The effects of timbre on neural responses to musical emotion

Zhang, W., Liu, F. ORCID: https://orcid.org/0000-0002-7776-0222, Zhou, L., Wang, W., Jiang, H. and Jiang, C. (2019) The effects of timbre on neural responses to musical emotion. Music Perception, 37 (2). pp. 134-146. ISSN 0730-7829 doi: 10.1525/mp.2019.37.2.134 Available at https://centaur.reading.ac.uk/85954/

To link to this article DOI: http://dx.doi.org/10.1525/mp.2019.37.2.134

Publisher: University of California Press

The effects of timbre on neural responses to musical emotion

Weixia Zhang[a], Fang Liu[b], Linshu Zhou[c], Wanqi Wang[c], Hanyuan Jiang[d], and Cunmei Jiang[c,e]*

[a] *Department of Psychology, College of Education, Shanghai Normal University, Shanghai, China, 200234*

[b] *School of Psychology and Clinical Language Sciences, University of Reading, Reading, UK, RG6 6AL*

[c] *Music College, Shanghai Normal University, Shanghai, China, 200234*

[d] *Faculty of Humanities and Arts, Macau University of Science and Technology, Macau, China, 519020*

[e] *Institute of Psychology, Shanghai Normal University, Shanghai, China, 200234*

Running head: Effects of timbre on musical emotion processing

* Corresponding author to:
Dr. Cunmei Jiang
Music College
Shanghai Normal University
100 E. Guilin Road
Shanghai, 200234, China
Tel: 0086-21-64322990; Fax: 0086-21-64322935
Electronic mail: cunmeijiang@126.com

Abstract

Timbre is an important factor that affects the perception of emotion in music. To date, little is known about the effects of timbre on neural responses to musical emotion. To address this issue, we used ERPs to investigate whether there are different neural responses to musical emotion when the same melodies are presented in different timbres. With a cross-modal affective priming paradigm, target faces were primed by affectively congruent or incongruent melodies without lyrics presented in violin, flute, and the voice. Results showed a larger P3 and a larger left anterior distributed LPC in response to affectively incongruent versus congruent trials in the voice version. For the flute version, however, only the LPC effect was found, which was distributed over centro-parietal electrodes. Unlike the voice and flute versions, an N400 effect was observed in the violin version. These findings revealed different patterns of neural responses to emotional processing of music when the same melodies were presented in different timbres, and provide evidence to confirm the hypothesis that there are specialized neural responses to the human voice.

Keywords: timbre, affective priming, N400, LPC, P3

Introduction

Timbre is one of the most important acoustic attributes in our environment (Menon et al., 2002). Through timbre, listeners can distinguish two sounds of identical pitch, duration, and intensity (Griffiths & Warren, 2004; McAdams, Cunible, Carlyon, Darwin, & Russell, 1992). Among different timbres, the human voice has been shown to be associated with specialized neural activities (for a review see Belin, Fecteau, & Bédard, 2004). In particular, fMRI studies (Belin, Zatorre, & Ahad, 2002; Belin, Zatorre, Lafaille, Ahad, & Pike, 2000) have indicated that the superior temporal sulcus (STS) shows greater response to the human voice than to non-vocal stimuli. Electrophysiological evidence has also revealed a larger amplitude of voice-specific response (VSR) (Levy, Granot, & Bentin, 2001) and a larger fronto-temporal positivity to voice (FTPV) (Bruneau et al., 2013; Capilla, Belin, & Gross, 2012; Charest et al., 2009) in response to vocal compared with non-vocal stimuli such as environmental sounds. For non-vocal stimuli, it has also been suggested that the perception of different timbres involves different electrophysiological correlates (Aramaki, Besson, Kronland-Martinet, & Ystad, 2008; Crummer, Walton, Wayman, Hantz, & Frisina, 1994). For example, metal sounds elicited a smaller P200, larger N280 and negative slow wave than wood and glass sounds, whereas the latter of which did not differ from each other (Aramaki et al., 2008). In short, the aforementioned studies suggest different neural activities during the discrimination of sound stimuli of different timbre.

In the music domain, timbre is thought to be an important factor that affects the perception of emotions (Alluri & Toiviainen, 2010; Balkwill & Thompson, 1999; Barthet, Depalle, Kronland-Martinet, & Ystad, 2010; Bowman & Yamauchi, 2016; Eerola, Ferrer, & Alluri, 2012; Eerola, Friberg, & Bresin, 2013; Hailstone et al., 2009). Indeed, timbre has a robust contribution to emotional expressions in music (Eerola et al., 2013), and it can also enhance listeners' sensitivity to musical emotions (Balkwill & Thompson, 1999). It has also been suggested that certain emotions are best expressed by certain timbres but not by others (Behrens & Green, 1993; Gabrielsson

& Juslin, 1996; Paquette, Peretz, & Belin, 2013). For example, anger is best expressed by timpani rather than the singing voice, while fear is best expressed by the violin rather than the singing voice or timpani (Behrens & Green, 1993).

A few previous studies have examined neural responses to musical emotion by using a cross-modal affective priming paradigm. This paradigm is thought to be an appropriate method to examine the activation of affective representations (Hermans, De Houwer, & Eelen, 2001; Herring, Taylor, White, & Crites Jr, 2011). It has been widely employed in the investigations of musical emotion, where music excerpts are used to prime affectively congruent/incongruent words (Goerlich et al., 2012; Sollberge, Rebe, & Eckstein, 2003; Steinbeis & Koelsch, 2009) or pictures (Lense, Gordon, Key, & Dykens, 2012; Logeswaran & Bhattacharya, 2009). Therefore, unlike the semantic priming paradigm that focuses on the association of meaning between the primes and targets, the affective priming paradigm focuses on the association between the primes and targets in emotional features such as valence (Herring et al., 2011; Timmers & Crook, 2014).

The N400 or late positive component (LPC) effect has been reported by previous ERP studies as an indicator of the affective priming effect (e.g., Herring et al., 2011; Schirmer, Kotz, & Friederici, 2002; Zhang, Li, Gold, & Jiang, 2010). Although the N400 was initially interpreted as representing semantic integration (for a review see Kutas & Federmeier, 2010; Kutas & Hillyard, 1980), it has also been implicated in the processing of affective incongruence (Schirmer et al., 2002; Zhang, Lawson, Guo, & Jiang, 2006) and difficulty in affective integration between the primes and targets (Kamiyama, Abla, Iwanaga, & Okanoya, 2013; Zhang et al., 2010). Likewise, the LPC reflects increased attentional involvement activated by affective incongruence between the primes and targets when distributed over centro-parietal electrodes (Herring et al., 2011; Hinojosa, Carretié, Méndez-Bértolo, Míguez, & Pozo, 2009; Zhang, Kong, & Jiang, 2012), while the frontally distributed LPC reflects controlled attentional engagement (Leutgeb, Schäfer, Köchel, & Schienle, 2012).

Although previous research has suggested that the perception of vocal and non-vocal stimuli involves distinct neural processes, no study has yet examined whether emotional processing of musical melodies that are presented in different timbres would also show neural differences. In the present study, we investigated the effects of timbre on neural responses to musical emotion using the affective priming paradigm, with a 2 congruency (congruent vs. incongruent) × 3 timbre (violin, voice, flute) within-subjects design. Target facial expressions were primed by affectively congruent or incongruent melodies, which were presented in three timbre versions: the voice, violin, and flute. Two pretests were conducted to ensure the validity of the stimuli. The first pretest was to balance the performance level (i.e., how good the performance was) and performance style (including rubato, intensity, and phrasing) among the three versions, while the second pretest was to assess whether the musical stimuli were affectively congruent or incongruent with the faces and to confirm the validity of the stimuli.

We chose the voice, flute, and violin as different timbre conditions based on the following considerations. First, although previous research has shown that perception of the human voice involves distinct neural correlates compared with non-vocal sounds (Belin et al., 2000; Charest et al., 2009; Levy et al., 2001), it remains unknown whether there are different neural responses to musical emotion when the same melodies are presented in the voice and in other timbres. The inclusion of the voice in the present study would allow us to address this issue and to test the hypothesis that there is specialized neural processing of musical emotion in the human voice. Second, we chose the flute as another timbre condition because it is not only representative of wind instruments in an orchestra, but also resembles the voice in terms of structure and sound production. That is, the resonances of the vocal tract generate the timbre for both the voice and flute (Wolfe, Garnier, & Smith, 2009). Therefore, by comparing the flute with the voice, we would reveal whether timbres created by similar energy sources would still lead to different neural activities during emotion processing. Finally, the violin is representative of string instruments in an orchestra, which is different from the flute and voice in structure and sound production. Therefore, the inclusion of the voice, flute, and violin would allow us to compare the effects of the vocal and two representative instrumental timbres on neural responses to musical emotion. It is predicted that the neural processing of musical emotion may be different across the three timbre versions, given that timbre has shown significant effects on the processing of musical emotion at the behavioral level (e.g., Behrens & Green, 1993; Eerola et al., 2012; Hailstone et al., 2009).

Experimental procedure

Participants

Twenty-eight university students (*Mage* = 23.36 years, *SD* = 1.32 years; 14 males) participated as paid volunteers. All participants were right-handed, native speakers of Chinese, with no history of psychiatric or neurological diseases. All reported having normal hearing and normal or corrected-to-normal vision. None of them had received any extracurricular training in music. All participants signed a written consent form before the experiment.

Stimuli

Musical melodies and facial images were used as primes and targets, respectively. First, a total of 120 melodies were selected from European operas composed during classical or Romantic musical periods from around 1750–1900. Given that happy and sad melodies may reflect different musical scale (or interval) structures, we included 60 happy and 60 sad melodic stimuli in order to avoid the possible interactions between timbre type and scale structure of the melodies. The selections were based on musicological analysis and self-reports from the composers (e.g., a melody associated with the lyrics of “crying” was intended to express the emotion of *sadness*).

In order to enhance the ecological validity of the musical stimuli, we used recordings of performances, following previous studies on musical emotion and timbre (Balkwill & Thompson, 1999; Behrens & Green, 1993; Hailstone et al., 2009). Each melody was played by a violinist and a flutist, as well as sung by a vocalist with the syllable “la”, resulting in 360 musical stimuli in three versions. The performers were all female, and had all received professional music training over 18 years. They were informed that the three versions of each melody should be performed in a similar style with regard to rubato, intensity and phrasing, and they discussed how to perform each melody before recording. After recording, all musical stimuli were edited using Adobe Audition CS6 (Adobe Systems Inc) with 22.05 kHz sampling rate and 16-bit resolution. The mean duration of the melodies was on average 7 s, ranging from 2 to 9 s. The loudness of the melodies was normalized to approximately 68 dB SPL, fading out in 1s.

A pretest was conducted in order to rule out the differences in performance levels or styles among the three timbre conditions. Eight musicians (all with more than 10 years of professional music training) rated the similarity among the three versions of each melody with regard to rubato, intensity, phrasing, and the overall performance level using a 7-point scale (1 = very incongruent, 4 = not sure, 7 = very congruent). Only melodies with a mean value above 4 (the minimum standard) were chosen as experimental stimuli, which led to the selection of 240 musical stimuli (i.e., three versions of 80 melodies) as the potential prime stimuli. The results of the pretest showed that the three versions of the musical stimuli were performed in similar manners at similar performance levels (rubato: $M = 5.02$, $SD = 0.22$; intensity: $M = 5.19$, $SD = 0.38$; phrase: $M = 5.70$, $SD = 0.49$; overall performance level: $M = 5.25$, $SD = 0.33$).

240 emotional faces were selected as potential target stimuli from the Chinese Facial Affective Picture System (CFAPS) (Gong, Huang, Wang, & Luo, 2011), of which 120 expressed happiness and 120 expressed sadness. Each of the 240 musical stimuli was presented twice, followed by either an affectively congruent face or an incongruent face, thus resulting in 480 trials (see Figure 1 for examples).

Insert Figure 1, about here.

Another pretest was conducted to assess whether the musical stimuli were affectively congruent or incongruent with the faces. Twenty non-musicians who did not participate in the EEG experiment were asked to rate each prime-target pair regarding affective congruency using a 9-point scale (1 = very incongruent, 5 = not sure, 9 = very congruent). We used the 9-point instead of 7-point scale to get more detailed rating information. In the end, 360 prime-target pairs from the extreme ends (congruent pairs: ratings ≥ 6; incongruent pairs: ratings ≤ 4) of the continuous distribution of the congruent/incongruent values were selected as the final experimental stimuli with the constraints that each prime and target would be used twice, and that all three versions of the same melody would be selected.

An ANOVA taking congruency and timbre version as within-subjects factors was conducted on the affective congruency ratings to confirm whether the melody-face pairs were affectively congruent or incongruent. Results showed a main effect of congruency ($F_{(1,19)} = 297.13$, $p < .001$, $\eta_p^2 = .94$), reflecting a higher rating on the affectively congruent ($M = 7.28$, $SD = 0.73$) than the incongruent trials ($M = 2.48$, $SD = 0.61$). No other main effect or interaction was found (ps > .25). These results confirmed that our manipulation of congruency was valid, and the congruency ratings of the prime-target pairs (and thus the task difficulties) did not differ across the three timbre versions.

Procedure

There were six experimental conditions: affectively congruent and incongruent conditions for the voice version, affectively congruent and incongruent conditions for the violin version, and affectively congruent and incongruent conditions for the flute version. There were 360 trials in total, with 60 trials in each condition. To ensure that each stimulus only appeared once for each participant, two lists were created using a Latin square design, where each melody and emotional face was presented in either the congruent or incongruent condition within each list. Each list thus consisted of 180 trials, with 30 trials in each condition. The trials were presented in pseudo-randomized order in each list. During the experiment, the two lists were equally distributed across the 28 participants.

Each trial started with a black fixation in the middle of the screen with a white background. After 1000 ms, the prime was presented binaurally through Philips SHM1900 headphones. After the presentation of the prime, the target appeared on the screen for 1000 ms. Following the disappearance of the target, the response interface appeared on the screen. Such a design would avoid any contamination from artefacts associated with the action of button-pressing. Participants were instructed to judge whether the prime-target pairs were affectively congruent or not by pressing one of the two response buttons. The association between the hand side (left or right) and the response (congruent or incongruent) was counterbalanced across the participants. Before the formal experiment, six practice trials were given to familiarize the participants with the stimuli and procedure. In order to check whether there would be possible effects of familiarity with the stimuli on the results, following the EEG experiment, participants were asked to report if they had heard any of the melodies before. None of the participants reported being familiar with any of the melodies.

EEG recording and preprocessing

The EEG was recorded from 64 Ag/AgCl electrodes organized according to the international 10/20 system, referenced to the left mastoid. The electrode in front of the Fz served as ground. Vertical and horizontal electrooculograms (EOGs) were recorded by placing electrodes supra- and infraorbitally at the left eye and at the outer canthi of both eyes respectively. Impedances of all electrodes were kept below 5 kΩ during recording. The sampling rate was 500 Hz, with a band-pass filter of 0.05-100 Hz.

During preprocessing, EEG was re-referenced to the average of bipolar mastoid, and eye movements were corrected using the NeuroScan software 4.4 (Semlitsch, Anderer, Schuster, & Presslich, 1986). A band-pass filter of 0.1-30 Hz (24-dB/oct slope) was applied offline. EEG epochs from -200 to 1000 ms relative to the target onset were time-locked and baseline corrected (-200–0 ms). Trials with voltage amplitudes more than ±80 μV were treated as artifacts and rejected. Following previous studies using the affective priming paradigm (e.g., Hinojosa et al., 2009; Werheid, Alpay, Jentzsch, & Sommer, 2005), we excluded the trials with incorrect responses (less than 20%). On average, 29.02% of the trials were rejected, and 21 trials (*SD* = 1) were retained per condition.

ERP Data analysis

All ERP analyses were based on the mean amplitude values of each participant in each condition. Based on visual inspection and previous studies (e.g., Daltrozzo & Schön, 2008; Herring et al., 2011; Hinojosa et al., 2009; Steinbeis & Koelsch, 2009), the time windows of 280–440 ms and 500–600 ms after the onset of target stimulus were used for statistical analysis.

Although emotion type (happiness vs. sadness) was considered as a factor and included in the behavioral analysis, this factor was excluded from the ERP analysis due to the small number of trials in each condition. In our design, there were only 15 trials per condition if each timbre was divided into happy and sad emotions. To ensure that the number of trials was enough to get reliable results (Luck, 2005), we excluded emotion type from the current ERP analysis. Repeated measures ANOVAs were conducted for the midline and lateral electrodes separately (the selected electrodes are shown in Figure 2). For the midline electrodes, congruency (congruent vs. incongruent), timbre (violin, voice, and flute), and anteriority (anterior, central, and posterior) were considered as within-subjects factors. For the lateral electrodes, hemisphere (left vs. right) was added as an additional within-subjects factor. The mean of the respective electrodes in each region of interest was computed for analysis. The ANOVAs were followed by simple effects tests if there were any significant interactions, and all pairwise comparisons were adjusted by Bonferroni correction. Greenhouse–Geisser correction was applied when the degree of freedom in the numerator was greater than 1, and in these cases, the original degrees of freedom with corrected *p* values were reported. Only the significant effects containing the main experimental variables (congruency and timbre) are reported.

Insert Figure 2, about here.

Results

Behavioral results

To avoid response bias, sensitivity (*d'*) from signal-detection theory was used to measure judgment scores of each participant (Macmillan & Creelman, 2004). Higher values of *d'* represent better judgment. A hit was defined when a congruent pair was judged as congruent, and a false alarm was defined when an incongruent pair was judged as congruent. The log-linear rule was used for corrections of extreme proportions to avoid the biasing effect on the values of *d'* (Hautus, 1995).

A two-way repeated measures ANOVA taking emotion type and timbre version as within-subjects factors was conducted. As shown in Figure 3, the results showed a main effect of timbre ($F_{(2, 54)} = 4.62$, $p = .01$, $\eta_p^2 = .15$), as *d'* values for the violin and flute versions were significantly higher than the voice version, [violin (2.13 ± 0.73) > voice (1.71 ± 0.93), $p < .05$; flute (2.18 ± 0.67) > voice (1.71 ± 0.93), $p < .05$]. No significant difference was found between the violin and flute versions ($p > .10$). There was an interaction between timbre and emotion type ($F_{(2, 54)} = 8.70$, $p = .001$, $\eta_p^2 = .24$), with a higher *d'* for the happy voice version than the sad voice version ($F_{(1, 27)} = 11.82$, $p = .002$, $\eta_p^2 = .30$). However, no difference between happy and sad emotions was found for the violin ($F_{(1, 27)} = 0.001$, $p > .10$) or flute version ($F_{(1, 27)} = 0.004$, $p > .10$). We also computed the accuracy of performance, and the results showed the same pattern as above. These behavioral results indicate the effect of timbre on the processing of musical emotion.

Insert Figure 3, about here.

Electrophysiological results

As stated above, emotion type was excluded for the ERP analysis due to the small number of trials in each condition. Figure 4 shows the grand average waveforms elicited by affectively congruent and incongruent emotional faces for the violin, voice, and flute version, respectively. Figure 5 shows the scalp distribution of incongruent-minus-congruent difference waves for the three versions. As can be seen for the violin version, a larger negativity was induced by the incongruent than congruent trials in the time window of 280–440 ms. For the voice version, two larger positivities were induced by the incongruent than congruent trials in the time windows of 280–320 ms and 500–600 ms, respectively. However, only a larger late positivity was induced by the incongruent than congruent trials in the time window of 500–600 ms for the flute version.

280–440 ms time window. The results showed significant interactions between timbre and congruency on both midline ($F_{(1.40,\,37.66)} = 5.85$; $p = .01$, $\eta_p^2 = .18$) and lateral electrodes ($F_{(2,\,54)} = 6.18$; $p = .004$, $\eta_p^2 = .19$), owing to a larger negativity elicited by the affectively incongruent than congruent trials in the violin version (midline: $F_{(1,\,27)} = 7.82$; $p = .01$, $\eta_p^2 = .23$; lateral: $F_{(1,\,27)} = 11.62$; $p = .002$, $\eta_p^2 = .30$), but not in the voice (midline: $F_{(1,\,27)} = 2.66$; $p = .11$; lateral: $F_{(1,\,27)} = 1.68$; $p = .21$) or flute (midline: $F_{(1,\,27)} = 3.54$; $p = .07$; lateral: $F_{(1,\,27)} = 3.67$; $p = .07$) version.

Based on visual inspection, there might be a positive ERP effect for the voice version during 280–320 ms. We therefore preformed data analysis in sliding windows over 280–440 ms with a length of 40 ms (four time windows in total: 280–320 ms, 320–360 ms, 360–400 ms and 400–440 ms). The results showed significant interactions between timbre and congruency on both lateral (*ps* < .05) and midline electrodes (*ps* < .05) at each time window, owing to larger negativities in the affectively incongruent than congruent trials for the violin version in all time windows (*ps* < .05). In the voice version, a larger positivity in the affectively incongruent than congruent trials was observed only in the time window of 280–320 ms (midline: $F_{(1, 27)} = 4.66$; $p = .04$, $\eta_p^2 = .15$; lateral: $F_{(1, 27)} = 5.42$; $p = .03$, $\eta_p^2 = .17$). In the flute version, however, no significant difference was found between incongruent and congruent trials across all the sliding windows (*ps* > .05).

500–600 ms time window. The results revealed a significant main effect of congruency ($F_{(1, 27)} = 4.33$; $p = .047$, $\eta_p^2 = .14$), as the affectively incongruent trials elicited a larger LPC than the congruent trials. The effect of timbre was also significant on midline electrodes ($F_{(1, 27)} = 3.32$; $p = .04$, $\eta_p^2 = .11$), although pair-wise comparisons did not reveal any significant differences between the three timbre versions (*p*s > .05). Significant interactions between congruency and timbre were observed on both midline ($F_{(2, 54)} = 4.62$; $p = .01$, $\eta_p^2 = .15$) and lateral electrodes ($F_{(2, 54)} = 3.56$; $p = .04$, $\eta_p^2 = .12$), as a larger LPC was elicited by the affectively incongruent than congruent trials for the voice (midline: $F_{(1, 27)} = 4.26$; $p = .049$, $\eta_p^2 = .14$; lateral: $F_{(1, 27)} = 4.60$; $p = .04$, $\eta_p^2 = .15$) and flute versions (midline: $F_{(1, 27)} = 6.38$; $p = .02$, $\eta_p^2 = .19$; lateral: $F_{(1, 27)} = 3.76$; $p = .06$, $\eta_p^2 = .12$), but not for the violin version (midline: $F_{(1, 27)} = 2.59$; $p = .12$; lateral: $F_{(1, 27)} = 2.00$; $p = .17$). Furthermore, there was a significant interaction among congruency, anteriority, and hemisphere ($F_{(2, 54)} = 4.77$; $p = .01$, $\eta_p^2 = .15$), as there was a slight left-anterior weighting for the LPC effect. No other significant interactions containing the main experimental variables (congruency, timbre) were observed (*p*s > .05).

Although the interactions between timbre and anteriority or hemisphere were not significant in the aforementioned repeated measures ANOVAs, there were theoretical motivations to examine the neural responses to congruent and incongruent trials within each of the three timbres (Maxwell & Delaney, 2003). As mentioned before, given the effects of timbre on the processing of musical emotion at the behavioral level (e. g., Balkwill & Thompson, 1999; Eerola et al., 2012; Hailstone et al., 2009) and the differences in neural activities during the perception of vocal and non-vocal stimuli (Belin et al., 2000; Bruneau et al., 2013; Capilla et al., 2012), it would be expected that the voice and flute have different effects on neural responses to musical emotion. Therefore, a 2 congruency (congruent vs. incongruent) × 3 anteriority (anterior, central, and posterior) repeated measures ANOVA for the midline and a 2 congruency (congruent vs. incongruent) × 3 anteriority (anterior, central, and posterior) × 2 hemisphere (left vs. right) repeated measures ANOVA for the lateral electrodes were conducted for the voice and flute versions, separately. For the voice version, only a significant interaction between congruency, anteriority and hemisphere was found ($F_{(2, 54)} = 3.90$; $p = .03$, $\eta_p^2 = .13$), as there was a left-anterior weighting for the LPC effect ($F_{(1, 27)} = 7.33$; $p = .01$, $\eta_p^2 = .21$). For the flute version, only a significant main effect of anteriority was found ($F_{(2, 54)} = 7.24$; $p = .002$, $\eta_p^2 = .21$), with the largest amplitudes of LPC in central parietal sites.

Insert Figures 4 and 5, about here.

Discussion

Using ERPs and the cross-modal affective priming paradigm, we investigated the effects of timbre on neural responses to musical emotion. Like the behavioral results, our ERP data showed the effect of timbre on musical emotion processing. For the voice version, we found a larger P3 in the time window of 280–320 ms and a larger left anterior distributed LPC in 500–600 ms in response to the incongruent than congruent trials. For the flute version, however, only the LPC effect was found, which was distributed over centro-parietal electrodes. Unlike the voice and flute versions, a larger N400 in 280–440 ms was elicited in response to the incongruent than congruent trials in the violin version. These findings suggest that timbre influences the neural responses to musical emotion.

The main finding of this study was that there were distinct neural responses to musical emotion, when the same melodies were presented in different timbres. For the voice version, a larger P3 and LPC were elicited by incongruent than congruent trials. Such a P3 reflects that more attentional demands were needed to integrate the affectively incongruent information between vocal music and face, given that the P3 reflects attention allocation (Abrahamse, Duthoo, Notebaert, & Risko, 2013; Polich & Kok, 1995). On the other hand, the P3 is usually followed by the LPC, which reflects sustained attention (Foti, Hajcak, & Dien, 2009; Kujawa, Weinberg, Hajcak, & Klein, 2013; Weinberg & Hajcak, 2011). It has also been suggested that the frontal LPC in response to emotional stimuli reflects an improvement in controlled attentional engagement (Leutgeb et al., 2012). Taken together, the present LPC in the voice version might reflect the sustained, controlled attentional allocation which was needed to integrate the affectively incongruent information between vocal music and face.

It is worth noting that the VSR (voice-specific response) has been considered as an ERP signature of vocal discrimination (Levy et al., 2001; Levy, Granot, & Bentin, 2003). The latency of the present P3 is highly consistent with the VSR, both of which peaked at around 300 ms. Furthermore, similar to the P3, the VSR also reflects the allocation of attention in vocal stimuli (Levy et al., 2003). Therefore, our study confirmed the hypothesis that there are specialized neural responses to the human voice (Belin et al., 2004; Charest et al., 2009; Levy et al., 2001), including the neural processing of musical emotion in the human voice.

Similar to the voice version, there was also an LPC effect in the flute version. However, unlike the voice version, the scalp distribution of this effect was distributed over centro-parietal electrodes. It has been suggested that the centro-parietal LPC reflects increased attentional involvement activated by affectively incongruent stimuli (Herring et al., 2011; Hinojosa et al., 2009; Zhang et al., 2012). In this case, the LPC effect elicited by the flute version might indicate the enhanced attention induced by affectively incongruent trials, which was different from the left anterior distributed LPC in the voice version.

Unlike the voice and flute versions, however, a larger N400 was elicited in response to affectively incongruent versus congruent trials in the violin version, which reflects the activation of representations of affective meanings in the affective priming paradigm (Daltrozzo & Schön, 2008; Eder, Leuthold, Rothermund, & Schweinberger, 2011; Goerlich et al., 2012). That is, the N400 effect was an indication that the primes activated the representations of affectively related targets in the present study. Alternatively, given that the amplitude of the N400 is correlated with the difficulty in affective integration between the primes and targets (Kamiyama et al., 2013; Zhang et al., 2010), the N400 effect observed in the present study might reflect the difficulty in affective integration for affectively incongruent trials in the violin version.

Overall, our results revealed distinct neural responses to musical emotion presented in the voice, violin and flute. Specifically, although the flute and voice share similarities in structure and sound production (Wolfe, 2018), the voice version elicited a P3 and a left anterior distributed LPC, whereas the flute version only elicited a centro-parietal distributed LPC effect. Such a difference may be attributed to listeners' familiarity with the human voice. Indeed, familiarity with the stimuli affects attentional processing (Calvo & Eysenck, 2008; Griffiths, Brockmark, Höjesjö, & Johnsson, 2004), and it has also been suggested that familiar human voices elicited more attentional engagement than unfamiliar human voices (Beauchemin et al., 2006). This may account for the attentional engagement from the relatively early to the later processing stage in the voice version. Alternatively, the differences in neural responses between the voice and flute versions may be interpreted from the evolutionary perspective. It has been assumed that the human voice has an evolutionary significance (Andics, Gάcsi, Faragό, Kis, & Miklόsi, 2014; Grossmann, Oberecker, Koch, & Friederici, 2010; Petkov et al., 2008). Indeed, the processing of information contained in vocalizations from conspecific individuals is crucial for making decisions in behavioral contexts, such as territory disputes, mate choice, or hierarchy-related challenges (Owings & Morton, 1998).

On the other hand, even for the instrumental timbres, violin and flute exhibited different patterns of neural activities. The violin version elicited an N400 effect at the time window of 280–440 ms, whereas the flute version elicited an LPC at the time window of 500–600 ms. Such a difference might be attributed to the differences in acoustic features between the violin and flute. Indeed, previous studies have shown that neural responses to brighter (Toiviainen et al., 1998) and rougher sounds exhibited a shorter latency than sounds that were less bright and less rough (De Baene, Vandierendonck, Leman, Widmann, & Tervaniemi, 2004). In the present study, the violin version was brighter (violin: $M = .45$, $SD = .07$; flute: $M = 22$, $SD = .04$) and rougher than the flute (violin: $M = 224.26$, $SD = 147.17$; flute: $M = 74.96$, $SD = 105.97$) (see Supplementary Tables 1 and 2 for details). These differences might account for the distinct neural responses between the violin and flute versions.

Finally, we observed an effect of emotion type (happiness vs. sadness) on participants' behavioral performance in the voice version, but not in the violin or flute version. That is, participants achieved a higher *d'* in the happy voice condition than the sad voice condition. Such a difference might be due to the fact that the human voice is best at expressing sadness (Behrens & Green, 1993), which may make it easier for listeners to perceive sadness in a voice than any other emotion types. Given that listeners have a tendency to seek positive emotions after perceiving negative emotions (Erber & Erber, 1994), the subsequent sad events are likely to be perceived as less sad than they would normally be. Applying this possibility to our affective priming paradigm in the present study, owing to the salient sadness expression in the human voice, our participants might have judged the congruent sad music-face pairs as less congruent (and thus led to worse performance) compared to the congruent happy music-face pairs. However, future research is needed to confirm this possibility. In addition, due to the limited number of trials in our EEG experiment, we were unable to examine the effect of emotion type on neural processing of musical emotion across different timbres. Future studies are required to investigate this question further.

In conclusion, the present findings revealed different patterns of neural responses to emotional processing of music, when the same melodies were presented in different timbres: the voice, violin, and flute. Our findings confirmed the hypothesis that there are specialized neural responses to the human voice. Moreover, our findings also provided insights into the neural correlates underlying the processing of musical emotion, and how timbre played a role in this processing.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 31470972 to C. J. and F. L., and Grant No. 31500876 to L. Z.), and the European Research Council Starting Grant to F. L. and C. J. (CAASD, No. 678733). We wish to thank Dr. Carolyn Wu for her help with the revisions on an earlier version of this manuscript.

Declarations of interest

None.

References

Abrahamse, E. L., Duthoo, W., Notebaert, W., & Risko, E. F. (2013). Attention modulation by proportion congruency: The asymmetrical list shifting effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(5), 1552–1562.

Alluri, V., & Toiviainen, P. (2010). Exploring perceptual and acoustical correlates of polyphonic timbre. *Music Perception, 27*(3), 223–242.

Andics, A., Gácsi, M., Faragó, T., Kis, A., & Miklósi, Á. (2014). Voice-sensitive regions in the dog and human brain are revealed by comparative fMRI. *Current Biology, 24*(5), 574–578.

Aramaki, M., Besson, M., Kronland-Martinet, R., & Ystad, S. (2008). Timbre perception of sounds from impacted materials: Behavioral, electrophysiological and acoustic approaches. In S. Ystad, R. Kronland-Martinet, & K. Jensen (Eds.), *Lecture Notes in Computer Science: Vol. 5493*. Computer Music Modeling and Retrieval. Genesis of Meaning in Sound and Music (pp. 1–17). https://doi.org/10.1007/978-3-642-02518-1_1

Balkwill, L. L., & Thompson, W. F. (1999). A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. *Music Perception, 17*(1), 43–64.

Barthet, M., Depalle, P., Kronland-Martinet, R., & Ystad, S. (2010). Acoustical correlates of timbre and expressiveness in clarinet performance. *Music Perception*, *28*(2), 135–154.

Beauchemin, M., De Beaumont, L., Vannasing, P., Turcotte, A., Arcand, C., Belin, P., & Lassonde, M. (2006). Electrophysiological markers of voice familiarity. *European Journal of Neuroscience, 23*(11), 3081–3086.

Behrens, G. A., & Green, S. B. (1993). The ability to identify emotional content of solo improvisations performed vocally and on three different instruments. *Psychology of Music, 21*(1), 20–33.

Belin, P., Fecteau, S., & Bedard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences, 8*(3), 129–135.

Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Cognitive Brain Research, 13*(1), 17–26.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature, 403*(6767), 309–311.

Bowman, C., & Yamauchi, T. (2016). Perceiving categorical emotion in sound: The role of timbre. *Psychomusicology: Music, Mind, and Brain, 26*(1), 15–25.

Bruneau, N., Roux, S., Cléry, H., Rogier, O., Bidet-Caulet, A., & Barthélémy, C. (2013). Early neurophysiological correlates of vocal versus non-vocal sound processing in adults. *Brain Research, 1528*, 20–27.

Calvo, M. G., & Eysenck, M. W. (2008). Affective significance enhances covert attention: Roles of anxiety and word familiarity. *Quarterly Journal of Experimental Psychology, 61*(11), 1669–1686.

Capilla, A., Belin, P., & Gross, J. (2012). The early spatio-temporal correlates and task independence of cerebral voice processing studied with MEG. *Cerebral Cortex, 23*(6), 1388–1395.

Charest, I., Pernet, C. R., Rousselet, G. A., Quiñones, I., Latinus, M., Fillion-Bilodeau, S., Chartrand, J. P., & Belin, P. (2009). Electrophysiological evidence for an early processing of human voices. *BMC Neuroscience, 10*(1), 127. https://doi.org/10.1186/1471-2202-10-127

Crummer, G. C., Walton, J. P., Wayman, J. W., Hantz, E. C., & Frisina, R. D. (1994). Neural processing of musical timbre by musicians, nonmusicians, and musicians possessing absolute pitch. *The Journal of the Acoustical Society of America, 95*(5), 2720–2727.

Daltrozzo, J., & Schön, D. (2008). Conceptual processing in music as revealed by N400 effects on words and musical targets. *Journal of Cognitive Neuroscience, 21*(10), 1882–1892.

De Baene, W., Vandierendonck, A., Leman, M., Widmann, A., & Tervaniemi, M. (2004). Roughness perception in sounds: Behavioral and ERP evidence. *Biological Psychology, 67*(3), 319–330.

Eder, A. B., Leuthold, H., Rothermund, K., & Schweinberger, S. R. (2011). Automatic response activation in sequential affective priming: An ERP study. *Social Cognitive and Affective Neuroscience, 7*(4), 436–445.

Eerola, T., Ferrer, R., & Alluri, V. (2012). Timbre and affect dimensions: Evidence from affect and similarity ratings and acoustic correlates of isolated instrument sounds. *Music Perception, 30*(1), 49–70.

Eerola, T., Friberg, A., & Bresin, R. (2013). Emotional expression in music: Contribution, linearity, and additivity of primary musical cues. *Frontiers in Psychology, 4*(487). https://doi.org/10.3389/fpsyg.2013.00487

Erber, R., & Erber, M. W. (1994). Beyond mood and social judgment: Mood incongruent recall and mood regulation. *European Journal of Social Psychology, 24*(1), 79–88.

Foti, D., Hajcak, G., & Dien, J. (2009). Differentiating neural responses to emotional pictures: Evidence from temporal ‐ spatial PCA. *Psychophysiology, 46*(3), 521–530.

Gabrielsson, A., & Juslin, P. N. (1996). Emotional expression in music performance: Between the performer's intention and the listener's experience. *Psychology of Music, 24*(1), 68–91.

Goerlich, K. S., Witteman, J., Schiller, N. O., Van Heuven, V. J., Aleman, A., & Martens, S. (2012). The nature of affective priming in music and speech. *Journal of Cognitive Neuroscience, 24*(8), 1725–1741.

Gong, X., Huang, Y., Wang, Y., & Luo, Y. (2011). Revision of the Chinese facial affective picture system. *Chinese Mental Health Journal, 25*(1), 40–46.

Griffiths, S. W., Brockmark, S., Höjesjö, J., & Johnsson, J. I. (2004). Coping with divided attention: The advantage of familiarity. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 271*(1540), 695–699.

Griffiths, T. D., & Warren, J. D. (2004). What is an auditory object? *Nature Reviews Neuroscience, 5*(11), 887–892.

Grossmann, T., Oberecker, R., Koch, S. P., & Friederici, A. D. (2010). The developmental origins of voice processing in the human brain. *Neuron, 65*(6), 852–858.

Hailstone, J. C., Omar, R., Henley, S. M., Frost, C., Kenward, M. G., & Warren, J. D. (2009). It's not what you play, it's how you play it: Timbre affects perception of emotion in music. *Quarterly Journal of Experimental Psychology, 62*(11), 2141–2155.

Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d'. *Behavior Research Methods, 27*(1), 46–51.

Hermans, D., De Hower, J., & Eelen, P. (2001). A time course analysis of the affective priming effect. *Cognition and Emotion, 15*(2), 143–165.

Herring, D. R., Taylor, J. H., White, K. R., & Crites, S. L. (2011). Electrophysiological responses to evaluative priming: The LPP is sensitive to incongruity. *Emotion, 11*(4), 794–806.

Hinojosa, J. A., Carretié, L., Méndezbértolo, C., Míguez, A., & Pozo, M. A. (2009). Arousal contributions to affective priming: Electrophysiological correlates. *Emotion, 9*(2), 164–171.

Kamiyama, K. S., Abla, D., Iwanaga, K., & Okanoya, K. (2013). Interaction between musical emotion and facial expression as measured by event-related potentials. *Neuropsychologia, 51*(3), 500–505.

Kujawa, A., Weinberg, A., Hajcak, G., & Klein, D. N. (2013). Differentiating event‐related potential components sensitive to emotion in middle childhood: Evidence from temporal-spatial PCA. *Developmental Psychobiology, 55*(5), 539–550.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology, 62*(1), 621–647.

Kutas, M., & Hillyard, S. A. (1980). Reading between the lines: Event-related brain potentials during natural sentence processing. *Brain and Language, 11*(2), 354–373.

Lense, M. D., Gordon, R. L., Key, A. P., & Dykens, E. M. (2012, July). *Neural correlates of musical timbre perception in Williams syndrome*. Paper presented at the Twelfth International Conference on Music Perception and Cognition, Thessaloniki, Greece.

Leutgeb, V., Schäfer, A., Köchel, A., & Schienle, A. (2012). Exposure therapy leads to enhanced late frontal positivity in 8-to 13-year-old spider phobic girls. *Biological Psychology, 90*(1), 97–104.

Levy, D. A., Granot, R., & Bentin, S. (2001). Processing specificity for human voice stimuli: Electrophysiological evidence. *NeuroReport, 12*(12), 2653–2657.

Levy, D. A., Granot, R., & Bentin, S. (2003). Neural sensitivity to human voices: ERP evidence of task and attentional influences. *Psychophysiology, 40*(2), 291–305.

Logeswaran, N., & Bhattacharya, J. (2009). Crossmodal transfer of emotion by music. *Neuroscience Letters, 455*(2), 129–133.

Luck, S. J. (2005). Ten simple rules for designing and interpreting ERP experiments. In T. C. Handy (Eds.), *Event-related potentials: A methods handbook* (pp. 17–32). Cambridge, MA: The MIT Press.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Maxwell, S. E., & Delaney, H. D. (2003). *Designing experiments and analyzing data: A model comparison perspective*. New York, NY: Routledge.

McAdams, S., Cunible, J., Carlyon, R. P., Darwin, C. J., & Russell, I. J. (1992). Perception of timbral analogies. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 336*(1278), 383–389.

Menon, V., Levitin, D. J., Smith, B. K., Lembke, A., Krasnow, B. D., Glazer, D., Glover, G. H., & McAdams, S. (2002). Neural correlates of timbre change in harmonic sounds. *NeuroImage, 17*(4), 1742–1754.

Owings, D. H., & Morton, E. S. (1998). *Animal vocal communication: A new approach*. Cambridge, MA: Cambridge University Press.

Paquette, S., Peretz, I., & Belin, P. (2013). The “Musical Emotional Bursts”: A validated set of musical affect bursts to investigate auditory affective processing. *Frontiers in Psychology, 4*(509). https://doi.org/10.3389/fpsyg.2013.00509

Petkov, C. I., Kayser, C., Steudel, T., Whittingstall, K., Augath, M., & Logothetis, N. K. (2008). A voice region in the monkey brain. *Nature Neuroscience, 11*(3), 367–374.

Polich, J., & Kok, A. (1995). Cognitive and biological determinants of P300: An integrative review. *Biological Psychology, 41*(2), 103–146.

Schirmer, A., Kotz, S. A., & Friederici, A. D. (2002). Sex differentiates the role of emotional prosody during word processing. *Cognitive Brain Research, 14*(2), 228–233.

Semlitsch, H. V., Anderer, P., Schuster, P., & Presslich, O. (1986). A solution for reliable and valid reduction of ocular artifacts, applied to the P300 ERP. *Psychophysiology, 23*(6), 695–703.

Sollberge, B., Rebe, R., & Eckstein, D. (2003). Musical chords as affective priming context in a word-evaluation task. *Music Perception*, *20*(3), 263–282.

Steinbeis, N., & Koelsch, S. (2009). Affective priming effects of musical sounds on the processing of word meaning. *Journal of Cognitive Neuroscience, 23*(3), 604–621.

Timmers, R., & Crook, H. (2014). Affective priming in music listening: Emotions as a source of musical expectation. *Music Perception*, *31*(5), 470–484.

Toiviainen, P., Tervaniemi, M., Louhivuori, J., Saher, M., Huotilainen, M., & Näätänen, R. (1998). Timbre similarity: Convergence of neural, behavioral, and computational approaches. *Music Perception*, *16*(2), 223–241.

Weinberg, A., & Hajcak, G. (2011). The late positive potential predicts subsequent interference with target processing. *Journal of Cognitive Neuroscience, 23*(10), 2994–3007.

Werheid, K., Alpay, G., Jentzsch, I., & Sommer, W. (2005). Priming emotional facial expressions as evidenced by event-related brain potentials. *International Journal of Psychophysiology, 55*(2), 209–219.

Wolfe, J. (2018). The acoustics of woodwind musical instruments. *Acoustics Today, 14*, 50–56.

Wolfe, J., Garnier, M., & Smith, J. (2009). Vocal tract resonances in speech, singing, and playing musical instruments. *Human Frontier Science Program, 3*(1), 6–23.

Zhang, Q., Kong, L., & Yang, J. (2012). The interaction of arousal and valence in affective priming: Behavioral and electrophysiological evidence. *Brain Research, 1474*, 60–72.

Zhang, Q., Lawson, A., Guo, C., & Jiang, Y. (2006). Electrophysiological correlates of visual affective priming. *Brain Research Bulletin, 71*(3), 316–323.

Zhang, Q., Li, X., Gold, B. T., & Jiang, Y. (2010). Neural correlates of cross-domain affective priming. *Brain Research, 1329*, 142–151.

Figure Captions

Figure 1. Design of the cross-modal affective priming paradigm. Musical melodies were used as primes, and emotional faces as targets. Facial images were affectively congruent or incongruent with the prime melodies.

Figure 2. Electrode layout on the scalp. Six regions were selected for statistical analysis of lateral electrodes: left and right anterior, left and right central, and left and right posterior.

Figure 3. Values of *d*' under each condition. The error bars refer to the standard errors.

Figure 4. Grand mean ERP waveforms elicited by affectively congruent and incongruent facial images preceded by melodies presented in the violin version, the voice version, and the flute version. Gray-shaded areas indicate the time windows used for statistical analysis.

Figure 5. Scalp distribution of the affectively incongruent-minus-congruent difference waves in the 280–320 ms, 280–440 ms and 500–600 ms time windows for the violin version, the voice version, and the flute version.

Figure 1

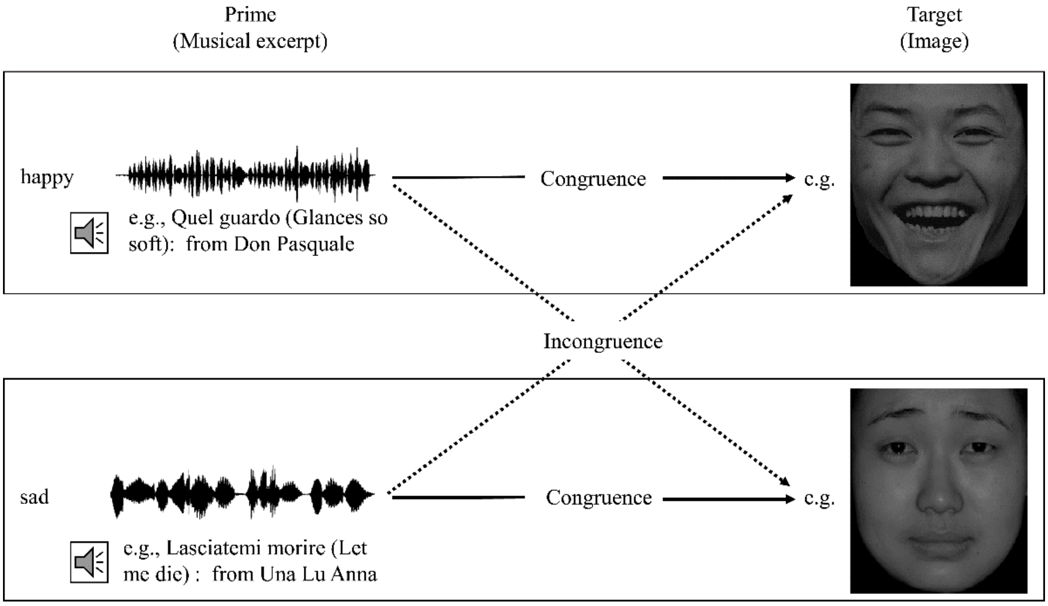
Prime
(Musical excerpt)
Target
(Image)
happy
e.g., Quel guardo (Glances so soft): from Don Pasquale
Congruence
c.g.
Incongruence
sad
e.g., Lasciatemi morire (Let me die) : from Una Lu Anna
Congruence
c.g.

Figure 2

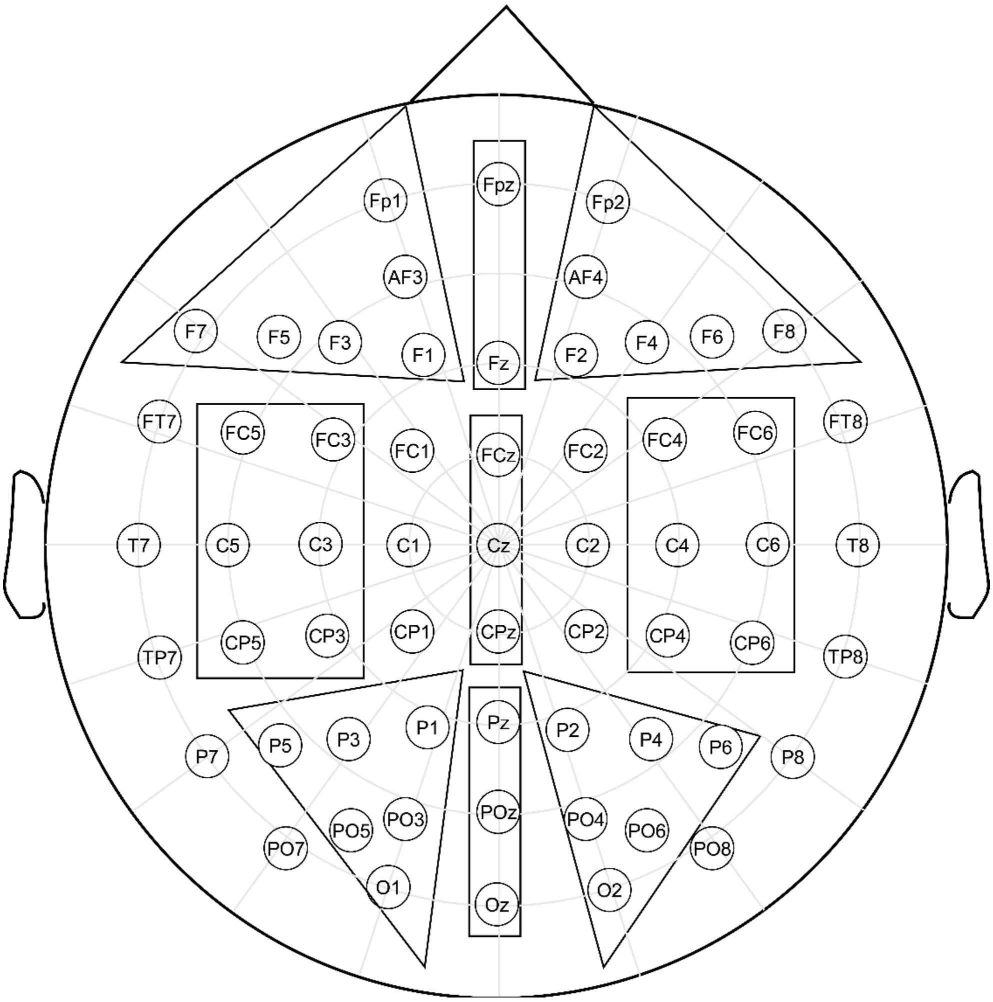
Fpz
Fp1
Fp2
AF3
AF4
F7
F5
F3
F1
Fz
F2
F4
F6
F8
FT7
FC5
FC3
FC1
FCz
FC2
FC4
FC6
FT8
T7
C5
C3
C1
Cz
C2
C4
C6
T8
TP7
CP5
CP3
CP1
CPz
CP2
CP4
CP6
TP8
P7
P5
P3
P1
Pz
P2
P4
P6
P8
PO7
PO5
PO3
POz
PO4
PO6
PO8
O1
Oz
O2

Figure 3

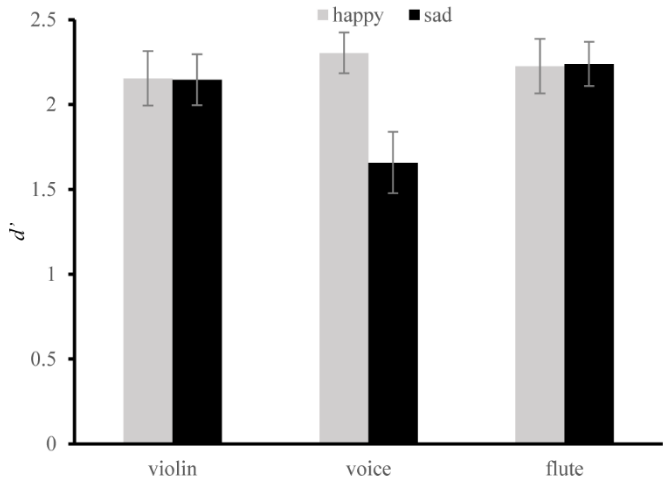

Figure 4

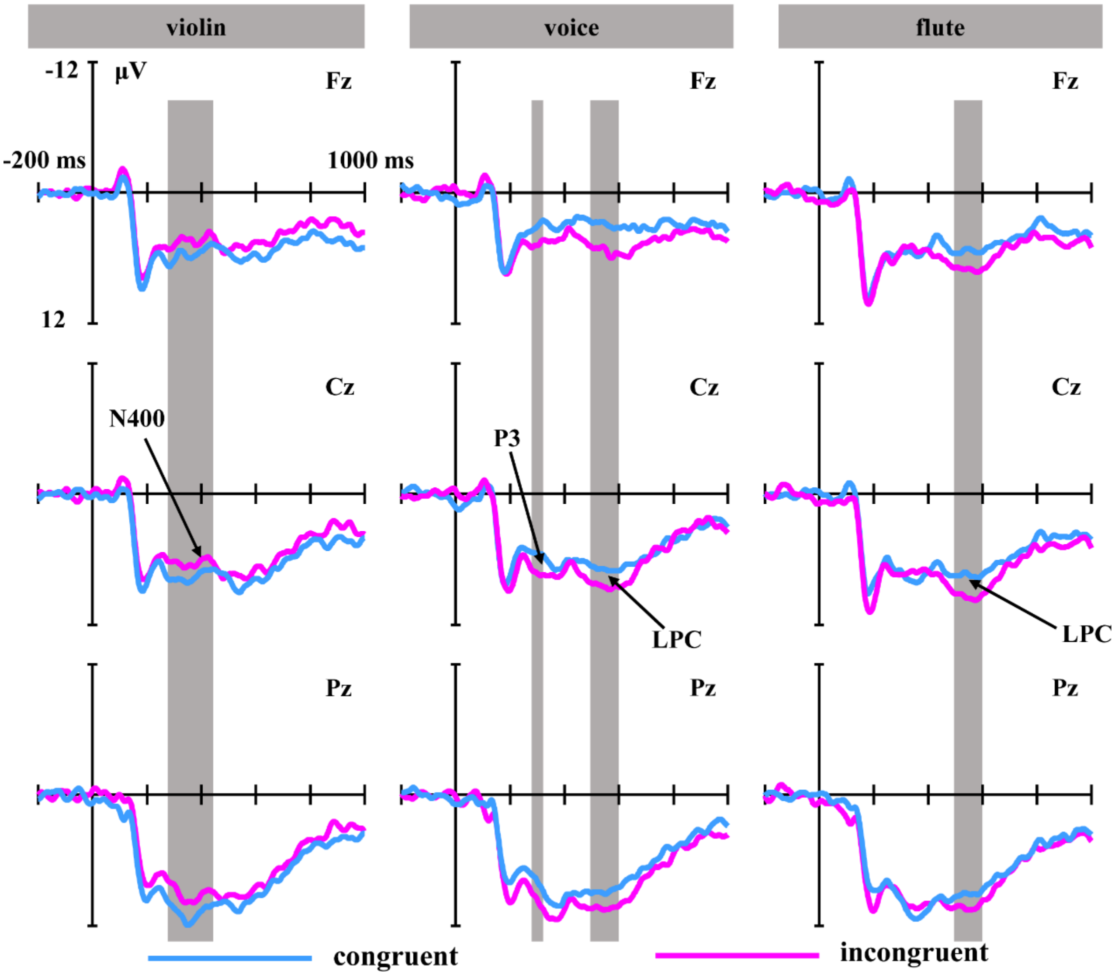
violin
voice
flute
-12
µV
12
-200 ms
1000 ms
Fz
Cz
Pz
N400
P3
LPC
LPC
congruent
incongruent

Figure 5