A., Moulson, M.C., Harkness, K.L. (2004). Neural correlates of mental state decoding in human adults: an event-related potential study. *Journal of Cognitive Neuroscience*, **16**, 415–26.
- Salovey, P., Mayer, J.D. (1989). Emotional intelligence. *Imagination, Cognition and Personality*, **9**, 185–211.
- Stefanics, G., Csukly, G., Komlósi, S., Czobor, P., Czigler, I. (2012). Processing of unattended facial emotions: a visual mismatch negativity study. *NeuroImage*, **59**, 3042–9.
- Stein, B.E., Stanford, T.R., Ramachandran, R., Perrault, T.J., Jr., Rowland, B.A. (2009). Challenges in quantifying multisensory integration: alternative criteria, models, and inverse effectiveness. *Experimental Brain Research*, **198**, 113–26.
- Stekelenburg, J.J., Vroomen, J. (2007). Neural correlates of multisensory integration of ecologically valid audiovisual events. *Journal of Cognitive Neuroscience*, **19**, 1964–73.
- Thonnessen, H., Boers, F., Dammers, J., Chen, Y.H., Norra, C., Mathiak, K. (2010). Early sensory encoding of affective prosody: neuromagnetic tomography of emotional category changes. *NeuroImage*, **50**, 250–9.
- Tzagarakis, C., Ince, N.F., Leuthold, A.C., Pellizzer, G. (2010). Beta-band activity during motor planning reflects response uncertainty. *Journal of Neuroscience*, **30**, 11270–7.
- Tzur, G., Berger, A. (2007). When things look wrong: Theta activity in rule violation. *Neuropsychologia*, **45**, 3122–6.
- Van Kleef, G.A. (2009). How emotions regulate social life the emotions as social information (EASI) model. *Current Directions in Psychological Science*, **18**, 184–8.
- Wambacq, I.J.A., Jerger, J.F. (2004). Processing of affective prosody and lexical-semantics in spoken utterances as differentiated by event-related potentials. *Cognitive Brain Research*, **20**, 427–37.
- Watson, R., Latinus, M., Noguchi, T., Garrod, O., Crabbe, F., Belin, P. (2014). Crossmodal adaptation in right posterior superior temporal sulcus during face–voice emotional integration. *The Journal of Neuroscience*, **34**, 6813–21.
- Winkler, I., Denham, S., Nelken, I. (2009). Modeling the auditory scene: predictive regularity representations and perceptual objects. *Trends in Cognitive Sciences*, **13**, 532–40.
- Xu, Q., Yang, Y., Wang, P., Sun, G., Zhao, L. (2013). Gender differences in preattentive processing of facial expressions: an ERP study. *Brain Topography*, **26**, 488–500.
- Yordanova, J., Kolev, V., Polich, J. (2001). P300 and alpha event-related desynchronization (ERD). *Psychophysiology*, **38**, 143–52.
- Young, A., Perrett, D., Calder, A., Sprengelmeyer, R., Ekman, P. (2002). *Facial expressions of emotion: stimuli and tests (FEEST)*. Bury St. Edmunds, UK: Thames Valley Test Company.