An ERP Study on Emotional Processing of Vocal and Instrumental Music

Authors: Jiang Jun, Zhang Weixia, Wanqi Wang, Jiang Jun

Date: 2020-03-01T16:46:53+00:00

Abstract

This study examined the electrophysiological differences in emotion processing between vocal and instrumental music using a cross-modal emotional priming paradigm. The prime stimuli consisted of vocal and instrumental music pieces (violin performance), while the target stimuli were facial expression images that were either emotionally congruent or incongruent with the music. The results demonstrated that, compared to the congruent condition, instrumental music incongruent with facial emotions elicited an N400, whereas vocal music incongruent with facial emotions elicited an LPC. These findings indicate that differential electrophysiological responses underlie emotion processing in vocal versus instrumental music.

Full Text

An ERP Study on Emotional Processing in Vocal and Instrumental Music

Jiang Jun[1], Zhang Weixia[2], Wang Wanqi[3]
(1 Music College, Shanghai Normal University; 2 College of Education, Shanghai Normal University; 3 School of Finance and Business, Shanghai Normal University, Shanghai 200234, China)

Abstract

This study employed a crossmodal affective priming paradigm to investigate the electrophysiological differences in emotional processing between vocal and instrumental music. Prime stimuli were vocal and instrumental pieces (violin performances), while target stimuli were facial expression images that were either emotionally congruent or incongruent with the music. The results showed that, compared to congruent conditions, instrumental music that was emotionally incongruent with faces elicited an N400, whereas vocal music that was emotionally incongruent with faces elicited an LPC. These findings indicate that there are differential electrophysiological responses to emotional processing of vocal and instrumental music.

Keywords: vocal music; instrumental music; emotion processing; N400; LPC
Classification Numbers: B842; B845

Listening to music constitutes an important part of modern life. Vocal music (i.e., songs) and instrumental music (i.e., pure music) are two primary types of music in daily listening. However, some individuals prefer listening to instrumental music, while others favor vocal music, believing that lyrics help them understand musical meaning, such as its emotional significance. Do listeners process emotion differently in vocal and instrumental music? Previous behavioral studies have suggested differences in emotional processing between these two modalities. In Stratton and Zalanowski's (1994) study, college students listened to a song expressing sadness, its instrumental version, or its lyrics alone, and rated the pleasantness of each stimulus on a 10-point scale (1 = very unpleasant, 10 = very pleasant). Results showed that participants rated the instrumental version as significantly more pleasant than both the song and lyrics. Using a 5-point rating scale, Brattico et al. (2011) similarly found that participants rated instrumental music as more pleasant than songs. However, contradictory findings have also emerged. Loui, Bachorik, Li, and Schlaug (2013) had participants listen to 16 songs and their instrumental versions, rating musical pleasantness on a 5-point scale, and found no significant difference in pleasantness ratings between vocal and instrumental versions.

The aforementioned studies examined vocal and instrumental emotion processing from the perspective of emotional valence, while other research has explored this issue from the dimension of emotional intensity. In Ali and Peynircioğlu's (2006) study, college students listened to 16 songs and their corresponding instrumental versions, then rated the emotional intensity expressed by the music (1 = not at all, 9 = very much). Results indicated that instrumental music was rated as having higher emotional intensity than songs. However, Franco, Chew, and Swaine (2017) reached different conclusions. The researchers had three groups of college students listen to instrumental music, songs with instrumental accompaniment, or songs without instrumental accompaniment, and rate emotional intensity (1 = not at all, 7 = very much). Their findings showed that participants rated emotional intensity similarly across different musical conditions.

The inconsistent findings in previous research may be attributed to the subjective and relatively insensitive nature of behavioral measures. Therefore, using more objective and sensitive electrophysiological indicators (Kraus & Horowitz-Kraus, 2014; Luck, 2014) may better reveal whether differences exist in emotional processing between vocal and instrumental music. The crossmodal affective priming paradigm has been widely used in ERP studies of emotion processing (e.g., Diamond & Zhang, 2016; Steinbeis & Koelsch, 2011). Such research has primarily focused on the N400 and LPC components. The N400 reflects semantic or conceptual integration processes (Kutas & Federmeier, 2011), peaking around 400 ms post-stimulus, typically distributed over central and parietal electrode sites with a slight right-hemisphere lateralization. Nevertheless, in crossmodal affective priming studies, target stimuli that are emotionally incongruent with primes—whether musical (Goerlich et al., 2012; Steinbeis & Koelsch, 2011), prosodic (Paulmann & Pell, 2010; Schirmer, Kotz, & Friederici, 2005), or facial (Diamond & Zhang, 2016; Föcker & Röder, 2019)—have elicited larger N400 components. This suggests that the N400 also reflects the detection and integration of emotional conflict (Zhang, Lawson, Guo, & Jiang, 2006; Zhang, Li, Gold, & Jiang, 2010). The LPC is a positive wave occurring 300–1000 ms after emotional stimulus presentation, peaking at 500–600 ms, typically distributed over midline and parietal regions, and primarily reflecting attentional and integrative processing of emotional information (Hajcak, MacNamara, & Olvet, 2010). Specifically, emotional stimuli such as pictures (Hajcak & Olvet, 2008) and speech prosody (Paulmann & Uskul, 2017) elicit larger LPCs across the scalp compared to neutral stimuli. Due to their strong motivational significance, emotional stimuli attract greater attention (Schupp et al., 2000). Similarly, target stimuli that are emotionally incongruent with priming speech prosody (Pell et al., 2015; Zheng & Huang, 2013) and facial pictures (Diamond & Zhang, 2016) elicit larger LPCs over bilateral parieto-occipital regions or the entire scalp. This demonstrates that the LPC is associated with attention allocation and emotional integration (Herring, Taylor, White, & Crites Jr, 2011; Zhang, Kong, & Jiang, 2012; Zhang et al., 2010).

Recently, Zhang et al. (2018) used ERP technology to examine the influence of lyrics on musical emotion processing. Their study showed that humming without lyrics elicited an N400 across the scalp, while music with lyrics elicited an LPC. To date, no research has investigated the differences in emotional processing between vocal and instrumental music from an electrophysiological perspective. Based on this, the present study employed a crossmodal affective priming paradigm to explore this question. To eliminate differences in musical structure between vocal and instrumental pieces, we selected identical musical excerpts performed by both voice and violin, thus creating vocal and instrumental versions. Since the presence of lyrics elicits an LPC while lyricless humming elicits an N400 (Zhang et al., 2018), we hypothesized that vocal and instrumental music might produce LPC and N400 effects, respectively.

2.1 Participants

Based on the effect size for emotional congruence from previous research (Zhang et al., 2018) ($F(1, 31) = 9.80$, p 2= 0.24), G*Power 3.1.9.4 software (Faul, Erdfelder, Lang, & Buchner, 2007) calculated that a minimum sample size of 16 was required to achieve 90% statistical power at the 0.01 significance level. We therefore recruited 20 college students without professional musical training as participants. Due to excessive EEG artifacts, three participants were excluded from data analysis, leaving a final sample of 17 participants (9 males; age = 24.65 ± 1.31 years). All participants were right-handed with normal hearing and (corrected) vision, and no history of brain injury or mental illness. All participants provided informed consent and received monetary compensation.

2.2 Stimuli

Stimuli consisted of prime and target stimuli. Prime stimuli were 50 foreign opera excerpts expressing positive emotion and 50 expressing negative emotion selected from vocal music textbooks (duration: 10–25 s, mean = 17 s). The opera excerpts were recorded by a graduate student with 18 years of training in either vocal performance or violin, who performed the excerpts either by singing in Chinese or playing violin. This eliminated differences in musical structure between vocal and instrumental versions. The experiment included 100 vocal fragments and 100 instrumental fragments. These musical excerpts were recorded in mono at a sampling rate of 22050 Hz. Finally, Adobe Audition CS6 software was used to normalize the volume to -7 dB and apply a 1 s fade-out.

Target stimuli were two female facial expression pictures selected from the Chinese Facial Affective Picture System (Gong et al., 2011), depicting happiness and sadness respectively. The average emotion recognition accuracy for both pictures was 97.98%, with mean emotional intensity ratings of 7.24 and 5.88. Each musical excerpt was paired with the two facial pictures (see Figure 1 [Figure 1: see original paper]), creating 200 emotionally congruent and 200 emotionally incongruent pairings.

Figure 1. Example of the affective priming paradigm.

To ensure stimulus validity, we conducted three pilot tests. First, an additional 16 college students rated the emotional congruence of the 400 music-picture pairings (1 = very incongruent, 9 = very congruent). For each musical excerpt, paired t-tests were used to examine whether ratings differed significantly between emotionally congruent and incongruent pairings. Two hundred forty pairings were selected as formal stimuli: regardless of whether the prime was vocal ($t(59) = 24.73$, $p < .001$, $d = 3.19$) or instrumental ($t(59) = 21.56$, $p < .001$, $d = 2.78$), listeners rated emotionally congruent pairings (vocal: M = 7.21, SD = 0.79; instrumental: M = 6.94, SD = 0.87) as more congruent than incongruent pairings (vocal: M = 2.72, SD = 0.80; instrumental: M = 3.15, SD = 0.77). A repeated-measures ANOVA with emotional congruence (congruent, incongruent) and music type (vocal, instrumental) as within-subject variables revealed a significant main effect of emotional congruence ($F(1, 59) = 822.81$, $p < .001$, p 2 = .93) and a significant interaction between emotional congruence and music type ($F(1, 59) = 11.06$, $p = .002$, p 2 = .16), while the main effect of music type was not significant, $F(1, 59) = 1.14$, $p = .290$. These results indicate that the experimental materials were valid and that task difficulty was similar across the two music types. Additionally, these participants rated the clarity of lyrics in vocal music (1 = unclear, 5 = clear) as higher than 4 ($M = 4.45$, $SD = 0.23$). Second, an additional 16 music majors (all with over 18 years of professional musical training) rated the consistency between vocal and instrumental versions in terms of tempo rubato ($M = 5.24$, $SD = 0.42$), dynamics ($M = 5.38$, $SD = 0.40$), phrasing ($M = 5.65$, $SD = 0.36$), and overall performance level ($M = 5.21$, $SD = 0.40$) on a 7-point scale (1 = very inconsistent, 7 = very consistent), with all mean ratings above 5. Third, an additional 16 college students rated the emotional valence (1 = very unpleasant, 4 = neutral, 7 = very pleasant) and arousal level (1 = very calm, 4 = neutral, 7 = very excited) of the music. For emotional valence, a repeated-measures ANOVA with emotion type (positive, negative) and music type (vocal, instrumental) as within-subject variables revealed significant main effects of emotion type ($F(1, 29) = 2241.98$, $p < .001$, p 2 = .99) and music type ($F(1, 29) = 4.80$, $p = .037$, p 2 = .14), as well as a significant interaction ($F(1, 29) = 8.37$, $p = .007$, p 2 = .22). Simple effects analysis showed no difference in valence between instrumental ($M = 5.43$, $SD = 0.30$) and vocal ($M = 5.46$, $SD = 0.26$) versions for positive music ($F(1, 29) = 0.18$, $p = .674$). For negative music, however, instrumental versions ($M = 2.66$, $SD = 0.30$) were rated as more pleasant than vocal versions ($M = 2.41$, $SD = 0.35$), $F(1, 29) = 13.40$, $p = .001$, p 2 = .32. For arousal level, significant main effects of emotion type ($F(1, 29) = 111.87$, $p < .001$, p 2 = .79) and music type ($F(1, 29) = 419.34$, $p < .001$, p 2 = .94) and a significant interaction ($F(1, 29) = 18.97$, $p < .001$, p 2 = .40) were found. Simple effects analysis revealed that for both positive ($F(1, 29) = 36.07$, $p < .001$, p 2 = .55) and negative ($F(1, 29) = 206.79$, $p < .001$, p 2 = .88) music, instrumental versions (positive: $M = 4.87$, $SD = 0.59$; negative: $M = 3.32$, $SD = 0.41$) elicited lower arousal than vocal versions (positive: $M = 5.54$, $SD = 0.47$; negative: $M = 4.80$, $SD = 0.49$).

2.3 Procedure

First, a black fixation cross "+" appeared at the center of the computer screen for 1 s. After the fixation disappeared, the prime stimulus was presented and participants listened to the music through headphones. Following music playback, the target stimulus was presented for 1 s. Finally, a response screen appeared, prompting participants to judge whether the emotional valence of the music and facial picture were congruent, after which they pressed the spacebar to begin the next trial. After the experiment, participants were asked to report the titles of the musical pieces, but they failed to report any keywords from the titles.

2.4 Data Recording and Analysis

EEG signals were recorded using a NeuroScan Synamps 2 system with a 64-channel electrode cap arranged according to the 10-20 system. The left mastoid served as the reference electrode, and the right mastoid as the recording electrode. Signals were amplified by an AC amplifier with a bandpass filter of 0.05–

100 Hz and a sampling rate of 500 Hz. Electrode impedance was maintained below 5 kΩ.

Data were preprocessed using Neuroscan 4.5 software. First, raw data were re-referenced to the average of the bilateral mastoid electrodes. Artifacts such as muscle activity, cardiac signals, and drift were manually rejected through visual inspection, and ocular artifacts were automatically corrected using the software's regression procedure. Second, a bandpass filter of 0.1–30 Hz (24 dB/oct slope) was applied. Then, EEG epochs from -200 to 1000 ms relative to target stimulus onset were segmented and baseline-corrected (-200 to 0 ms). Finally, trials with amplitudes exceeding $±$80 V and incorrect responses were automatically rejected, and trials for each experimental condition were averaged separately.

Following preprocessing, to control for Type I error probability resulting from multiple comparisons across numerous electrodes and time points, we used cluster-based permutation tests from the FieldTrip toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011; Maris & Oostenveld, 2007) for statistical analysis. First, dependent samples t-tests were conducted to compare differences between emotionally congruent and incongruent conditions at each adjacent electrode-time point pair. If the difference reached significance at the .05 level, that pair was identified as a cluster. For each cluster, the sum of t-values across all electrode-time point pairs within the cluster served as the cluster-level statistic. We then randomly shuffled each participant's data between congruent and incongruent conditions 2000 times to construct a null distribution assuming no difference between conditions. Finally, we compared the observed cluster-level statistics against this null distribution and estimated p-values using Monte Carlo methods ($\alpha = .05$). To increase statistical sensitivity, based on previous research on N400 (Schirmer et al., 2005; Steinbeis & Koelsch, 2011) and LPC (Herring et al., 2011; Zhang et al., 2010) and visual inspection, we defined 200–500 ms and 400–700 ms post-face onset as the time windows for N400 and LPC, respectively.

3.1 Behavioral Results

We conducted separate repeated-measures ANOVAs on accuracy and reaction time with emotional congruence and music type as within-subject variables. For accuracy, the main effect of emotional congruence was significant (F(1, 16) = 13.50, p = .002, p 2 = .46), indicating that participants performed better on emotionally congruent pairings (M = 92.04%, SD = 5.93) than on incongruent pairings (M = 81.27%, SD = 14.40). The main effect of music type was also significant (F(1, 16) = 17.15, p = .001, p 2 = .52), showing that participants were more accurate in judging emotional congruence when primed by vocal music (M = 89.79%, SD = 8.94) than by instrumental music (M = 83.52%, SD = 14.21). However, the interaction between the two factors was not significant (F(1, 16) = 0.02, p = .903).

For reaction time, the main effect of emotional congruence was significant (F(1, 16) = 12.54, p = .003, p 2 = .44), indicating faster responses to emotionally congruent pairings (M = 993.40 ms, SD = 456.20) than to incongruent pairings (M = 1174.97 ms, SD = 595.25). However, neither the main effect of music type (F(1, 16) = 0.49, p = .495) nor the interaction (F(1, 16) = 0.27, p = .613) was significant.

3.2 ERP Results

In the N400 time window, cluster-based permutation tests revealed that when the prime was instrumental, emotionally incongruent face pictures elicited more negative brain potentials than congruent faces (p = .011; see Figure 2 [Figure 2: see original paper]). As shown in Figure 3 [Figure 3: see original paper], this effect primarily occurred at right frontotemporal and centroparietal electrodes between 281–471 ms. However, when the prime was vocal, no difference was observed between emotionally congruent and incongruent face pictures.

In the LPC time window, permutation tests indicated that when the prime was vocal, emotionally incongruent faces elicited a larger positive component than congruent faces (p = .019; see Figure 2). As shown in Figure 3, this component primarily occurred at left frontocentral and parieto-occipital electrodes between 473–677 ms. However, when the prime was instrumental, no difference was observed between emotionally congruent and incongruent faces.

Figure 2. Grand-averaged ERP waveforms elicited by face pictures when primed by instrumental (a) and vocal (b) music at nine representative electrodes. Shaded areas represent 95% CI of mean amplitude at each time point.

Figure 3. Scalp topographic maps of difference waves generated by face pictures in two time windows when primed by instrumental (a) and vocal (b) music. Asterisks indicate electrodes where ERP effects were significant within specific time windows.

This ERP study investigated whether differential electrophysiological responses exist for emotional processing of vocal versus instrumental music. The findings demonstrated that instrumental music emotionally incongruent with faces elicited an N400, whereas vocal music emotionally incongruent with faces elicited an LPC. These results indicate that vocal and instrumental music emotion processing involve distinct electrophysiological responses.

We found that when instrumental excerpts served as primes, emotionally incongruent face pictures produced a larger N400 than congruent faces. This result aligns with previous findings showing that emotionally incongruent targets—whether faces or emotionally meaningful words—elicited larger (N400-like) responses when primed by single chords (Steinbeis & Koelsch, 2011; Zhou et al., 2019) or instrumental excerpts (Goerlich et al., 2012; Zhang et al., 2019). The N400 in the present study likely reflects the detection and integration of emotionally conflicting information (Zhang et al., 2006, 2010). Specifically, prime stimuli pre-activated emotional representations at the conceptual level related to the target stimuli, thereby reducing N400 amplitudes for emotionally congruent targets (Goerlich et al., 2012). For emotionally incongruent targets, however, the relevant emotional representations were not pre-activated, requiring more cognitive resources for emotional information integration (Zhang et al., 2010) and resulting in larger N400 amplitudes. Since this activation occurs automatically (Zhang et al., 2006), the N400 is associated with automatic integrative processing of emotional information.

Furthermore, the N400 in this study was primarily generated at right frontotemporal and centroparietal electrodes, indicating right-hemisphere lateralization. This is consistent with previous research showing right-lateralized N400 effects when participants judged whether target and prime stimuli were emotionally (Schirmer et al., 2005) or semantically (Painter & Koelsch, 2011) congruent. Indeed, fMRI studies have also shown that instrumental music activates more right-hemisphere regions than songs, such as the middle frontal gyrus, inferior frontal gyrus, and cingulate gyrus (Brattico et al., 2011). However, our findings contradict ERP studies reporting that N400 effects for targets incongruent with chord or instrumental primes originated from whole-scalp (Steinbeis & Koelsch, 2011; Zhou et al., 2019) or anterior (Goerlich et al., 2012) electrodes. This discrepancy may stem from differences in task demands: in those studies, participants judged the emotional valence of target stimuli, whereas in our study, they judged the emotional congruence between primes and targets, likely directing greater attention to the prime's emotion. Future research should investigate this issue.

We also found that when vocal excerpts served as primes, emotionally incongruent faces elicited a larger LPC than congruent faces. This result is consistent with previous findings showing that faces incongruent with songs (Zhang et al., 2018) and speech prosody (Pell et al., 2015; Zheng & Huang, 2013) elicited larger LPCs. The LPC in this study may reflect attentional resource allocation (Zhang et al., 2010, 2012). When the prime was a song, the higher arousal level of vocal music likely captured more attention (Schupp et al., 2000). Additionally, our LPC may reflect emotional integration processes (Herring et al., 2011), as listeners needed to integrate the emotions from lyrics and music before matching them with the facial expression. Because this involves linguistic information, listeners must consciously process emotional information using propositional representations or thought (Strack & Deutsch, 2004), reflecting reflective integrative processing (Imbir, Spustek, & Żygierewicz, 2016). However, vocal music emotion processing did not produce an N400, replicating Zhang et al.'s (2018) finding that faces incongruent with songs did not elicit an N400. This may be because N400 and LPC reflect different emotional integration functions. This interpretation aligns with the dual-model perspective of emotion (Jarymowicz & Imbir, 2015), which posits that emotional responding involves two independent evaluation systems—an automatic evaluation system and a reflective evaluation system (Jarymowicz & Imbir, 2015): the former does not require verbal mediation, while the latter does (Imbir et al., 2016).

Regarding LPC scalp distribution, our LPC was primarily generated at left frontocentral and parieto-occipital electrodes, indicating left-hemisphere lateralization. This is consistent with previous research showing larger LPCs at left electrodes when word emotion is related to reflective evaluation systems (Imbir et al., 2016) or when prime and target faces are emotionally incongruent (Cheal et al., 2014). fMRI studies have also shown that songs activate more left-hemisphere regions than instrumental music, such as the superior temporal gyrus, putamen, cuneus, and cerebellar declive (Brattico et al., 2011). However, our findings contradict ERP studies reporting that LPCs for faces incongruent with songs or speech prosody originated from whole-scalp (Zhang et al., 2018) or bilateral parieto-occipital (Pell et al., 2015; Zheng & Huang, 2013) electrodes. This inconsistency may be related to differences in the emotional valence and arousal levels of priming stimuli across studies. Since these studies did not provide relevant data, we cannot determine the specific cause of these contradictory results.

Acknowledgments

We thank Professor Jiang Cunmei from the Music College of Shanghai Normal University for her assistance with this manuscript. We also thank Dr. Cai Danchao from Shanghai Public Health Clinical Center, Dr. Sun Lijun from the Institute of Psychology, Chinese Academy of Sciences, and Dr. Li Peiyang from Chongqing University of Posts and Telecommunications for their help with ERP data analysis and figure preparation.

References

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.

Zhang, W., Wang, W., Zhou, L., & Jiang, C. (2018). The influence of lyrics on musical emotion processing: Behavioral and ERP study. *Acta Psychologica Sinica, 50*(12), 1339–1350.

Zheng, Z., & Huang, X. (2013). Emotional speech modulates facial expression recognition: ERP evidence. *Psychological Science, 36*(1), 33–37.

Ali, S. O., & Peynircioğlu, Z. F. (2006). Songs and emotions: Are lyrics and melodies equal partners? *Psychology of Music, 34*(4), 511–534.

Brattico, E., Alluri, V., Bogert, B., Jacobsen, T., Vartiainen, N., Nieminen, S., & Tervaniemi, M. (2011). A functional MRI study of happy and sad emotions in music with and without lyrics. *Frontiers in Psychology, 2*. doi:10.3389/fpsyg.2011.00308

Cheal, J. L., Heisz, J. J., Walsh, J. A., Shedden, J. M., & Rutherford, M. D. (2014). Afterimage induced neural activity during emotional face perception. *Brain Research, 1549*, 11–21.

Diamond, E., & Zhang, Y. (2016). Cortical processing of phonetic and emotional information in speech: A cross-modal priming study. *Neuropsychologia, 82*, 110–122.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences.* Behavior Research Methods, 39*(2), 175–191.

Föcker, J., & Röder, B. (2019). Event-related potentials reveal evidence for late integration of emotional prosody and facial expression in dynamic stimuli: An ERP study. *Multisensory Research, 32*(6), 473–497.

Franco, F., Chew, M., & Swaine, J. S. (2017). Preschoolers' attribution of affect to music: A comparison between vocal and instrumental performance. *Psychology of Music, 45*(1), 131–149.

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.

Hajcak, G., & Olvet, D. M. (2008). The persistence of attention to emotion: Brain potentials during and after picture presentation. *Emotion, 8*(2), 250–255.

Hajcak, G., MacNamara, A., & Olvet, D. M. (2010). Event-related potentials, emotion, and emotion regulation: An integrative review. *Developmental Neuropsychology, 35*(2), 129–155.

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

Imbir, K. K., Spustek, T., & Żygierewicz, J. (2016). Effects of valence and origin of emotions in word processing evidenced by event related potential correlates in a lexical decision task. *Frontiers in Psychology, 7.* doi: 10.3389/fpsyg.2016.00271

Jarymowicz, M. T., & Imbir, K. K. (2015). Toward a human emotions taxonomy (based on their automatic vs. reflective origin). *Emotion Review, 7*(2), 183–188.

Kraus, D., & Horowitz-Kraus, T. (2014). The effect of learning on feedback-related potentials in adolescents with dyslexia: An EEG-ERP study. *PLoS ONE, 9*(6), e100486.doi:10.1371/journal.pone.0100486

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.

Loui, P., Bachorik, J. P., Li, H. C., & Schlaug, G. (2013). Effects of voice on emotional arousal. *Frontiers in Psychology, 4.* doi: 10.3389/fpsyg.2013.00675

Luck, S. J. (2014). *An introduction to the event-related potential technique* (2nd ed.). Cambridge, MA: The MIT Press.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods, 164*(1), 177-190.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, 1-9. doi:10.1155/2011/156869.

Painter, J. G., & Koelsch, S. (2011). Can out-of-context musical sounds convey meaning? An ERP study on the processing of meaning in music. *Psychophysiology, 48*(5), 645–655.

Paulmann, S., & Pell, M. D. (2010). Contextual influences of emotional speech prosody on face processing: How much is enough? *Cognitive, Affective, & Behavioral Neuroscience, 10*(2), 230–242.

Paulmann, S., & Uskul, A. K. (2017). Early and late brain signatures of emotional prosody among individuals with high versus low power. *Psychophysiology, 54*(4), 555–565.

Pell, M. D., Rothermich, K., Liu, P., Paulmann, S., Sethi, S., & Rigoulot, S. (2015). Preferential decoding of emotion from human non-linguistic vocalizations versus speech prosody. *Biological Psychology, 111*, 14–25.

Schirmer, A., Kotz, S. A., & Friederici, A. D. (2005). On the role of attention for the processing of emotions in speech: Sex differences revisited. *Cognitive Brain Research, 24*(3), 442–452.

Schupp, H. T., Cuthbert, B. N., Bradley, M. M., Cacioppo, J. T., Ito, T., & Lang, P. J. (2000). Affective picture processing: The late positive potential is modulated by motivational relevance. *Psychophysiology, 37*(2), 257–261.

Steinbeis, N., & Koelsch, S. (2011). Affective priming effects of musical sounds on the processing of word meaning. *Journal of Cognitive Neuroscience, 23*(3), 604–621.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review, 8*(3), 220–247.

Stratton, V. N., & Zalanowski, A. H. (1994). Affective impact of music vs. lyrics. *Empirical Studies of the Arts, 12*(2),173–184.

Zhang, Q., Kong, L., & Jiang, Y. (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*(1-3), 316–323.

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

Zhang, W., Liu, F., Zhou, L., Wang, W., Jiang, H., & Jiang, C. (2019). The effects of timbre on neural responses to musical emotion. *Music Perception, 37*(2), 134–146.

Zhou, L., Liu, F., Jiang, J., & Jiang, C. (2019). Impaired emotional processing of chords in congenital amusia: Electrophysiological and behavioral evidence. *Brain and Cognition, 135*, 103577. doi: 10.1016/j.bandc.2019.06.001

Note: Figure translations are in progress. See original paper for figures.