



Musical Tension Associated With Violations of Hierarchical Structure

Lijun Sun^{1,2}, Li Hu^{1,2}, Guiqin Ren³ and Yufang Yang^{1,2*}

¹Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China,

²Department of Psychology, University of Chinese Academy of Sciences, Beijing, China, ³College of Psychology, Liaoning Normal University, Dalian, China

Tension is one of the core principles of emotion evoked by music, linking objective musical events and subjective experience. The present study used continuous behavioral rating and electroencephalography (EEG) to investigate the dynamic process of tension generation and its underlying neurocognitive mechanisms; specifically, tension induced by structural violations at different music hierarchical levels. In the experiment, twenty-four musicians were required to rate felt tension continuously in real-time, while listening to music sequences with either well-formed structure, phrase violations, or period violations. The behavioral data showed that structural violations gave rise to increasing and accumulating tension experience as the music unfolded; tension was increased dramatically by structural violations. Correspondingly, structural violations elicited N5 at GFP peaks, and induced decreasing neural oscillations power in the alpha frequency band (8–13 Hz). Furthermore, compared to phrase violations, period violations elicited larger N5 and induced a longer-lasting decrease of power in the alpha band, suggesting a hierarchical manner of musical processing. These results demonstrate the important role of musical structure in the generation of the experience of tension, providing support to the dynamic view of musical emotion and the hierarchical manner of tension processing.

Keywords: musical tension, hierarchical structure, global field power, N5, alpha

OPEN ACCESS

Edited by:

Franco Delogu,
Lawrence Technological University,
United States

Reviewed by:

José Lino Oliveira Bueno,
University of São Paulo, Brazil
Brian A. Coffman,
University of New Mexico,
United States

*Correspondence:

Yufang Yang
yangyf@psych.ac.cn

Specialty section:

This article was submitted to
Cognitive Neuroscience,
a section of the journal
Frontiers in Human Neuroscience

Received: 30 June 2020

Accepted: 27 August 2020

Published: 18 September 2020

Citation:

Sun L, Hu L, Ren G and Yang Y
(2020) Musical Tension Associated
With Violations of
Hierarchical Structure.
Front. Hum. Neurosci. 14:578112.
doi: 10.3389/fnhum.2020.578112

INTRODUCTION

Music is universal and essential in human societies, owing to its strong power of emotion (e.g., Koelsch, 2014; Swaminathan and Schellenberg, 2015). However, it remains unclear how the auditory elements in music generate emotion responses (Juslin and Laukka, 2004; Juslin and Västfjäll, 2008; Juslin, 2013). In the literature of music theories and music psychology, tension is considered as the link between musical events and experienced emotions (Lehne et al., 2013; Lehne and Koelsch, 2015b) and so that as an appropriate point of penetration to investigate the dynamics and richness of musical emotion. Musical tension is an affective state that is associated with conflict, dissonance, instability, or uncertainty and create a yearning for resolution (Lehne and Koelsch, 2015b). It is believed that the recurrent alterations between tension and relaxation create a music-generated experience (Lerdahl and Jackendoff, 1983).

As to the relationship between musical structure and tension, presently most studies have primarily focused on the relationship between tonally hierarchical structure and tension, and found that tonal regularity plays a critical role in tension generation; moving away from tonal center would induce tension, whereas returning to the tonal center a sense of resolution (Bigand et al., 1996; Bigand and Parncutt, 1999; Steinbeis et al., 2006; Lerdahl and Krumhansl, 2007). For example, Bigand and his colleagues manipulated the stability of certain chords in music sequences and examined the tension rating of the key chords, suggesting that unstable chords gave rise to higher tension experience than stable chords (Bigand et al., 1996; Bigand and Parncutt, 1999). Steinbeis et al. (2006) constructed harmonic regular and irregular conditions by manipulating the last chord of short and real music pieces. The results showed that structural violations induced a strong sense of tension because the ending of the music did not fulfill the expectation. While these studies shed light on the relationship between tonal structure and the experience of tension, musical stimuli used in these studies were short of large-scale structures.

According to the generative theory of tonal music (GTTM) and tonal tension model (TTM), music expresses complicated meaning and emotion with hierarchical structures. Not only each musical event expresses specific tension and relaxation, but the events are also organized into a hierarchy with a prolongational reduction tree, creating a complex pattern of tension and resolution (Lerdahl and Jackendoff, 1983; Lerdahl and Krumhansl, 2007). In real music pieces, discrete elements are organized at multiple time scales to form structural units such as musical section, phrase, period and movement (Lerdahl and Jackendoff, 1983). Based on different hierarchical structural units, multiple local patterns of tension and relaxation constitute a global pattern of tension and relaxation. That is, small tension arches are always embedded in large tension arches based on the hierarchical structure with different timescales in music. Based on the structural relations in music, several tension arches overlap and interweave to develop the tension and resolution in large-scale structures (Koelsch, 2012).

Music psychology has explored the cognitive and neural bases of tension generation. The model of music tension proposed by Margulis (2005) suggests that listeners make predictions for the upcoming events continuously based on the musical context. The mismatch between the actual stimulus and the prediction in the brain will induce tension experience, such as tonal inconsistency and violations. Recently, a general model of tension and suspense has been proposed by Lehne and Koelsch (2015a) based on the predictive coding theory (Friston and Kiebel, 2009). They suggest that prediction errors enable the brain to update predictions for musical events in the process of music listening. Specifically, when an unpredicted event occurs in music progression, listeners' mental models are not maintained stable, but rather updated based on the current perceptual information. In other words, prediction violations in music could lead to predictive errors and event model resetting,

which dynamically engaged working memory and cognitive control processes, such as attention switching (Zacks et al., 2007; Kurby and Zacks, 2008).

In terms of the proposals as to cognitive bases underlying tension generation, we assumed that musical tension could be modulated by the hierarchical levels of musical structure. That is, compared with low-level units, the processing of high-level units would be more difficult and required more cognitive efforts. The difficulties may result from: first, much more information included in musical events at higher levels; second, long-distance dependency involved in integrating multiple small units into a large unit at higher levels. The evidence has been provided by previous studies. It was found that larger closure positive shift (CPS) was elicited by period boundaries than section and phrase boundaries in music, suggesting more retrospective processing of preceding information (Zhang et al., 2016); it was more difficult for a global structure to be integrated into the context than the local structure (Zhang et al., 2018).

How tension experience is modulated by hierarchical structures with different time scales? Whether higher tension would be induced by structural violations at a higher level than that at a lower level? What are the cognitive neural bases underlying the music tension generation? Using continuous behavioral rating and electroencephalography (EEG), the present study explored these questions by investigating the processing of tension experience and neural responses induced by structural violations at different hierarchical levels. In the experiment, four-phrase music sequences were used to organize hierarchical units at different timescales. There were three types of music structures, in which the consistency of the tonality at different hierarchical levels was manipulated, while the musical structure within each phrase was kept the same across conditions. In well-formed sequences, four phrases were organized into a structure with three hierarchical levels and had consistent tonality from the beginning to the end. In sequences with phrase violations and period violations, the tonal consistency was disrupted at either phrase or period boundary respectively.

We hypothesized that tension ratings could be raised by structural violations, and higher tension would be induced by higher-level violations than the lower ones. Furthermore, for the cognitive mechanism underlying tension experience, larger N5 response, a neurophysiological marker of the integrated process of musical structure, would be elicited by sequences with structural violations compared to the well-formed sequences (Poulin-Charronnat et al., 2006; Loui et al., 2007; Miranda and Ullman, 2007; Koelsch and Jentschke, 2010; Sun et al., 2020b). Also, we predicted that the power in the alpha frequency band would decrease in response to structural violations compared with the well-formed sequences since the alpha power was tightly associated with the cognitive resources and attention (Meyer et al., 2013; Sadaghiani and Kleinschmidt, 2016). Finally, we hypothesized that both the N5 component and the power of alpha-band would be influenced by the hierarchical level of structural violations because more cognitive resources were required for working memory updating and attention switching in large structural units than the small ones.

MATERIALS AND METHODS

Participants

Twenty-four musicians, highly proficient in Western tonal music ($M_{\text{age}} = 23.21$ years, $SD = 2.38$, 17 female) participated in the experiment. They had received formal Western instrumental training for an average of 15 years (ranging from 8 to 22 years), playing piano, violin, viola, cello, or tuba. They were all right-handed with normal hearing. None of them reported a history of neural impairment or psychiatric illness. The study was approved by the Institutional Review Board of the Institute of Psychology, Chinese Academy of Sciences, and accords with the ethical principles of the Declaration of Helsinki. All participants were provided with informed consent.

Stimuli

Ten original chorale sequences were composed in 2/4 meter. Four phrases were organized into a well-formed structure with multi-level hierarchies in each original sequence. As illustrated in **Figure 1A** (condition 1), the first two phrases and the last two phrases grouped into two music periods with a consistent tonality relationship. There were no pauses during the whole musical sequence. Musical units were mainly separated by the different durations of notes. The boundaries of phrase and period were signified by half notes, which lasted for 1,200 ms and were longer than other notes. Based on each original sequence, two modified versions (condition 2 and condition 3) were created (see **Figures 1B,C**). The modified versions had structural violations at phrase or period level, while the local structure within each phrase was consistent across conditions. In condition 2, the musical structure was violated at the phrase level by modifying the tonality consistency of each phrase. Thus, the expectation of the listeners cannot be fulfilled at the beginning of each phrase. In condition 3, the musical structure was violated at the period level by modifying the tonality consistency of the two periods. Under such a situation, the expectation of the listeners cannot be fulfilled at the beginning of the second period (that is also the beginning of the third phrase).

The 30 chorale sequences (10 sequences for each condition) were then transposed to three other keys to form a total of 120 sequences. In terms of the sequences, all stimuli were created using the Sibelius 7.5 software (Avid Tech. Incorporated) and were exported in mid format. Using Cubase 5.1 software, the mid files were set at a constant velocity of 100 and were archived with a Yamaha piano timbre at a tempo of 100 beats per minute in .wav format.

Procedure

There were three versions of sequences: original version with no violations, a modified version with violations at the phrase level, another modified version with violations at the period level. Each version has 40 sequences, resulting in 120 chorale sequences in total. They were presented in a pseudorandom order under two constraints: a given version could not be repeated more than three times in succession, and consecutive sequences were not to originate from the same original chorale sequence.

To reveal the dynamic qualities of the tension experience, participants were required to continuously make real-time judgments of the felt tension using the computer mouse. Tension values were recorded using the Psychpy 1.0. software interface and data were collected at a sampling rate of 20 Hz. Before the beginning of each trial, they will see a sliding bar, ranging from 0 to 100, located at the center of the screen. The task for participants was to move the slider up or down through the mouse to indicate the degree of tension they felt while listening to music. Consistent with previous studies (Steinbeis et al., 2006; Lehne et al., 2013), the meaning of tension was not defined by specific descriptions. The initial position of the slider was set at 25 to prevent the rating bar from reaching maximal values throughout the musical progression. The experiment was conducted in an acoustically and electrically shielded room. All stimuli were presented binaurally through Audio Technica CKR30iS headphones. The loudness was adjusted by participants to their comfort levels. Three practice trials were performed before the experimental session to familiarize participants with the stimuli and procedure.

EEG Recording and Analysis

Using Brain Products (Munich, Germany), EEG data were recorded from 64 Ag/AgCl electrodes mounted at International 10–20 system scalp locations. The data were digitized at a rate of 500 Hz with an additional notched filter at 50 Hz. The FCz was used as an online reference electrode, and an electrode placed between Fz and FPz served as the ground electrode. Blinks and vertical eye movements were recorded with an electrode below the right eye. The impedance of all electrodes was maintained less than 5 k Ω and EEG data were amplified with AC amplifiers.

The raw EEG data were preprocessed with EEGLAB (Delorme and Makeig, 2004) in a MATLAB environment. The continuous data were referenced offline to the algebraic mean of the left and right mastoid electrodes. For the ERPs analysis, the data were filtered offline with a band-pass filter of 1–30 Hz. The high-pass filter of 1 Hz was applied to be consistent with previous literature and assure methodological comparison with other studies (e.g., Hu et al., 2014; Kuhn et al., 2015). Then, the data were segmented into epochs of 22.2 s ranging from 1,000 ms before the onset of the first chord, to 2 s after the offset of each chorale sequence. Next, trials were baseline corrected using the 1,000 ms pre-stimulus interval. The data of each participant were then corrected using an Independent Component Analysis (ICA) algorithm (Makeig et al., 1997; Delorme and Makeig, 2004) implemented in EEGLAB to delete ocular and muscle activity-generated artifacts. After the procedure of ICA components removal, we inspected epochs for each participant visually and rejected the contaminated epochs. Epochs in the same experimental condition were averaged for each participant, yielding three average waveforms in each participant.

Single-subject average waveforms were subsequently calculated to obtain global field power (GFP) values at each instant. At any instant, GFP is calculated as the root of the mean of the squared potential differences between each electrode and the mean of the instantaneous potential across electrodes

Figure 1 displays three musical sequences (A, B, and C) in 2/4 time, each consisting of two systems of piano notation. Red brackets indicate hierarchical structures. Sequence A shows a well-formed hierarchy with a single bracket under the first system and a longer one under the second. Sequence B shows phrase violations with two brackets under the first system and two under the second. Sequence C shows a period violation with one long bracket under the first system and one under the second.

FIGURE 1 | Samples of the musical sequences used in the study. Condition 1: a sequence with a well-formed hierarchical structure (A). Condition 2: a sequence with phrase violations (B). Condition 3: a sequence with a period violation (C).

(Lehmann and Skrandies, 1980). To determine whether the GFP values elicited by the three conditions differed as a function of musical hierarchical structure, we performed a point-by-point

repeated measures analysis of variance (RM ANOVA) test. After the significance level (p -value) was corrected for using a false discovery rate (FDR) procedure (Durka et al., 2004), a

time course of p values was obtained. In searching for the time window corresponding to p values below 0.05, we decided on the latency of the EEG responses elicited by the three conditions. Then, within the time window identified as being statistically significant, overall ANOVA tests were followed up with further paired comparisons. Finally, given that the strongest filed potentials and highest topographic signal-to-noise ratios were represented by local maxima of the GFP curve (Lehmann and Skrandies, 1980), we calculated scalp topography around 20 ms, on average, from the GFP peaks and compared differences among the three conditions.

For the time-frequency analysis, the preprocessing steps remained the same as for the ERP analysis, other than filtering with a 1–100 Hz band-pass filter. To obtain time-frequency distributions, Fast Fourier Transforms (FFTs) with a fixed 600 ms Hanning window was applied to each epoch for each participant using time steps of 4 ms and frequency steps of 1 Hz. The time-frequency distributions were corrected with a baseline of an interval from 800 to 200 ms pre-stimulus to include subtle stimulus-induced changes in ongoing oscillatory power. Based on previous findings (Knyazev et al., 2008; Parvaz et al., 2012), we selected alpha band frequencies (8–13 Hz) and nine central-frontal electrodes including Fz, F1, F2, FCz, Fc1, Fc2, C1, Cz, C2 electrodes as the regions of interest (ROIs). A point-by-point RM ANOVA test was performed and the significance level (p -value) was corrected using a FDR procedure (Durka et al., 2004). The time window of statistical significance was defined as that in which the p values were below 0.05. Next, further paired comparisons were individually performed within each significant time window.

RESULTS

Behavioral Results

To minimize the differences in slider moving ranges across participants, data were normalized to a zero mean and unit standard deviation (z -score) for each participant. Tension ratings were averaged across participants under each condition at each time point. **Figure 2** shows average tension profiles under the three conditions, indicating dynamic changes in tension throughout the whole musical pieces. In all conditions, within each phrase, tension values always increased from the beginning to the middle, then declined until the end. Across all three conditions, tension values overlapped only in the first phrase and separated in the other three.

The tension values were averaged for each participant in the following five time-windows: 0.5–4.8, 5.3–9.6, 10.1–14.4, 14.9–19.2, and 19.7–21.2 s (corresponding to the time windows of phrases 1–4 and 1.5 s after the end of sequences). In order to reduce the influence of the response delay, data in the time windows of 0–500 ms after phrase onsets were not analyzed. Two-way RM ANOVA taking time window and condition as within-subject factors was conducted. The results showed that the interaction effect between condition and time window ($F_{(8,184)} = 30.07$; $p < 0.001$, partial $\eta^2 = 0.57$), and the main effects of condition ($F_{(2,46)} = 46.70$; $p < 0.001$, partial

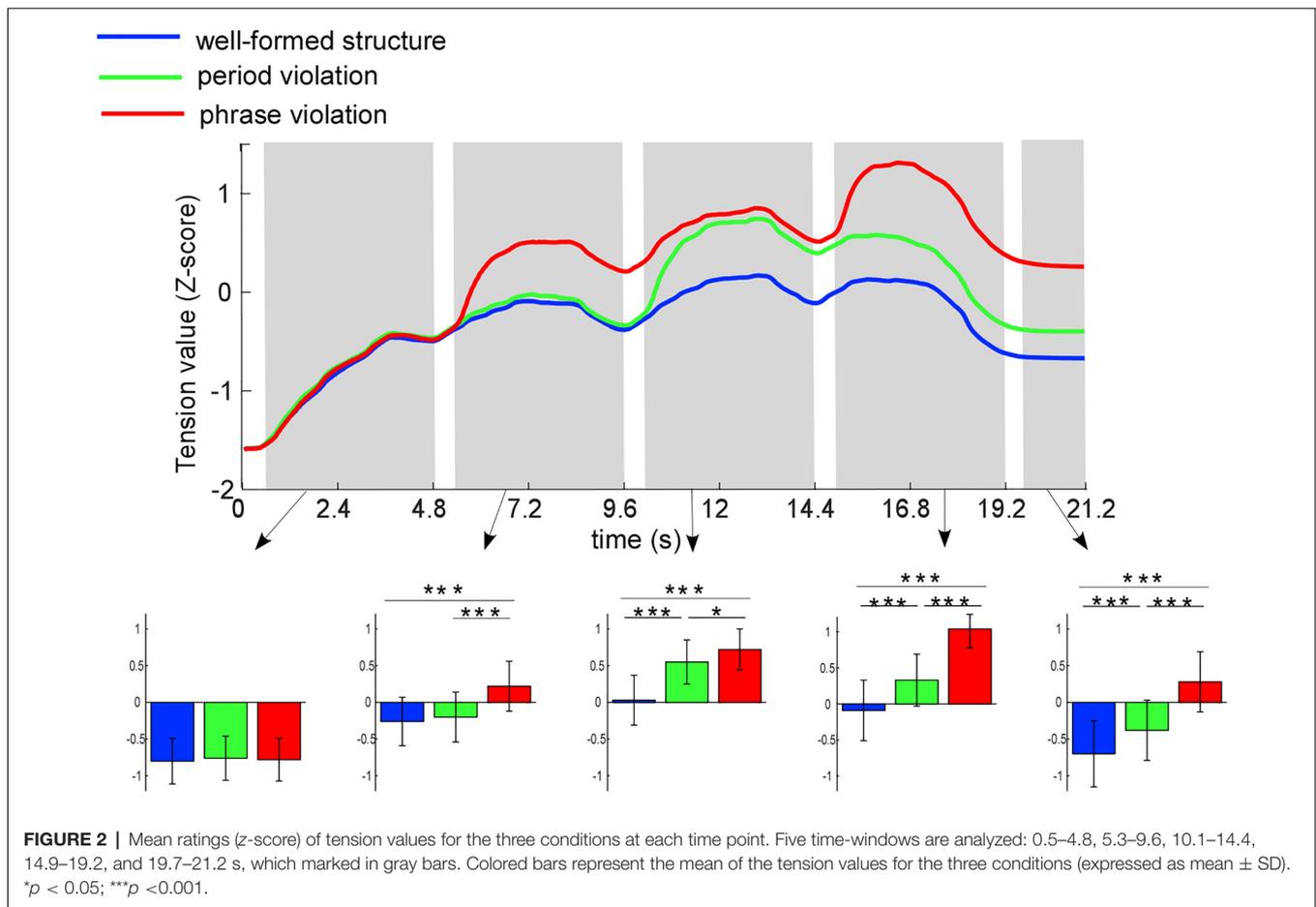
$\eta^2 = 0.67$) and time window ($F_{(4,92)} = 33.41$; $p < 0.001$, partial $\eta^2 = 0.59$) were significant. Then, we conducted ANOVA test taking condition as within-subject factor in each time window. The results revealed significant main effects for all time windows except for the first phrase (the first phrase: $p = 0.08$; the second phrase: $F_{(2,46)} = 32.22$; $p < 0.001$, partial $\eta^2 = 0.58$; the third phrase: $F_{(2,46)} = 42.76$; $p < 0.001$, partial $\eta^2 = 0.65$; the last phrase: $F_{(2,46)} = 45.62$; $p < 0.001$, partial $\eta^2 = 0.67$; post-ending: $F_{(2,46)} = 38.84$; $p < 0.001$, partial $\eta^2 = 0.63$).

Further paired comparisons among the three conditions showed that for the second phrase, more tension was induced by phrase violations than well-formed structures ($p < 0.001$) and period violations ($p < 0.001$); however, no significant difference of tension was found in conditions between well-formed structures and period violations ($p = 0.24$). For the third phrase, more tension was induced by phrase violations and period violations than well-formed structures ($ps < 0.001$), and more tension was induced by period violations than phrase violations ($p = 0.049$). For the last phrase and post-ending, more tension was induced by phrase violations than in the other two conditions ($ps < 0.001$), while more tension was induced by period violations than well-formed structures ($ps < 0.001$).

ERP Results

GFP profiles and scalp topography at the instants corresponding to the GFP peaks under different conditions are shown in **Figure 3**. The segregation of GFP profiles under the three conditions is shown around 700 ms after the onsets of two to four phrases.

Point-by-point RM ANOVA test found that the difference across conditions was significant in the time windows of 5,544–5,590 ms (744–790 ms after the onset of the second phrase), 10,342–10,396 ms (742–796 ms after the onset of the third phrase), and 14,960–15,204 ms (560–804 ms after the onset of the last phrase). Given that frontal-distributed negativities were elicited at the instants corresponding to GFP peaks, further paired comparisons were performed using the average amplitudes at frontal electrodes within the time window of 40 ms around each GFP peak. We conducted a one-way ANOVA taking condition as a within-subject factor in each time window. The results revealed significant main effects for all time windows (5,532–5,572 ms: $F_{(2,46)} = 7.03$; $p = 0.002$, partial $\eta^2 = 0.23$; 10,350–10,390: $F_{(2,46)} = 13.43$; $p < 0.001$, partial $\eta^2 = 0.37$; 15,132–15,172: $F_{(2,46)} = 4.73$; $p = 0.014$, partial $\eta^2 = 0.17$). Results of further paired comparisons showed that, for the time window of 5,532–5,572 ms, larger negativities were elicited by phrase violations than in the other two conditions (well-formed structures, $p = 0.001$; period violations, $p = 0.024$), whereas there was no significant difference in conditions between period violations and well-formed structures ($p = 0.437$). For the time window of 10,350–10,390 ms, larger negativities were elicited by phrase and period violations than in well-formed structures (phrase violations, $p = 0.032$; period violations, $p < 0.001$), while larger negativities were elicited by period violations than phrase violations ($p = 0.047$). For the time window of 15,132–15,172 ms, larger negativities were elicited by phrase violations than in the two other



conditions (well-formed structures, $p = 0.026$; period violations, $p = 0.036$), whereas there was no significant difference in conditions between well-formed structures and period violations ($p = 1.00$).

Time-Frequency Results

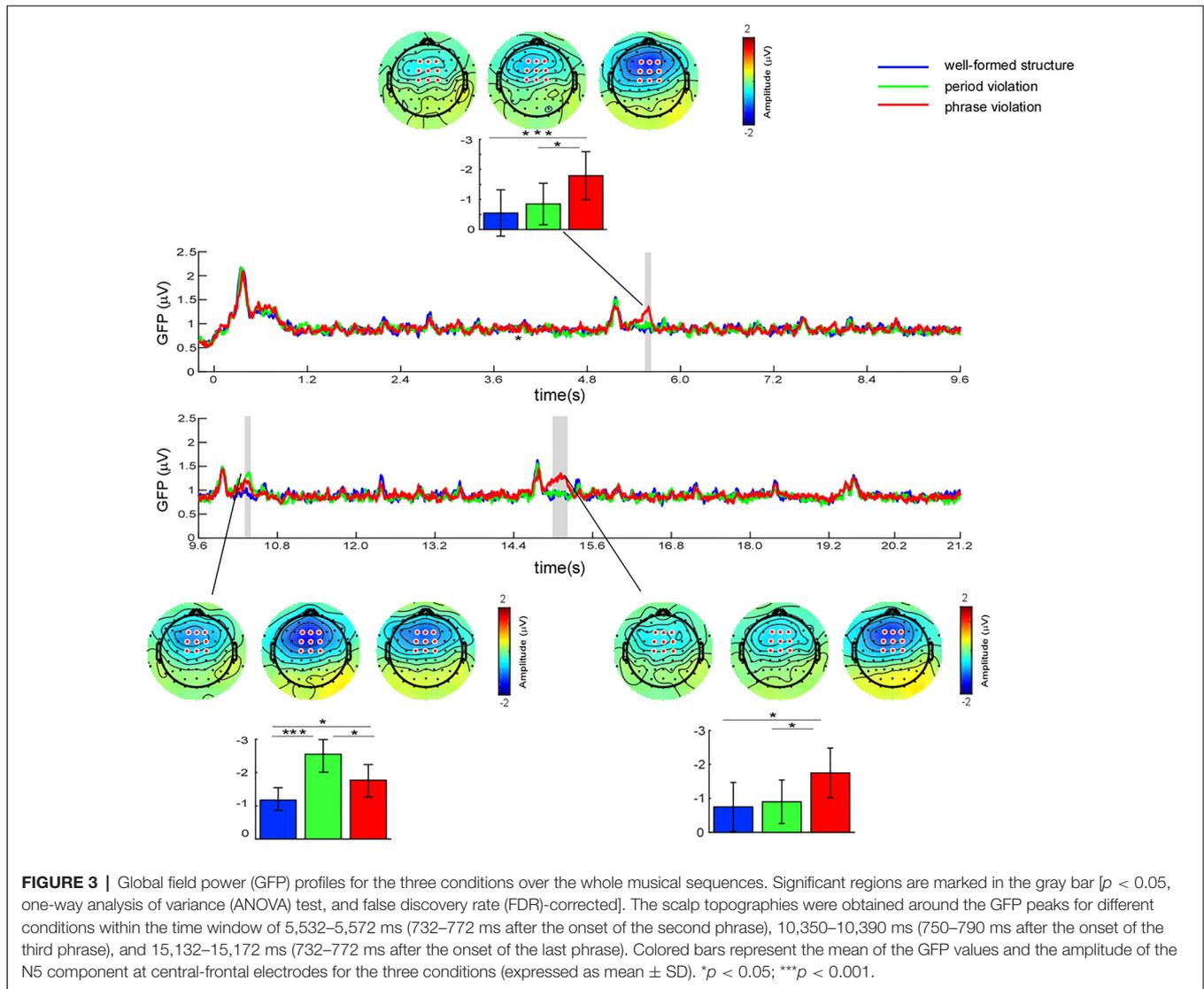
Figure 4 shows the time-frequency responses under different conditions. Significant time windows under different conditions for alpha frequency band (8–13 Hz) power are represented by gray-shaded areas. Scalp topographies at significant instants under different conditions reveal that the decrease in alpha power induced by phrase violations occurred around 800 ms after the onset of phrases 2–4, while this decrease lasted longer to period violations.

Results of a one-way RM ANOVA revealed significant main effects in the time windows of 5,536–5,808 ms (736–1,008 ms after the onset of the second phrase), 10,284–10,432 ms (684–832 ms after the onset of the third phrase), 11,100–11,332 ms (1,500–1,732 ms after the onset of the third phrase), and 15,300–15,532 ms (900–1,132 ms after the onset of the last phrase). For the time window of 5,536–5,808 ms, further paired comparisons indicated that the alpha band power induced by phrase violations was smaller than in the other two conditions (well-formed structures, $p = 0.048$; period violations,

$p = 0.007$), whereas there was no significant difference in conditions between well-formed structures and period violations ($p = 0.29$). For the time window of 10,284–10,432 ms, smaller alpha band power was induced by phrase and period violations than well-formed structures (phrase violations, $p = 0.038$; period violations, $p = 0.034$), whereas there was no significant difference in conditions between phrase violations and period violations ($p = 0.93$). For the time window of 11,100–11,332 ms, smaller alpha band power was induced by period violations than well-formed structures ($p = 0.014$), whereas there was no significant difference between other comparisons ($ps > 0.06$). For the time window of 15,300–15,532 ms, smaller power was induced by phrase violations than in the other two conditions (well-formed structures, $p = 0.050$; period violations, $p = 0.002$), whereas there was no significant difference in conditions between well-formed structures and period violations ($p = 0.51$).

DISCUSSION

The current experiment examined the effects of long-range hierarchical structure on tension experience and its underlying neurocognitive mechanisms. The behavioral data showed that the violations of hierarchical structure induced high tension experience, which was not completely resolved at structural



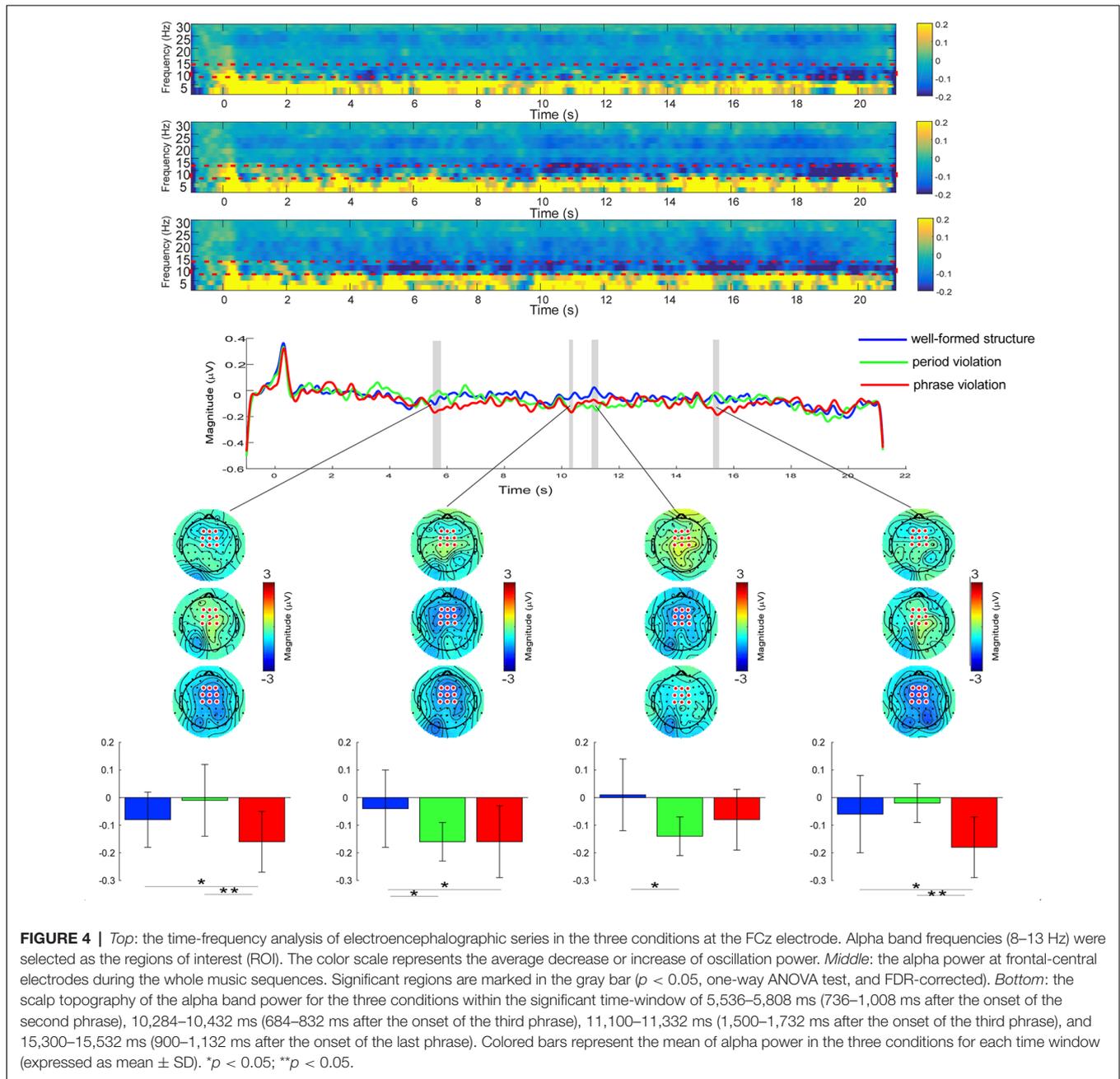
boundaries but accumulated during subsequent musical passages unfolding. The EEG results showed larger frontal-distributed negativities (N5) at GFP peaks, and decreasing power in the alpha band (8–13 Hz) in response to the structural violations. Compared to phrase violations, period violations elicited larger N5 and induced a longer-lasting decrease of power in the alpha band, indicating a hierarchical manner of musical processing. These results are discussed below in more detail.

Dynamic and Accumulating Tension Experience as Music Unfolding

The behavioral tension ratings demonstrated that the tension induced by music is rarely static, but changing constantly over time, in line with a dynamic view of music emotion (Scherer and Moors, 2019). Previous studies investigating music emotion were mainly focused on a limited set of discrete emotion categories (e.g., pleasant vs. unpleasant) to uncover the specific relationship between some musical features and

induced emotion (Menon and Levitin, 2005; Koelsch et al., 2006, 2013; Omigie et al., 2014). The methodology used in these studies cannot show the dynamic processes of musical emotion, therefore is inadequate to characterize the richness of emotion as the music unfolds over time. Our study was concerned with the dynamic process of musical tension as music unfolding in attempting to explore the development of musical tension and resolution.

Three tension curves showed a common trend across experimental conditions, in that musical tension always decreases towards the end of each phrase. The results suggested that musical boundaries were correlated with the resolution, resulting from the dominant or tonic chords in each phrase ending. Based on the relationship between the stability of chords and the strength of the generated tension (Bigand et al., 1996; Steinbeis et al., 2006), the harmonic cadence represented a strong conclusiveness, inducing the sense of resolution (Meyer, 1956; Lerdahl and Jackendoff, 1983).



A significant tension difference across conditions was found in the last three phrases. High tension experience was induced by structural violations, which can be explained by the Gestalt principles. Many aspects of music progression abide by Gestalt laws (Narmour, 1992). The sequences with phrase and period violations did not conform to the Gestalt principle of closure because of the tonality inconsistency. The violation of the Gestalt principle unfulfilled the prediction based on musical context, and in turn induced higher tension experience. In our study, the experimental design was careful and able to disentangle the influence of acoustic features and hierarchical structure on the experience of tension. In the first phrase, there was no significant

difference in tension experience across conditions, although their acoustic features were different. In contrast, in the second phrase, the tension experience differed between sequences with well-formed structure and phrase violations, despite they had the same acoustic features.

The experience of tension was descendent but neither resolved completely nor immediately towards the end of each phrase but rather accumulated throughout subsequent musical pieces. In sequences with phrase violations, the tension induced by each incongruous tonality accumulated and resulted in higher starting points for the subsequent phrases. A similar situation was found in sequences with period violations. Our previous

study investigating the tension experience induced by the nested structures also found the characteristics of tension experience accumulation (Sun et al., 2020a). The results could be explained by the event segmentation theory, which proposes a predictive error-based updating mechanism during the process of event segmentation (Zacks et al., 2007; Kurby and Zacks, 2008). When an unpredicted change occurs, event models are not maintained stable, but rather are updated based on current perceptual information. In our study, structural violations lead to predictive errors and event model resetting, which dynamically engaged working memory and cognitive control processes such as attention switching. This well explains why listeners' felt tension increased at phrase boundaries.

It is worth noting that different patterns of tension and resolution induced by hierarchical structures in the present study and nested structures in the previous study (Sun et al., 2020a). The tension curves showed that more tension arches were induced in hierarchical structures than nested structures, which might be due to the stronger dependence of tonality and more close connection in nested structures. However, it also might be because only one musical phrase was included in sequences used by our previous study (Sun et al., 2020a), whereas four phrases in the present study, implying the important role of the organization of musical units on tension experience. Future studies should further investigate the difference of tension experience induced by nested structures and hierarchical structures to discriminate tension experience delicately and shed new light on structural processing.

The Cognitive Neural Mechanism Underlying the Tension Experience

Given our experiment was the first study to investigate the dynamic EEG underlying musical tension experience, we adopted a data-driven method to determine the significant time windows. The topographical approach of GFP does not require any *a priori* hypothesis, such as the prerequisite of electrodes and time intervals, and therefore is amenable to determining significant time points statistically and reliably. GFP in EEG signal reflects stronger global brain activity, its value represents the strength of the electric potential over all EEG electrodes on the scalp, and has been used in previous studies to measure the global brain activity (Hu et al., 2014; Khanna et al., 2015). In our study, to assess the dynamic process evoked by the whole musical pieces, we analyzed epochs as long as 21.2 s. It would be inadequate to use traditional ERP analytical methods with low high pass filters because it was hard to exclude slow-wave drifts due to the long time-span of the musical pieces. GFP is superior in eliminating the influence of a poor signal-to-noise ratio (Milz et al., 2017), therefore suitable for characterizing rapid changes in brain activity.

It was found that frontal-distributed negativities at GFP peaks always occurred around 600–800 ms after the structural violations, the latency and scalp distributions were similar to the N5 component. Previous studies have suggested that the N5 is related to the violation of expectation and the integration of incongruent harmonic chord into musical context (Koelsch, 2005; Miranda and Ullman, 2007; Steinbeis and Koelsch, 2008;

Koelsch and Jentschke, 2010). In the time-frequency domain, structural violation induced a decrease in central-frontal alpha power. The decreased power in the alpha band may reflect orienting responses (Klimesch, 1999; Krause, 2006) and attention switching (Meyer et al., 2013; Sadaghiani and Kleinschmidt, 2016). In our experiment, listeners constructed a psychological model to predict the upcoming musical events based on musical contexts. Musical violations led to prediction errors, which required the brain to change the comparatively stable model to reduce prediction errors for the upcoming events. Event model resetting engaged more working memory and attention, the neural bases of which were reflected in the decreased power in the alpha band. Furthermore, more cognitive resources were devoted to the integration of harmonic structures, as indicated by the larger N5 component.

It is worth noting that the N5 effects elicited by period violations were larger than phrase violations because it is more difficult for listeners to integrate the period violations into the musical context (Zhang et al., 2018). Furthermore, longer-lasting decreases in alpha power were induced by period violations (around 800 ms to 1,500 ms) compared with the phrase violations (around 800 ms), suggesting more attentional resources were paid for the high-level violations. The results suggested a hierarchical manner of music processing. In line with previous studies in both music and language domains (Ding et al., 2016; Zhang et al., 2016; Harding et al., 2019), our study indicated that structural processing at different hierarchical levels might engage different neural mechanisms. Furthermore, the results supported the predictive coding theory (Friston, 2005, 2010), pointing to the hierarchical predictive coding process inherent to the auditory system, and active construction while listening to music (Koelsch, 2018).

The Relationship Between Tension Ratings and EEG Brain Responses

Compared with well-formed sequences, larger N5 and decreased power in the alpha band were shown when listening to the music sequences with structural violations at phrase and period levels. The effects of these distinct brain responses were in parallel with the changes of behavioral tension ratings, in that the close corresponding relationship is indicated by the similar time windows. The modulation of musical structure on tension ratings and brain responses can be explained by the predictive processes. The mismatch between the actual stimulus and the prediction induced tension experience (Meyer, 1956; Margulis, 2005; Huron, 2006), which required listeners to pay more cognitive resources to update working memory, switch attention (Zacks et al., 2007; Kurby and Zacks, 2008) and integrate the violated events into a musical context, as implied by large N5 and decrease power in the alpha band.

Although tension ratings and brain responses induced by music structural violations are closely related, the processing processes they reflect are different. In particular, tension ratings reflected the online tension experience, while the brain responds to the integrated processing of the musical event into the musical context. Thus, the tension rating induced by each structural violation was based on the ending point in the

previous phrase. That is, the tension value in one moment represented the tension induced by the present musical event and the accumulated tension from the previous musical context. In contrast, the N5 and the power in the alpha band did not exhibit the accumulation characteristics, but fading away within one phrase. The strength of the brain responses only reflected the cognitive effort to integrate the present musical event into the musical context.

In conclusion, the present study using behavioral rating and EEG investigated the influence of music hierarchical structure on the experience of tension and its underlying neural mechanism. Behavioral results showed that structural violations gave rise to an increasing and accumulating tension experience as indicated by tension curves. In parallel, brain responses including larger N5 effects at GFP peaks and decreased power in alpha frequency band were in response to the structural violations, reflecting more attention and cognitive resources engaged in the integrated processing. Moreover, compared with phrase violations, period violations elicited larger N5 and induced a longer-lasting decrease in alpha power, revealing the hierarchical manner of structure processing in music. The dynamic tension experience revealed by our findings is of significance to emotional regulation. Furthermore, the hierarchical manner of structure processing also reminded musicians to pay more attention to large-scale music structure while listening to music.

REFERENCES

- Bigand, E., and Parncutt, R. (1999). Perceiving musical tension in long chord sequences. *Psychol. Res.* 62, 237–254. doi: 10.1007/s004260050053
- Bigand, E., Parncutt, R., and Lerdahl, F. (1996). Perception of musical tension in short chord sequences: the influence of harmonic function, sensory dissonance, horizontal motion and musical training. *Percept. Psychophys.* 58, 124–141.
- Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009
- Ding, N., Melloni, L., Zhang, H., Tian, X., and Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nat. Neurosci.* 19, 158–164. doi: 10.1038/nn.4186
- Durka, P. J., Zygierevicz, J., Klekowicz, H., and Ginter, J. (2004). On the statistical significance of event-related EEG desynchronization and synchronization in the time-frequency plane. *IEEE Trans. Biomed. Eng.* 51, 1167–1175. doi: 10.1109/TBME.2004.827341
- Friston, K. (2005). A theory of cortical responses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 815–836. doi: 10.1098/rstb.2005.1622
- Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11:127. doi: 10.1038/nrn2787
- Friston, K., and Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 1211–1221. doi: 10.1098/rstb.2008.0300
- Harding, E., Sammler, D., Henry, M. J., Large, E. W., and Kotz, S. A. (2019). Cortical tracking of rhythm in music and speech. *NeuroImage* 185, 96–101. doi: 10.1016/j.neuroimage.2018.10.037
- Hu, L., Valentini, E., Zhang, Z. G., Liang, M., and Iannetti, G. D. (2014). The primary somatosensory cortex contributes to the latest part of the cortical response elicited by nociceptive somatosensory stimuli in humans. *NeuroImage* 84, 383–393. doi: 10.1016/j.neuroimage.2013.08.057
- Huron, D. (2006). *Sweet Anticipation: Music and the Psychology of Expectation*. Cambridge, MA: MIT Press.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Board of the Institute of Psychology, Chinese Academy of Sciences. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

LS and YY proposed the study and designed the experiment. LS and GR conducted the tests. LS and LH contributed to data analysis. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Natural Science Foundation of China (Grant Nos. 31470985 and 31971034) and the China Postdoctoral Science Foundation (Y9BH032001).

- Juslin, P. N. (2013). From everyday emotions to aesthetic emotions: towards a unified theory of musical emotions. *Phys. Life Rev.* 10, 235–266. doi: 10.1016/j.plrev.2013.05.008
- Juslin, P. N., and Laukka, P. (2004). Expression, perception and induction of musical emotions: a review and a questionnaire study of everyday listening. *J. New Music Res.* 33, 217–238. doi: 10.1080/0929821042000317813
- Juslin, P. N., and Västfjäll, D. (2008). Emotional responses to music: the need to consider underlying mechanisms. *Behav. Brain Sci.* 31, 559–575. doi: 10.1017/S0140525X08005293
- Khanna, A., Pascual-Leone, A., Michel, C. M., and Farzan, F. (2015). Microstates in resting-state EEG: current status and future directions. *Neurosci. Biobehav. Rev.* 49, 105–113. doi: 10.1016/j.neubiorev.2014.12.010
- Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Res. Brain Res. Rev.* 29, 169–195. doi: 10.1016/s0165-0173(98)00056-3
- Knyazev, G. G., Bocharov, A. V., Levin, E. A., Savostyanov, A. N., and Slobodskoj-Plusnin, J. Y. (2008). Anxiety and oscillatory responses to emotional facial expressions. *Brain Res.* 1227, 174–188. doi: 10.1016/j.brainres.2008.06.108
- Koelsch, S. (2005). Neural substrates of processing syntax and semantics in music. *Curr. Opin. Neurobiol.* 15, 207–212. doi: 10.1016/j.conb.2005.03.005
- Koelsch, S. (2012). *Brain and Music*. Oxford, UK: John Wiley and Sons.
- Koelsch, S. (2014). Brain correlates of music-evoked emotions. *Nat. Rev. Neurosci.* 15, 170–180. doi: 10.1038/nrn3666
- Koelsch, S. (2018). Investigating the neural encoding of emotion with music. *Neuron* 98, 1075–1079. doi: 10.1016/j.neuron.2018.04.029
- Koelsch, S., and Jentschke, S. (2010). Differences in electric brain responses to melodies and chords. *J. Cogn. Neurosci.* 22, 2251–2262. doi: 10.1162/jocn.2009.21338
- Koelsch, S., Fritz, T., Cramon, D. Y., Müller, K., and Friederici, A. D. (2006). Investigating emotion with music: an fMRI study. *Hum. Brain Mapp.* 27, 239–250. doi: 10.1002/hbm.20180

- Koelsch, S., Skouras, S., Fritz, T., Herrera, P., Bonhage, C., Küssner, M. B., et al. (2013). The roles of superficial amygdala and auditory cortex in music-evoked fear and joy. *NeuroImage* 81, 49–60. doi: 10.1016/j.neuroimage.2013.05.008
- Krause, C. M. (2006). Cognition- and memory-related ERD/ERS responses in the auditory stimulus modality. *Prog. Brain Res.* 159, 197–207. doi: 10.1016/s0079-6123(06)59013-2
- Kuhn, A., Brodbeck, V., Tagliazucchi, E., Morzelewski, A., von-Wegner, F., and Laufs, H. (2015). Narcoleptic patients show fragmented EEG-microstructure during early nrem sleep. *Brain Topogr.* 28, 619–635. doi: 10.1007/s10548-014-0387-1
- Kurby, C. A., and Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends Cogn. Sci.* 12, 72–79. doi: 10.1016/j.tics.2007.11.004
- Lehmann, D., and Skrandies, W. (1980). Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroencephalogr. Clin. Neurophysiol.* 48, 609–621. doi: 10.1016/0013-4694(80)90419-8
- Lehne, M., and Koelsch, S. (2015a). “Tension-resolution patterns as a key element of aesthetic experience: psychological principles and underlying brain mechanisms,” in *Art, Aesthetics and the Brain*, eds J. P. Huston, M. Nadal, F. Mora, L. F. Agnati, and C. J. C. Conde (Oxford: Oxford University Press).
- Lehne, M., and Koelsch, S. (2015b). Toward a general psychological model of tension and suspense. *Front. Psychol.* 6:79. doi: 10.3389/fpsyg.2015.00079
- Lehne, M., Rohrmeier, M., Gollmann, D., and Koelsch, S. (2013). The influence of different structural features on felt musical tension in two piano pieces by mozart and mendelssohn. *Music Percept.* 31, 171–185. doi: 10.1525/mp.2013.31.2.171
- Lerdahl, F., and Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.
- Lerdahl, F., and Krumhansl, C. L. (2007). Modeling tonal tension. *Music Percept.* 24, 329–366. doi: 10.1525/mp.2007.24.4.329
- Loui, P., Grent, T., Torpey, D., and Woldorff, M. (2007). Effects of attention on the neural processing of harmonic syntax in western music. *Cogn. Brain Res.* 25, 678–687. doi: 10.1016/j.cogbrainres.2005.08.019
- Makeig, S., Jung, T. P., Bell, A. J., Ghahremani, D., and Sejnowski, T. J. (1997). Blind separation of auditory event-related brain responses into independent components. *Proc. Natl. Acad. Sci. U S A* 94, 10979–10984. doi: 10.1073/pnas.94.20.10979
- Margulis, E. H. (2005). A model of melodic expectation. *Music Percept.* 22, 663–714. doi: 10.1525/mp.2005.22.4.663
- Menon, V., and Levitin, D. J. (2005). The rewards of music listening: response and physiological connectivity of the mesolimbic system. *NeuroImage* 28, 175–184. doi: 10.1016/j.neuroimage.2005.05.053
- Meyer, L. B. (1956). *Emotion and Meaning in Music*. Chicago, IL: University of Chicago Press.
- Meyer, L., Obleser, J., and Friederici, A. D. (2013). Left parietal enhancement during working memory-intensive sentence processing. *Cortex* 49, 711–721. doi: 10.1016/j.cortex.2012.03.006
- Milz, P., Pascualmarqui, R. D., Achermann, P., Kochi, K., and Faber, P. L. (2017). The EEG microstate topography is predominantly determined by intracortical sources in the alpha band. *NeuroImage* 162, 353–361. doi: 10.1016/j.neuroimage.2017.08.058
- Miranda, R. A., and Ullman, M. T. (2007). Double dissociation between rules and memory in music: an event-related potential study. *NeuroImage* 38, 331–345. doi: 10.1016/j.neuroimage.2007.07.034
- Narmour, E. (1992). *The Analysis and Cognition of Basic Melodic Structures: The Implication-Realization Model*. Chicago, IL: University of Chicago Press.
- Omigie, D., Dellacherie, D., Hasboun, D., George, N., Clement, S., Adam, C., et al. (2014). An intracranial EEG study of the neural dynamics of musical valence processing. *Cereb. Cortex* 25, 4038–4047. doi: 10.1093/cercor/bhu118
- Parvaz, M. A., Macnamara, A., Goldstein, R. Z., and Hajcak, G. (2012). Event-related induced frontal alpha as a marker of lateral prefrontal cortex activation during cognitive reappraisal. *Cogn. Affect. Behav. Neurosci.* 12, 730–740. doi: 10.3758/s13415-012-0107-9
- Poulin-Charronnat, B., Bigand, E., and Koelsch, S. (2006). Processing of musical syntax tonic versus subdominant: an event-related potential study. *J. Cogn. Neurosci.* 18, 1545–1554. doi: 10.1162/jocn.2006.18.9.1545
- Sadaghiani, S., and Kleinschmidt, A. (2016). Brain networks and α -oscillations: structural and functional foundations of cognitive control. *Trends Cogn. Sci.* 20, 805–817. doi: 10.1016/j.tics.2016.09.004
- Scherer, K. R., and Moors, A. (2019). The emotion process: event appraisal and component differentiation. *Annu. Rev. Psychol.* 70, 719–745. doi: 10.1146/annurev-psych-122216-011854
- Steinbeis, N., and Koelsch, S. (2008). Shared neural resources between music and language indicate semantic processing of musical tension-resolution patterns. *Cereb. Cortex* 18, 1169–1178. doi: 10.1093/cercor/bhm149
- Steinbeis, N., Koelsch, S., and Sloboda, J. A. (2006). The role of harmonic expectancy violations in musical emotions: evidence from subjective, physiological and neural responses. *J. Cogn. Neurosci.* 18, 1380–1393. doi: 10.1162/jocn.2006.18.8.1380
- Sun, L., Feng, C., and Yang, Y. (2020a). Tension experience induced by nested structures in music. *Front. Hum. Neurosci.* 14:210. doi: 10.3389/fnhum.2020.00210
- Sun, L., Thompson, W. F., Liu, F., Zhou, L., and Jiang, C. (2020b). The human brain processes hierarchical structures of meter and harmony differently: evidence from musicians and nonmusicians. *Psychophysiology* 57:e13598. doi: 10.1111/psyp.13598
- Swaminathan, S., and Schellenberg, E. G. (2015). Current emotion research in music psychology. *Emot. Rev.* 7, 189–197. doi: 10.1177/1754073914558282
- Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., and Reynolds, J. R. (2007). Event perception: a mind-brain perspective. *Psychol. Bull.* 133, 273–293. doi: 10.1037/0033-2909.133.2.273
- Zhang, J., Jiang, C., Zhou, L., and Yang, Y. (2016). Perception of hierarchical boundaries in music and its modulation by expertise. *Neuropsychologia* 91, 490–498. doi: 10.1016/j.neuropsychologia.2016.09.013
- Zhang, J., Zhou, X., Chang, R., and Yang, Y. (2018). Effects of global and local contexts on chord processing: an ERP study. *Neuropsychologia* 109, 149–154. doi: 10.1016/j.neuropsychologia.2017.12.016

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Sun, Hu, Ren and Yang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.