

# *The effects of timbre on neural responses to musical emotion*

Article

Accepted Version

Zhang, W., Liu, F., 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: <https://doi.org/10.1525/mp.2019.37.2.134> Available at <http://centaur.reading.ac.uk/85954/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

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

Publisher: University of California Press

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

## **CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

---

# 1 **The effects of timbre on neural responses to musical emotion**

2 Weixia Zhang<sup>a</sup>, Fang Liu<sup>b</sup>, Linshu Zhou<sup>c</sup>, Wanqi Wang<sup>c</sup>, Hanyuan Jiang<sup>d</sup>, and  
3 Cunmei Jiang<sup>c,e\*</sup>

4  
5 *<sup>a</sup> Department of Psychology, College of Education, Shanghai Normal University,*  
6 *Shanghai, China, 200234*

7 *<sup>b</sup> School of Psychology and Clinical Language Sciences, University of Reading, Reading,*  
8 *UK, RG6 6AL*

9 *<sup>c</sup> Music College, Shanghai Normal University, Shanghai, China, 200234*

10 *<sup>d</sup> Faculty of Humanities and Arts, Macau University of Science and Technology, Macau,*  
11 *China, 519020*

12 *<sup>e</sup> Institute of Psychology, Shanghai Normal University, Shanghai, China, 200234*

13  
14 Running head: Effects of timbre on musical emotion processing

15  
16 \* Corresponding author to:

17 Dr. Cunmei Jiang

18 Music College

19 Shanghai Normal University

20 100 E. Guilin Road

21 Shanghai, 200234, China

22 Tel: 0086-21-64322990; Fax: 0086-21-64322935

23 Electronic mail: [cunmeijiang@126.com](mailto:cunmeijiang@126.com)

24

# 1 Abstract

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

16

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

18

19

20

21

22

23

24

25

26

27

---

# 1 Introduction

2 Timbre is one of the most important acoustic attributes in our environment  
3 (Menon et al., 2002). Through timbre, listeners can distinguish two sounds of  
4 identical pitch, duration, and intensity (Griffiths & Warren, 2004; McAdams, Cunible,  
5 Carlyon, Darwin, & Russell, 1992). Among different timbres, the human voice has  
6 been shown to be associated with specialized neural activities (for a review see Belin,  
7 Fecteau, & Bédard, 2004). In particular, fMRI studies (Belin, Zatorre, & Ahad, 2002;  
8 Belin, Zatorre, Lafaille, Ahad, & Pike, 2000) have indicated that the superior  
9 temporal sulcus (STS) shows greater response to the human voice than to non-vocal  
10 stimuli. Electrophysiological evidence has also revealed a larger amplitude of  
11 voice-specific response (VSR) (Levy, Granot, & Bentin, 2001) and a larger  
12 fronto-temporal positivity to voice (FTPV) (Bruneau et al., 2013; Capilla, Belin, &  
13 Gross, 2012; Charest et al., 2009) in response to vocal compared with non-vocal  
14 stimuli such as environmental sounds. For non-vocal stimuli, it has also been  
15 suggested that the perception of different timbres involves different  
16 electrophysiological correlates (Aramaki, Besson, Kronland-Martinet, & Ystad, 2008;  
17 Crummer, Walton, Wayman, Hantz, & Frisina, 1994). For example, metal sounds  
18 elicited a smaller P200, larger N280 and negative slow wave than wood and glass  
19 sounds, whereas the latter of which did not differ from each other (Aramaki et al.,  
20 2008). In short, the aforementioned studies suggest different neural activities during  
21 the discrimination of sound stimuli of different timbre.

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

---

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

4 A few previous studies have examined neural responses to musical emotion by  
5 using a cross-modal affective priming paradigm. This paradigm is thought to be an  
6 appropriate method to examine the activation of affective representations (Hermans,  
7 De Houwer, & Eelen, 2001; Herring, Taylor, White, & Crites Jr, 2011). It has been  
8 widely employed in the investigations of musical emotion, where music excerpts are  
9 used to prime affectively congruent/incongruent words (Goerlich et al., 2012;  
10 Sollberge, Rebe, & Eckstein, 2003; Steinbeis & Koelsch, 2009) or pictures (Lense,  
11 Gordon, Key, & Dykens, 2012; Logeswaran & Bhattacharya, 2009). Therefore, unlike  
12 the semantic priming paradigm that focuses on the association of meaning between  
13 the primes and targets, the affective priming paradigm focuses on the association  
14 between the primes and targets in emotional features such as valence (Herring et al.,  
15 2011; Timmers & Crook, 2014).

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

29 Although previous research has suggested that the perception of vocal and  
30 non-vocal stimuli involves distinct neural processes, no study has yet examined

1 whether emotional processing of musical melodies that are presented in different  
2 timbres would also show neural differences. In the present study, we investigated the  
3 effects of timbre on neural responses to musical emotion using the affective priming  
4 paradigm, with a 2 congruency (congruent vs. incongruent)  $\times$  3 timbre (violin, voice,  
5 flute) within-subjects design. Target facial expressions were primed by affectively  
6 congruent or incongruent melodies, which were presented in three timbre versions:  
7 the voice, violin, and flute. Two pretests were conducted to ensure the validity of the  
8 stimuli. The first pretest was to balance the performance level (i.e., how good the  
9 performance was) and performance style (including rubato, intensity, and phrasing)  
10 among the three versions, while the second pretest was to assess whether the musical  
11 stimuli were affectively congruent or incongruent with the faces and to confirm the  
12 validity of the stimuli.

13 We chose the voice, flute, and violin as different timbre conditions based on the  
14 following considerations. First, although previous research has shown that perception  
15 of the human voice involves distinct neural correlates compared with non-vocal  
16 sounds (Belin et al., 2000; Charest et al., 2009; Levy et al., 2001), it remains unknown  
17 whether there are different neural responses to musical emotion when the same  
18 melodies are presented in the voice and in other timbres. The inclusion of the voice in  
19 the present study would allow us to address this issue and to test the hypothesis that  
20 there is specialized neural processing of musical emotion in the human voice. Second,  
21 we chose the flute as another timbre condition because it is not only representative of  
22 wind instruments in an orchestra, but also resembles the voice in terms of structure  
23 and sound production. That is, the resonances of the vocal tract generate the timbre  
24 for both the voice and flute (Wolfe, Garnier, & Smith, 2009). Therefore, by  
25 comparing the flute with the voice, we would reveal whether timbres created by  
26 similar energy sources would still lead to different neural activities during emotion  
27 processing. Finally, the violin is representative of string instruments in an orchestra,  
28 which is different from the flute and voice in structure and sound production.  
29 Therefore, the inclusion of the voice, flute, and violin would allow us to compare the  
30 effects of the vocal and two representative instrumental timbres on neural responses to

1 musical emotion. It is predicted that the neural processing of musical emotion may be  
2 different across the three timbre versions, given that timbre has shown significant  
3 effects on the processing of musical emotion at the behavioral level (e.g., Behrens &  
4 Green, 1993; Eerola et al., 2012; Hailstone et al., 2009).

## 5 **Experimental procedure**

### 6 **Participants**

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

### 13 **Stimuli**

14 Musical melodies and facial images were used as primes and targets, respectively.  
15 First, a total of 120 melodies were selected from European operas composed during  
16 classical or Romantic musical periods from around 1750–1900. Given that happy and  
17 sad melodies may reflect different musical scale (or interval) structures, we included  
18 60 happy and 60 sad melodic stimuli in order to avoid the possible interactions  
19 between timbre type and scale structure of the melodies. The selections were based on  
20 musicological analysis and self-reports from the composers (e.g., a melody associated  
21 with the lyrics of “crying” was intended to express the emotion of *sadness*).

22 In order to enhance the ecological validity of the musical stimuli, we used  
23 recordings of performances, following previous studies on musical emotion and  
24 timbre (Balkwill & Thompson, 1999; Behrens & Green, 1993; Hailstone et al., 2009).  
25 Each melody was played by a violinist and a flutist, as well as sung by a vocalist with  
26 the syllable “la”, resulting in 360 musical stimuli in three versions. The performers  
27 were all female, and had all received professional music training over 18 years. They



1 were informed that the three versions of each melody should be performed in a similar  
2 style with regard to rubato, intensity and phrasing, and they discussed how to perform  
3 each melody before recording. After recording, all musical stimuli were edited using  
4 Adobe Audition CS6 (Adobe Systems Inc) with 22.05 kHz sampling rate and 16-bit  
5 resolution. The mean duration of the melodies was on average 7 s, ranging from 2 to 9  
6 s. The loudness of the melodies was normalized to approximately 68 dB SPL, fading  
7 out in 1s.

8 A pretest was conducted in order to rule out the differences in performance  
9 levels or styles among the three timbre conditions. Eight musicians (all with more  
10 than 10 years of professional music training) rated the similarity among the three  
11 versions of each melody with regard to rubato, intensity, phrasing, and the overall  
12 performance level using a 7-point scale (1 = very incongruent, 4 = not sure, 7 = very  
13 congruent). Only melodies with a mean value above 4 (the minimum standard) were  
14 chosen as experimental stimuli, which led to the selection of 240 musical stimuli (i.e.,  
15 three versions of 80 melodies) as the potential prime stimuli. The results of the pretest  
16 showed that the three versions of the musical stimuli were performed in similar  
17 manners at similar performance levels (rubato:  $M = 5.02$ ,  $SD = 0.22$ ; intensity:  $M =$   
18  $5.19$ ,  $SD = 0.38$ ; phrase:  $M = 5.70$ ,  $SD = 0.49$ ; overall performance level:  $M = 5.25$ ,  
19  $SD = 0.33$ ).

20 240 emotional faces were selected as potential target stimuli from the Chinese  
21 Facial Affective Picture System (CFAPS) (Gong, Huang, Wang, & Luo, 2011), of  
22 which 120 expressed happiness and 120 expressed sadness. Each of the 240 musical  
23 stimuli was presented twice, followed by either an affectively congruent face or an  
24 incongruent face, thus resulting in 480 trials (see Figure 1 for examples).

25 -----  
26 Insert Figure 1, about here.  
27 -----

28 Another pretest was conducted to assess whether the musical stimuli were  
29 affectively congruent or incongruent with the faces. Twenty non-musicians who did  
30 not participate in the EEG experiment were asked to rate each prime-target pair

1 regarding affective congruency using a 9-point scale (1 = very incongruent, 5 = not  
2 sure, 9 = very congruent). We used the 9-point instead of 7-point scale to get more  
3 detailed rating information. In the end, 360 prime-target pairs from the extreme ends  
4 (congruent pairs: ratings  $\geq 6$ ; incongruent pairs: ratings  $\leq 4$ ) of the continuous  
5 distribution of the congruent/incongruent values were selected as the final  
6 experimental stimuli with the constraints that each prime and target would be used  
7 twice, and that all three versions of the same melody would be selected.

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

## 17 Procedure

18 There were six experimental conditions: affectively congruent and incongruent  
19 conditions for the voice version, affectively congruent and incongruent conditions for  
20 the violin version, and affectively congruent and incongruent conditions for the flute  
21 version. There were 360 trials in total, with 60 trials in each condition. To ensure that  
22 each stimulus only appeared once for each participant, two lists were created using a  
23 Latin square design, where each melody and emotional face was presented in either  
24 the congruent or incongruent condition within each list. Each list thus consisted of  
25 180 trials, with 30 trials in each condition. The trials were presented in  
26 pseudo-randomized order in each list. During the experiment, the two lists were  
27 equally distributed across the 28 participants.

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

## 15 **EEG recording and preprocessing**

16 The EEG was recorded from 64 Ag/AgCl electrodes organized according to the  
17 international 10/20 system, referenced to the left mastoid. The electrode in front of  
18 the Fz served as ground. Vertical and horizontal electrooculograms (EOGs) were  
19 recorded by placing electrodes supra- and infraorbitally at the left eye and at the  
20 outer canthi of both eyes respectively. Impedances of all electrodes were kept below  
21 5 k $\Omega$  during recording. The sampling rate was 500 Hz, with a band-pass filter of  
22 0.05-100 Hz.

23 During preprocessing, EEG was re-referenced to the average of bipolar mastoid,  
24 and eye movements were corrected using the NeuroScan software 4.4 (Semlitsch,  
25 Anderer, Schuster, & Presslich, 1986). A band-pass filter of 0.1-30 Hz (24-dB/oct  
26 slope) was applied offline. EEG epochs from -200 to 1000 ms relative to the target  
27 onset were time-locked and baseline corrected (-200–0 ms). Trials with voltage  
28 amplitudes more than  $\pm 80$   $\mu$ V were treated as artifacts and rejected. Following

1 previous studies using the affective priming paradigm (e.g., Hinojosa et al., 2009;  
2 Werheid, Alpay, Jentzsch, & Sommer, 2005), we excluded the trials with incorrect  
3 responses (less than 20%). On average, 29.02% of the trials were rejected, and 21  
4 trials ( $SD = 1$ ) were retained per condition.

## 5 **ERP Data analysis**

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

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

1 were reported. Only the significant effects containing the main experimental  
2 variables (congruency and timbre) are reported.

3  
4 -----  
5 Insert Figure 2, about here.  
6 -----  
7

## 8 Results

### 9 Behavioral results

10 To avoid response bias, sensitivity ( $d'$ ) from signal-detection theory was  
11 used to measure judgment scores of each participant (Macmillan & Creelman,  
12 2004). Higher values of  $d'$  represent better judgment. A hit was defined when a  
13 congruent pair was judged as congruent, and a false alarm was defined when an  
14 incongruent pair was judged as congruent. The log-linear rule was used for  
15 corrections of extreme proportions to avoid the biasing effect on the values of  $d'$   
16 (Hautus, 1995).

17 A two-way repeated measures ANOVA taking emotion type and timbre version  
18 as within-subjects factors was conducted. As shown in Figure 3, the results showed a  
19 main effect of timbre ( $F_{(2, 54)} = 4.62, p = .01, \eta_p^2 = .15$ ), as  $d'$  values for the violin and  
20 flute versions were significantly higher than the voice version, [violin ( $2.13 \pm 0.73$ ) >  
21 voice ( $1.71 \pm 0.93$ ),  $p < .05$ ; flute ( $2.18 \pm 0.67$ ) > voice ( $1.71 \pm 0.93$ ),  $p < .05$ ]. No  
22 significant difference was found between the violin and flute versions ( $p > .10$ ). There  
23 was an interaction between timbre and emotion type ( $F_{(2, 54)} = 8.70, p = .001, \eta_p^2$   
24  $= .24$ ), with a higher  $d'$  for the happy voice version than the sad voice version ( $F_{(1, 27)}$   
25  $= 11.82, p = .002, \eta_p^2 = .30$ ). However, no difference between happy and sad  
26 emotions was found for the violin ( $F_{(1, 27)} = 0.001, p > .10$ ) or flute version ( $F_{(1, 27)} =$   
27  $0.004, p > .10$ ). We also computed the accuracy of performance, and the results

1 showed the same pattern as above. These behavioral results indicate the effect of  
2 timbre on the processing of musical emotion.

3 -----  
4 Insert Figure 3, about here.  
5 -----  
6

## 7 **Electrophysiological results**

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

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

26 Based on visual inspection, there might be a positive ERP effect for the voice  
27 version during 280–320 ms. We therefore preformed data analysis in sliding windows  
28 over 280–440 ms with a length of 40 ms (four time windows in total: 280–320 ms,

1 320–360 ms, 360–400 ms and 400–440 ms). The results showed significant  
 2 interactions between timbre and congruency on both lateral ( $ps < .05$ ) and midline  
 3 electrodes ( $ps < .05$ ) at each time window, owing to larger negativities in the  
 4 affectively incongruent than congruent trials for the violin version in all time windows  
 5 ( $ps < .05$ ). In the voice version, a larger positivity in the affectively incongruent than  
 6 congruent trials was observed only in the time window of 280–320 ms (midline:  $F_{(1, 27)}$   
 7  $= 4.66$ ;  $p = .04$ ,  $\eta_p^2 = .15$ ; lateral:  $F_{(1, 27)} = 5.42$ ;  $p = .03$ ,  $\eta_p^2 = .17$ ). In the flute version,  
 8 however, no significant difference was found between incongruent and congruent  
 9 trials across all the sliding windows ( $ps > .05$ ).

10 **500–600 ms time window.** The results revealed a significant main effect of  
 11 congruency ( $F_{(1, 27)} = 4.33$ ;  $p = .047$ ,  $\eta_p^2 = .14$ ), as the affectively incongruent trials  
 12 elicited a larger LPC than the congruent trials. The effect of timbre was also  
 13 significant on midline electrodes ( $F_{(1, 27)} = 3.32$ ;  $p = .04$ ,  $\eta_p^2 = .11$ ), although pair-wise  
 14 comparisons did not reveal any significant differences between the three timbre  
 15 versions ( $ps > .05$ ). Significant interactions between congruency and timbre were  
 16 observed on both midline ( $F_{(2, 54)} = 4.62$ ;  $p = .01$ ,  $\eta_p^2 = .15$ ) and lateral electrodes ( $F_{(2,$   
 17  $54)} = 3.56$ ;  $p = .04$ ,  $\eta_p^2 = .12$ ), as a larger LPC was elicited by the affectively  
 18 incongruent than congruent trials for the voice (midline:  $F_{(1, 27)} = 4.26$ ;  $p = .049$ ,  $\eta_p^2$   
 19  $= .14$ ; lateral:  $F_{(1, 27)} = 4.60$ ;  $p = .04$ ,  $\eta_p^2 = .15$ ) and flute versions (midline:  $F_{(1, 27)} =$   
 20  $6.38$ ;  $p = .02$ ,  $\eta_p^2 = .19$ ; lateral:  $F_{(1, 27)} = 3.76$ ;  $p = .06$ ,  $\eta_p^2 = .12$ ), but not for the violin  
 21 version (midline:  $F_{(1, 27)} = 2.59$ ;  $p = .12$ ; lateral:  $F_{(1, 27)} = 2.00$ ;  $p = .17$ ). Furthermore,  
 22 there was a significant interaction among congruency, anteriority, and hemisphere ( $F_{(2,$   
 23  $54)} = 4.77$ ;  $p = .01$ ,  $\eta_p^2 = .15$ ), as there was a slight left-anterior weighting for the LPC  
 24 effect. No other significant interactions containing the main experimental variables  
 25 (congruency, timbre) were observed ( $ps > .05$ ).

26 Although the interactions between timbre and anteriority or hemisphere were not  
 27 significant in the aforementioned repeated measures ANOVAs, there were theoretical  
 28 motivations to examine the neural responses to congruent and incongruent trials  
 29 within each of the three timbres (Maxwell & Delaney, 2003). As mentioned before,  
 30 given the effects of timbre on the processing of musical emotion at the behavioral

1 level (e. g., Balkwill & Thompson, 1999; Eerola et al., 2012; Hailstone et al., 2009)  
2 and the differences in neural activities during the perception of vocal and non-vocal  
3 stimuli (Belin et al., 2000; Bruneau et al., 2013; Capilla et al., 2012), it would be  
4 expected that the voice and flute have different effects on neural responses to musical  
5 emotion. Therefore, a 2 congruency (congruent vs. incongruent)  $\times$  3 anteriority  
6 (anterior, central, and posterior) repeated measures ANOVA for the midline and a 2  
7 congruency (congruent vs. incongruent)  $\times$  3 anteriority (anterior, central, and  
8 posterior)  $\times$  2 hemisphere (left vs. right) repeated measures ANOVA for the lateral  
9 electrodes were conducted for the voice and flute versions, separately. For the voice  
10 version, only a significant interaction between congruency, anteriority and  
11 hemisphere was found ( $F_{(2, 54)} = 3.90$ ;  $p = .03$ ,  $\eta_p^2 = .13$ ), as there was a left-anterior  
12 weighting for the LPC effect ( $F_{(1, 27)} = 7.33$ ;  $p = .01$ ,  $\eta_p^2 = .21$ ). For the flute version,  
13 only a significant main effect of anteriority was found ( $F_{(2, 54)} = 7.24$ ;  $p = .002$ ,  $\eta_p^2$   
14 = .21), with the largest amplitudes of LPC in central parietal sites.

15 -----  
16 Insert Figures 4 and 5, about here.  
17 -----

## 18 Discussion

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



1 The main finding of this study was that there were distinct neural responses to  
2 musical emotion, when the same melodies were presented in different timbres. For the  
3 voice version, a larger P3 and LPC were elicited by incongruent than congruent trials.  
4 Such a P3 reflects that more attentional demands were needed to integrate the  
5 affectively incongruent information between vocal music and face, given that the P3  
6 reflects attention allocation (Abrahamse, Duthoo, Notebaert, & Risko, 2013; Polich &  
7 Kok, 1995). On the other hand, the P3 is usually followed by the LPC, which reflects  
8 sustained attention (Foti, Hajcak, & Dien, 2009; Kujawa, Weinberg, Hajcak, & Klein,  
9 2013; Weinberg & Hajcak, 2011). It has also been suggested that the frontal LPC in  
10 response to emotional stimuli reflects an improvement in controlled attentional  
11 engagement (Leutgeb et al., 2012). Taken together, the present LPC in the voice  
12 version might reflect the sustained, controlled attentional allocation which was needed  
13 to integrate the affectively incongruent information between vocal music and face.

14 It is worth noting that the VSR (voice-specific response) has been considered as  
15 an ERP signature of vocal discrimination (Levy et al., 2001; Levy, Granot, & Bentin,  
16 2003). The latency of the present P3 is highly consistent with the VSR, both of which  
17 peaked at around 300 ms. Furthermore, similar to the P3, the VSR also reflects the  
18 allocation of attention in vocal stimuli (Levy et al., 2003). Therefore, our study  
19 confirmed the hypothesis that there are specialized neural responses to the human  
20 voice (Belin et al., 2004; Charest et al., 2009; Levy et al., 2001), including the neural  
21 processing of musical emotion in the human voice.

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

1 Unlike the voice and flute versions, however, a larger N400 was elicited in  
2 response to affectively incongruent versus congruent trials in the violin version, which  
3 reflects the activation of representations of affective meanings in the affective priming  
4 paradigm (Daltrozzo & Schön, 2008; Eder, Leuthold, Rothermund, & Schweinberger,  
5 2011; Goerlich et al., 2012). That is, the N400 effect was an indication that the primes  
6 activated the representations of affectively related targets in the present study.  
7 Alternatively, given that the amplitude of the N400 is correlated with the difficulty in  
8 affective integration between the primes and targets (Kamiyama et al., 2013; Zhang et  
9 al., 2010), the N400 effect observed in the present study might reflect the difficulty in  
10 affective integration for affectively incongruent trials in the violin version.

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

29 On the other hand, even for the instrumental timbres, violin and flute exhibited  
30 different patterns of neural activities. The violin version elicited an N400 effect at the

1 time window of 280–440 ms, whereas the flute version elicited an LPC at the time  
2 window of 500–600 ms. Such a difference might be attributed to the differences in  
3 acoustic features between the violin and flute. Indeed, previous studies have shown  
4 that neural responses to brighter (Toiviainen et al., 1998) and rougher sounds  
5 exhibited a shorter latency than sounds that were less bright and less rough (De Baene,  
6 Vandierendonck, Leman, Widmann, & Tervaniemi, 2004). In the present study, the  
7 violin version was brighter (violin:  $M = .45$ ,  $SD = .07$ ; flute:  $M = .22$ ,  $SD = .04$ ) and  
8 rougher than the flute (violin:  $M = 224.26$ ,  $SD = 147.17$ ; flute:  $M = 74.96$ ,  $SD =$   
9  $105.97$ ) (see Supplementary Tables 1 and 2 for details). These differences might  
10 account for the distinct neural responses between the violin and flute versions.

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

28 In conclusion, the present findings revealed different patterns of neural responses  
29 to emotional processing of music, when the same melodies were presented in different  
30 timbres: the voice, violin, and flute. Our findings confirmed the hypothesis that there

1 are specialized neural responses to the human voice. Moreover, our findings also  
2 provided insights into the neural correlates underlying the processing of musical  
3 emotion, and how timbre played a role in this processing.

4

# 1 **Acknowledgments**

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

7

1 **Declarations of interest**

2       None.

3

---

## 1   **References**

- 2   Abrahamse, E. L., Duthoo, W., Notebaert, W., & Risko, E. F. (2013). Attention  
3       modulation by proportion congruency: The asymmetrical list shifting effect.  
4       *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(5),  
5       1552–1562.
- 6   Alluri, V., & Toiviainen, P. (2010). Exploring perceptual and acoustical correlates of  
7       polyphonic timbre. *Music Perception*, *27*(3), 223–242.
- 8   Andics, A., Gácsi, M., Faragó, T., Kis, A., & Miklósi, Á. (2014). Voice-sensitive  
9       regions in the dog and human brain are revealed by comparative fMRI. *Current*  
10      *Biology*, *24*(5), 574–578.
- 11  Aramaki, M., Besson, M., Kronland-Martinnet, R., & Ystad, S. (2008). Timbre  
12      perception of sounds from impacted materials: Behavioral, electrophysiological  
13      and acoustic approaches. In S. Ystad, R. Kronland-Martinnet, & K. Jensen (Eds.),  
14      *Lecture Notes in Computer Science: Vol. 5493. Computer Music Modeling and*  
15      *Retrieval. Genesis of Meaning in Sound and Music* (pp. 1–17).  
16      [https://doi.org/10.1007/978-3-642-02518-1\\_1](https://doi.org/10.1007/978-3-642-02518-1_1)
- 17  Balkwill, L. L., & Thompson, W. F. (1999). A cross-cultural investigation of the  
18      perception of emotion in music: Psychophysical and cultural cues. *Music*  
19      *Perception*, *17*(1), 43–64.
- 20  Barthet, M., Depalle, P., Kronland-Martinnet, R., & Ystad, S. (2010). Acoustical  
21      correlates of timbre and expressiveness in clarinet performance. *Music*  
22      *Perception*, *28*(2), 135–154.
- 23  Beauchemin, M., De Beaumont, L., Vannasing, P., Turcotte, A., Arcand, C., Belin, P.,  
24      & Lassonde, M. (2006). Electrophysiological markers of voice familiarity.  
25      *European Journal of Neuroscience*, *23*(11), 3081–3086.
- 26  Behrens, G. A., & Green, S. B. (1993). The ability to identify emotional content of  
27      solo improvisations performed vocally and on three different instruments.  
28      *Psychology of Music*, *21*(1), 20–33.
- 29  Belin, P., Fecteau, S., & Bedard, C. (2004). Thinking the voice: Neural correlates of

- 1 voice perception. *Trends in Cognitive Sciences*, 8(3), 129–135.
- 2 Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal  
3 sounds. *Cognitive Brain Research*, 13(1), 17–26.
- 4 Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective  
5 areas in human auditory cortex. *Nature*, 403(6767), 309–311.
- 6 Bowman, C., & Yamauchi, T. (2016). Perceiving categorical emotion in sound: The  
7 role of timbre. *Psychomusicology: Music, Mind, and Brain*, 26(1), 15–25.
- 8 Bruneau, N., Roux, S., Cléry, H., Rogier, O., Bidet-Caulet, A., & Barthélémy, C.  
9 (2013). Early neurophysiological correlates of vocal versus non-vocal sound  
10 processing in adults. *Brain Research*, 1528, 20–27.
- 11 Calvo, M. G., & Eysenck, M. W. (2008). Affective significance enhances covert  
12 attention: Roles of anxiety and word familiarity. *Quarterly Journal of*  
13 *Experimental Psychology*, 61(11), 1669–1686.
- 14 Capilla, A., Belin, P., & Gross, J. (2012). The early spatio-temporal correlates and  
15 task independence of cerebral voice processing studied with MEG. *Cerebral*  
16 *Cortex*, 23(6), 1388–1395.
- 17 Charest, I., Pernet, C. R., Rousselet, G. A., Quiñones, I., Latinus, M., Fillion-Bilodeau,  
18 S., Chartrand, J. P., & Belin, P. (2009). Electrophysiological evidence for an  
19 early processing of human voices. *BMC Neuroscience*, 10(1), 127.  
20 <https://doi.org/10.1186/1471-2202-10-127>
- 21 Crummer, G. C., Walton, J. P., Wayman, J. W., Hantz, E. C., & Frisina, R. D. (1994).  
22 Neural processing of musical timbre by musicians, nonmusicians, and musicians  
23 possessing absolute pitch. *The Journal of the Acoustical Society of America*,  
24 95(5), 2720–2727.
- 25 Daltrozzo, J., & Schön, D. (2008). Conceptual processing in music as revealed by  
26 N400 effects on words and musical targets. *Journal of Cognitive Neuroscience*,  
27 21(10), 1882–1892.
- 28 De Baene, W., Vandierendonck, A., Leman, M., Widmann, A., & Tervaniemi, M.  
29 (2004). Roughness perception in sounds: Behavioral and ERP evidence.  
30 *Biological Psychology*, 67(3), 319–330.



- 
- 1 Eder, A. B., Leuthold, H., Rothermund, K., & Schweinberger, S. R. (2011).  
2 Automatic response activation in sequential affective priming: An ERP study.  
3 *Social Cognitive and Affective Neuroscience*, 7(4), 436–445.
- 4 Eerola, T., Ferrer, R., & Alluri, V. (2012). Timbre and affect dimensions: Evidence  
5 from affect and similarity ratings and acoustic correlates of isolated instrument  
6 sounds. *Music Perception*, 30(1), 49–70.
- 7 Eerola, T., Friberg, A., & Bresin, R. (2013). Emotional expression in music:  
8 Contribution, linearity, and additivity of primary musical cues. *Frontiers in*  
9 *Psychology*, 4(487). <https://doi.org/10.3389/fpsyg.2013.00487>
- 10 Erber, R., & Erber, M. W. (1994). Beyond mood and social judgment: Mood  
11 incongruent recall and mood regulation. *European Journal of Social Psychology*,  
12 24(1), 79–88.
- 13 Foti, D., Hajcak, G., & Dien, J. (2009). Differentiating neural responses to emotional  
14 pictures: Evidence from temporal - spatial PCA. *Psychophysiology*, 46(3), 521–  
15 530.
- 16 Gabrielsson, A., & Juslin, P. N. (1996). Emotional expression in music performance:  
17 Between the performer's intention and the listener's experience. *Psychology of*  
18 *Music*, 24(1), 68–91.
- 19 Goerlich, K. S., Witteman, J., Schiller, N. O., Van Heuven, V. J., Aleman, A., &  
20 Martens, S. (2012). The nature of affective priming in music and speech. *Journal*  
21 *of Cognitive Neuroscience*, 24(8), 1725–1741.
- 22 Gong, X., Huang, Y., Wang, Y., & Luo, Y. (2011). Revision of the Chinese facial  
23 affective picture system. *Chinese Mental Health Journal*, 25(1), 40–46.
- 24 Griffiths, S. W., Brockmark, S., Höjesjö, J., & Johnsson, J. I. (2004). Coping with  
25 divided attention: The advantage of familiarity. *Proceedings of the Royal Society*  
26 *of London. Series B: Biological Sciences*, 271(1540), 695–699.
- 27 Griffiths, T. D., & Warren, J. D. (2004). What is an auditory object? *Nature Reviews*  
28 *Neuroscience*, 5(11), 887–892.

- 
- 1 Grossmann, T., Oberecker, R., Koch, S. P., & Friederici, A. D. (2010). The  
2 developmental origins of voice processing in the human brain. *Neuron*, *65*(6),  
3 852–858.
- 4 Hailstone, J. C., Omar, R., Henley, S. M., Frost, C., Kenward, M. G., & Warren, J. D.  
5 (2009). It's not what you play, it's how you play it: Timbre affects perception of  
6 emotion in music. *Quarterly Journal of Experimental Psychology*, *62*(11), 2141–  
7 2155.
- 8 Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on  
9 estimated values of  $d'$ . *Behavior Research Methods*, *27*(1), 46–51.
- 10 Hermans, D., De Hower, J., & Eelen, P. (2001). A time course analysis of the  
11 affective priming effect. *Cognition and Emotion*, *15*(2), 143–165.
- 12 Herring, D. R., Taylor, J. H., White, K. R., & Crites, S. L. (2011).  
13 Electrophysiological responses to evaluative priming: The LPP is sensitive to  
14 incongruity. *Emotion*, *11*(4), 794–806.
- 15 Hinojosa, J. A., Carretié, L., Méndezbértolo, C., Míguez, A., & Pozo, M. A. (2009).  
16 Arousal contributions to affective priming: Electrophysiological correlates.  
17 *Emotion*, *9*(2), 164–171.
- 18 Kamiyama, K. S., Abla, D., Iwanaga, K., & Okanoya, K. (2013). Interaction between  
19 musical emotion and facial expression as measured by event-related potentials.  
20 *Neuropsychologia*, *51*(3), 500–505.
- 21 Kujawa, A., Weinberg, A., Hajcak, G., & Klein, D. N. (2013). Differentiating event -  
22 related potential components sensitive to emotion in middle childhood: Evidence  
23 from temporal-spatial PCA. *Developmental Psychobiology*, *55*(5), 539–550.
- 24 Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning  
25 in the N400 component of the event related brain potential (ERP). *Annual*  
26 *Review of Psychology*, *62*(1), 621–647.
- 27 Kutas, M., & Hillyard, S. A. (1980). Reading between the lines: Event-related brain  
28 potentials during natural sentence processing. *Brain and Language*, *11*(2), 354–  
29 373.

- 
- 1 Lense, M. D., Gordon, R. L., Key, A. P., & Dykens, E. M. (2012, July). *Neural*  
2 *correlates of musical timbre perception in Williams syndrome*. Paper presented  
3 at the Twelfth International Conference on Music Perception and Cognition,  
4 Thessaloniki, Greece.
- 5 Leutgeb, V., Schäfer, A., Köchel, A., & Schienle, A. (2012). Exposure therapy leads  
6 to enhanced late frontal positivity in 8-to 13-year-old spider phobic girls.  
7 *Biological Psychology, 90*(1), 97–104.
- 8 Levy, D. A., Granot, R., & Bentin, S. (2001). Processing specificity for human voice  
9 stimuli: Electrophysiological evidence. *NeuroReport, 12*(12), 2653–2657.
- 10 Levy, D. A., Granot, R., & Bentin, S. (2003). Neural sensitivity to human voices: ERP  
11 evidence of task and attentional influences. *Psychophysiology, 40*(2), 291–305.
- 12 Logeswaran, N., & Bhattacharya, J. (2009). Crossmodal transfer of emotion by music.  
13 *Neuroscience Letters, 455*(2), 129–133.
- 14 Luck, S. J. (2005). Ten simple rules for designing and interpreting ERP experiments.  
15 In T. C. Handy (Eds.), *Event-related potentials: A methods handbook* (pp. 17–  
16 32). Cambridge, MA: The MIT Press.
- 17 Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd  
18 ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- 19 Maxwell, S. E., & Delaney, H. D. (2003). *Designing experiments and analyzing data:*  
20 *A model comparison perspective*. New York, NY: Routledge.
- 21 McAdams, S., Cunible, J., Carlyon, R. P., Darwin, C. J., & Russell, I. J. (1992).  
22 Perception of timbral analogies. *Philosophical Transactions of the Royal Society*  
23 *of London. Series B: Biological Sciences, 336*(1278), 383–389.
- 24 Menon, V., Levitin, D. J., Smith, B. K., Lembke, A., Krasnow, B. D., Glazer, D.,  
25 Glover, G. H., & McAdams, S. (2002). Neural correlates of timbre change in  
26 harmonic sounds. *NeuroImage, 17*(4), 1742–1754.
- 27 Owings, D. H., & Morton, E. S. (1998). *Animal vocal communication: A new*  
28 *approach*. Cambridge, MA: Cambridge University Press.

- 
- 1 Paquette, S., Peretz, I., & Belin, P. (2013). The “Musical Emotional Bursts”: A  
2 validated set of musical affect bursts to investigate auditory affective processing.  
3 *Frontiers in Psychology, 4*(509). <https://doi.org/10.3389/fpsyg.2013.00509>
- 4 Petkov, C. I., Kayser, C., Steudel, T., Whittingstall, K., Augath, M., & Logothetis, N.  
5 K. (2008). A voice region in the monkey brain. *Nature Neuroscience, 11*(3),  
6 367–374.
- 7 Polich, J., & Kok, A. (1995). Cognitive and biological determinants of P300: An  
8 integrative review. *Biological Psychology, 41*(2), 103–146.
- 9 Schirmer, A., Kotz, S. A., & Friederici, A. D. (2002). Sex differentiates the role of  
10 emotional prosody during word processing. *Cognitive Brain Research, 14*(2),  
11 228–233.
- 12 Semlitsch, H. V., Anderer, P., Schuster, P., & Presslich, O. (1986). A solution for  
13 reliable and valid reduction of ocular artifacts, applied to the P300 ERP.  
14 *Psychophysiology, 23*(6), 695–703.
- 15 Sollberge, B., Rebe, R., & Eckstein, D. (2003). Musical chords as affective priming  
16 context in a word-evaluation task. *Music Perception, 20*(3), 263–282.
- 17 Steinbeis, N., & Koelsch, S. (2009). Affective priming effects of musical sounds on  
18 the processing of word meaning. *Journal of Cognitive Neuroscience, 23*(3), 604–  
19 621.
- 20 Timmers, R., & Crook, H. (2014). Affective priming in music listening: Emotions as  
21 a source of musical expectation. *Music Perception, 31*(5), 470–484.
- 22 Toiviainen, P., Tervaniemi, M., Louhivuori, J., Saher, M., Huotilainen, M., &  
23 Näätänen, R. (1998). Timbre similarity: Convergence of neural, behavioral, and  
24 computational approaches. *Music Perception, 16*(2), 223–241.
- 25 Weinberg, A., & Hajcak, G. (2011). The late positive potential predicts subsequent  
26 interference with target processing. *Journal of Cognitive Neuroscience, 23*(10),  
27 2994–3007.
- 28 Werheid, K., Alpay, G., Jentsch, I., & Sommer, W. (2005). Priming emotional facial  
29 expressions as evidenced by event-related brain potentials. *International Journal*  
30 *of Psychophysiology, 55*(2), 209–219.

- 
- 1 Wolfe, J. (2018). The acoustics of woodwind musical instruments. *Acoustics Today*,  
2 14, 50–56.
- 3 Wolfe, J., Garnier, M., & Smith, J. (2009). Vocal tract resonances in speech, singing,  
4 and playing musical instruments. *Human Frontier Science Program*, 3(1), 6–23.
- 5 Zhang, Q., Kong, L., & Yang, J. (2012). The interaction of arousal and valence in  
6 affective priming: Behavioral and electrophysiological evidence. *Brain Research*,  
7 1474, 60–72.
- 8 Zhang, Q., Lawson, A., Guo, C., & Jiang, Y. (2006). Electrophysiological correlates  
9 of visual affective priming. *Brain Research Bulletin*, 71(3), 316–323.
- 10 Zhang, Q., Li, X., Gold, B. T., & Jiang, Y. (2010). Neural correlates of cross-domain  
11 affective priming. *Brain Research*, 1329, 142–151.
- 12

---

## 1 **Figure Captions**

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

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

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

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

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

Figure 1

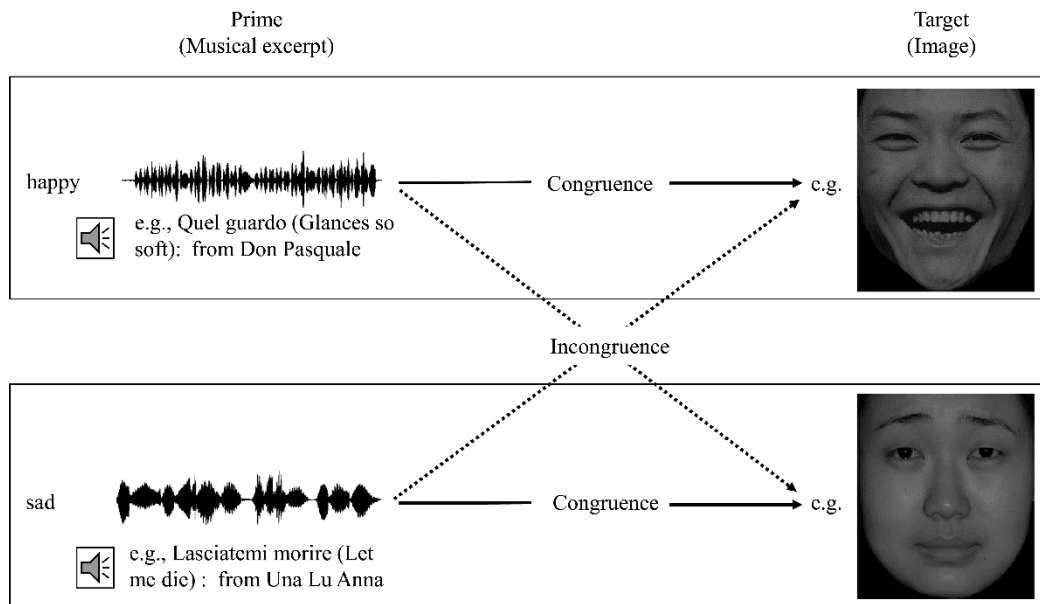
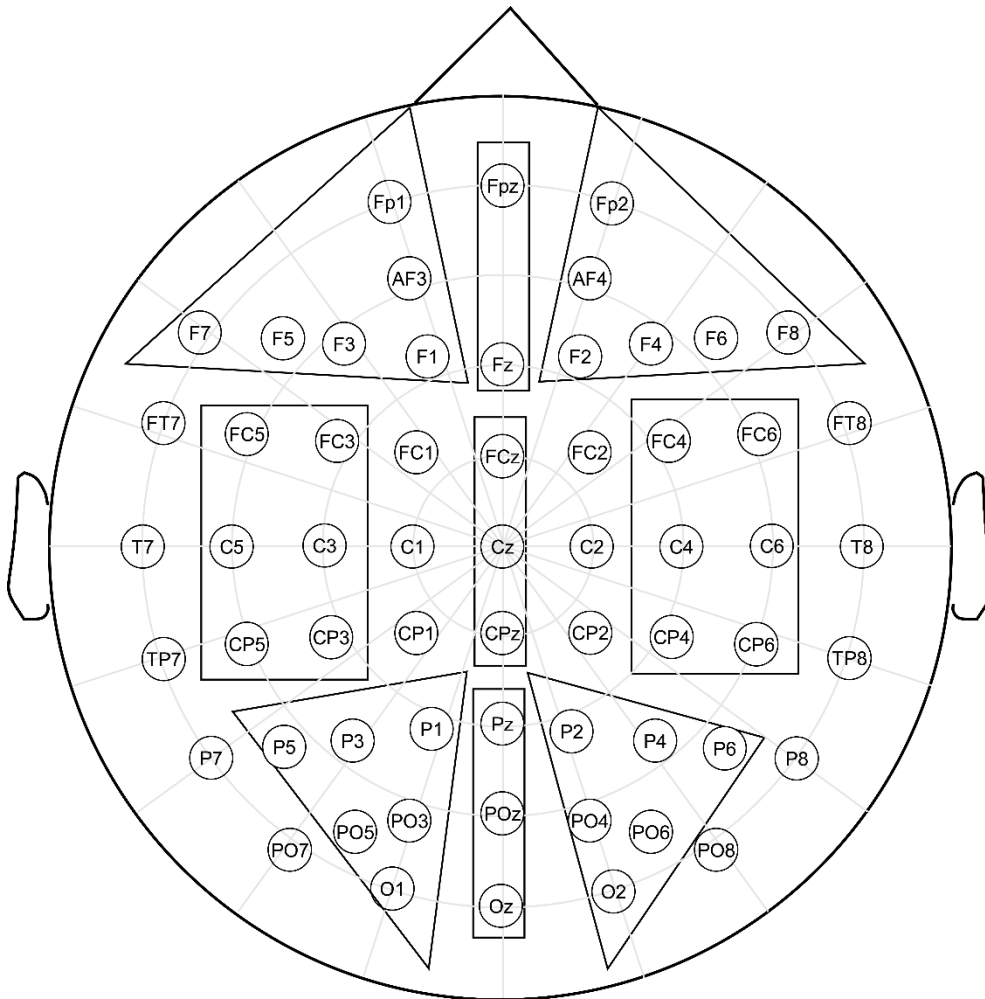


Figure 2





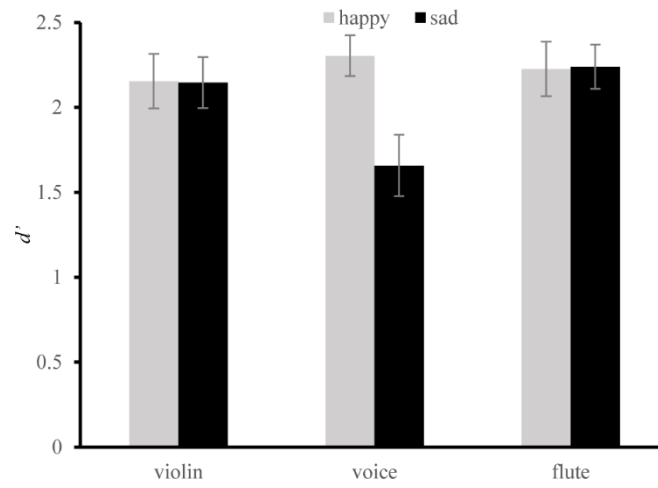
**Figure 3**

Figure 4

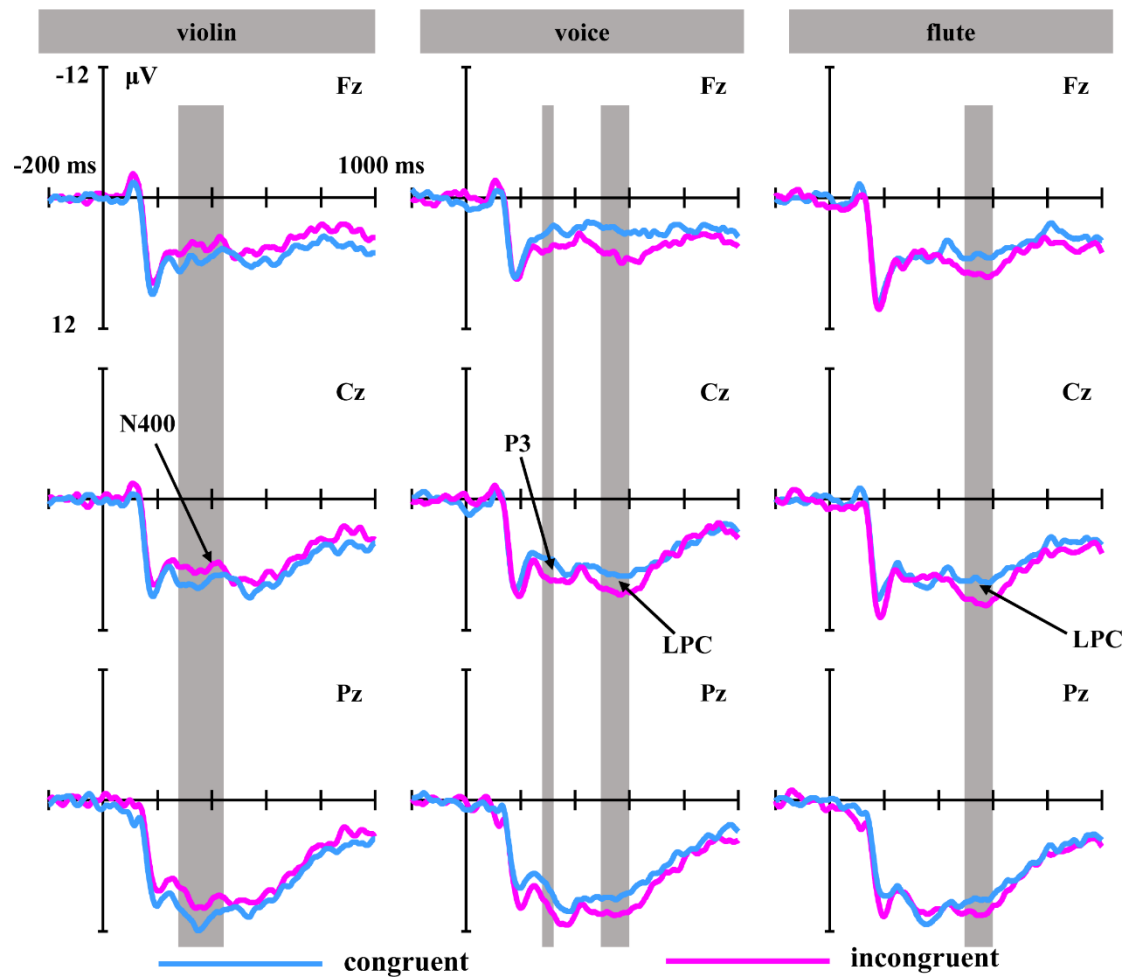


Figure 5

