Time-varying similarity of neural responses to musical tension is shaped by physical features and musical themes

Article

Accepted Version

Time-varying Similarity of Neural Responses to Musical Tension is Shaped by Physical Features and Musical Themes

Chao Xue [1], Yiran Chen[2], William Forde Thompson[3], Fang Liu[4], and Cunmei Jiang[5*]

[1] Department of Psychology, Shanghai Normal University, Shanghai, 200234, China

[2] Research Institute of the McGill University Health Centre, Montreal, QC, Canada

[3] Bond University, Gold Coast, QLD, 4226, Australia

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

[5] Music College, Shanghai Normal University, Shanghai, 200234, China

* Corresponding author:

Cunmei Jiang, Music College, Shanghai Normal University, No.100 Guilin Rd. Shanghai, China, 200234.

E-mail: cunmeijiang@shnu.edu.cn

Abstract

The similarity of understanding is important for music experience and communication, but little is understood about the sources of this common knowledge. Although neural responses to the same piece of music are known to be similar across listeners, it remains unclear whether this neural response similarity is linked to musical understanding and the role of dynamic musical attributes in shaping it. Our study addresses this gap by investigating the relationship between neural response similarity, musical tension, and dynamic musical attributes. Using electroencephalography-based inter-subject correlation (EEG-ISC), we examined how the neural response similarity among listeners varies throughout the evaluation of musical tension in the first movement of Beethoven's Piano Sonata No. 8. Participants continuously rated the degree of alignment between musical events and their expectations, while neural activity was recorded using electroencephalography (EEG). The results showed that neural response similarity fluctuated in tandem with musical tension, with increased similarity observed during moments of heightened tension. This time-varying neural response similarity was influenced by two dynamic attributes contributing to musical tension: physical features and musical themes. Specifically, its fluctuation was driven by physical features, and the patterns of its variation were modulated by musical themes, with similar time-varying patterns observed across similar thematic materials. These findings offer valuable insight into the role of dynamic musical attributes in shaping neural response similarity, and reveal an important source and mechanism of shared musical understandings.

Keywords: Similarity in understanding; EEG-ISC method; Musical tension; Physical features; Musical themes

1. Introduction

Musical understanding is a complex and multidimensional process that involves comprehending and interpreting various aspects of music, including its structural elements, expressive qualities, historical context, and cultural significance, and intertwines with listeners' cognitive, aesthetic, emotional responses (Hallam & Papageorgi, 2016; Martin, 1966). Thus, it is considered to be idiosyncratic (Kopiez, 2006). However, similarity in understanding is an important aspect in comprehending music that provides a basis for a deeper and more effective musical communication. To date, there is little understanding of the sources of this common knowledge.

The similarity in understanding is indicative of intersubjectivity (Margulis et al., 2022; Reynaert, 2001), which refers to the shared understandings (Göncü, 1993), expectations (Garfinkel, 1967), or subjective states (Scheff et al., 2006) among multiple people in a community. The process by which humans achieve intersubjectivity has been discussed in fields such as philosophy and social cognition, and has long been assumed to depend on intersubjective interactions. In particular, Edmund Husserl, a founder of phenomenology, proposed that intersubjectivity depends on empathy, the process by which individuals come to understand the experiences of other people (Husserl, 1977). Indeed, shared understandings can arise through social interactions between infants and caregivers (Schore, 2021; Terrace et al., 2022; Trevarthen, 2010) and between social actors (Raymond, 2019; Stone et al., 2012). Moreover, the mirror neuron and mentalizing systems (Vogeley, 2017) and the default mode network (Marchetti & Koster, 2014) have been proposed as the neural bases of intersubjective interactions.

While social interactions are clearly involved in intersubjectivity, philosophers with an ontological perspective have posited that the external world can drive intersubjectivity by providing a shared context (Buber, 2012; Heidegger, 2010) or focus (Schutz, 1972). In fact, interpersonal interaction can establish a shared sense of feelings between individuals in response to a common external stimulus (Echterhoff et al., 2009; Higgins et al., 2021). Others' attitudes toward a shared stimulus facilitate similar feelings between individuals (Higgins et al., 2007). When adopting the same perspective to watch a video (Lahnakoski et al., 2014), and providing similar interpretations of an ambiguous video (Nguyen et al., 2019), individuals show increased similarity in neural responses. These studies suggest that a shared stimulus is beneficial for establishing intersubjectivity. The world is dynamic, however, and this overall neural response similarity is insufficient to explain how a dynamically changing stimulus produces neural response similarity across individuals throughout the entire process of understanding.

Musical events unfold over time. The lack of explicit semantic information in music poses a challenge in achieving musical intersubjectivity through interactions. The central focus, therefore, is on how the same music piece produces a similar understanding of music. Previous studies have shown that listeners of the same culture tend to provide similar imagined narratives to the same musical excerpts (Margulis et al., 2022; McAuley et al., 2021). Alfred Schütz, a social phenomenologist, proposed that musical intersubjectivity is formed by a shared flux of inner time musical experience between composers, performers, and listeners (Schütz, 1951). Such a shared flux of musical experience may be associated with dynamic musical features. Therefore, it is crucial to examine how dynamically changing music produces similarities in understanding throughout the entire process of musical listening.

The Inter-subject correlation (ISC) method offers a unique way to explore the similarity of neural activity evoked by a stimulus across multiple individuals (Dmochowski et al., 2012; Hasson et al., 2004). While functional magnetic resonance imaging (fMRI)-based ISC (fMRI-ISC) research on emotion suggests that neural synchronization is influenced by musical features (Sachs et al., 2020; Trost et al., 2015), electroencephalography (EEG)-based ISC (EEG-ISC) research has shown an overall similarity of neural responses during listening to natural music (Madsen et al., 2019) and music retaining basic features (Kaneshiro et al., 2020). To date, only two studies have focused on the dynamic features of music and demonstrated a time-varying neural response similarity in response to music (Dauer et al., 2021; Kaneshiro et al., 2021) within groups of listeners, including large numbers of trained musicians.

However, since no neural markers of musical understanding were identified in these studies, any neural response similarity could be attributed to similar perceptual inputs rather than comparable musical understandings. Thus far, how dynamically changing music influences the time-varying similarities of neural responses associated with understanding remains uncertain.

The present study used EEG-ISC to examine how neural response similarity among listeners to the same music piece varies throughout the entire evaluation of musical tension. Musical tension is an affective state that arises from expectations related to various musical elements (Krumhansl, 2002; Lehne & Koelsch, 2015). It evokes specific emotional and cognitive responses in listeners (Koelsch, 2012), and serves as a fundamental psychological experience reflecting musical understanding (Huron, 2006; Meyer, 1956). Therefore, this study focused on the neural response similarity produced by this experience. We initially evaluated the neural response similarity in response to natural music while ensuring its association with the experience of musical tension, which served as the cornerstone of our study. Since physical features and themes are two primary dynamic aspects that contribute to musical tension, we focused on these aspects to investigate how these factors influence the evolving similarities of neural responses associated with the experience of musical tension.

Accordingly, our study opted for five changing physical features that have been shown to contribute to creating musical tension (Lartillot, 2019), namely root mean square (RMS), fluctuation peaks, key clarity, harmonic change detection function (HCDF), and novelty. RMS and fluctuation peaks refer to dynamics and rhythm, respectively (Moore, 2012; Pampalk et al., 2002), while key clarity and HCDF are related to tonal information (Degani et al., 2015; Gómez, 2006; Krumhansl, 2001; Saari et al., 2013), and novelty is a specific structural description of the temporal progression of moments (Foote & Cooper, 2003). All physical features, except for novelty, function at a low level.

Although these physical features contribute to generating musical tension, they may not have any cognitive interpretation. Conversely, themes serve as basic structural units that can convey the essence of music. They offer an additional dimension by capturing dynamic changes in music through repetitions and variations. Variations in themes lead to experiences of different patterns of dynamic tension at the structural level, as these themes include a blend of elemental traits like melody, rhythm, and harmony. In order to examine how these themes influence changes in neural similarities during musical comprehension, we selected the first movement of Beethoven's Piano Sonata No. 8 as the musical stimulus. This choice was based on the presence of primary and secondary themes within it. These themes serve as the core materials, and carry the main ideas of music, but differ from each other in style and key. In this sonata, themes are repeated, and varied, and act as the building blocks for larger musical structures, such as exposition, development, and recapitulation.

By recognizing the roles of shared context in establishing intersubjectivity (Buber, 2012; Heidegger, 2010) or focus (Schutz, 1972), it is expected that the same dynamically changing musical piece would produce the time-varying similarities of neural responses, associated with the musical tension experienced by listeners. More importantly, since musical tension ratings are influenced by musical elements, such as RMS, fluctuation peaks, key clarity, HCDF, and novelty (Lartillot, 2019), these neural similarities would be driven by musical features. Likewise, the patterns of neural similarities would be affected by musical themes, as variations in themes reflect different patterns of dynamic tension experiences at the structural level.

2. Materials and Methods

2.1. Participants

This study included 41 right-handed, non-musician, native Mandarin Chinese speakers (21 females and 20 males, age M = 23.56 years, SD = 1.95). None of them had received extracurricular music training or reported any neurological, hearing, or psychological disorders. Due to the possibility of individuals with musical anhedonia among non-musicians (Martínez-Molina et al., 2016; Mas-Herrero et al., 2013), participants were selected based on self-reports of enjoying listening to music. The Edinburgh

Handedness Inventory (Oldfield, 1971) was used to confirm the right-handedness of all participants. This study was approved by the Institutional Review Board of Shanghai Normal University (Shanghai, China). All participants provided written informed consent and were remunerated for their participation.

2.2. Stimuli

The first movement of Ludwig van Beethoven's Piano Sonata No. 8 in C minor, Op. 13, performed by Emil Gilels (with a duration of 9 min and 9 s) (Beethoven, 1981), was chosen as the musical stimulus. The stimulus was obtained from the NetEase CloudMusic platform (https://music.163.com). As shown in Table 1, the selected musical piece followed a typical sonata form, consisting of 19 thematic sections with three main parts (exposition, development, and recapitulation) and two framing modules (introduction and coda). To control for the overall neural response similarity (measured by EEG-ISC) arising from physical stimulus features rather than the structural element of the music stimulus, a phase-scrambled version (hereafter referred to as "phase") was created by preserving spectral density but not time-dependent fluctuations of the original sonata (hereafter referred to as "original") to serve as a control version. Following a previous study (Abrams et al., 2013), we used a Fourier transform to the original stimulus and then randomized the phase of each frequency by randomly shifting the phase between 0 and 2π (Prichard & Theiler, 1994). This procedure allows us to preserve the magnitude spectra of the phase stimulus while disrupting time-dependent fluctuations. The waveform characteristics of the original and phase stimuli are shown in Figure 1.

Insert Table 1, about here.

Insert Figure 1, about here.

2.3. Procedure

The experimental stimuli were presented to participants through Edifier R1200T speakers (Beijing Edifier Technology Company, Ltd., Beijing, China) for the EEG session and Philips SHM7410 headphones (Philips N.V., Amsterdam, the Netherlands) for the behavioral session. Prior to the experiment, participants were allowed to adjust the loudness of the stimuli to their individual comfort level. They completed both EEG and behavioral sessions, with the EEG session always preceding the behavioral one. The sessions were not randomized in order because repeated listening can diminish EEG-ISC (Madsen et al., 2019), while musical repetitions have minimal impact on tension ratings (Bigand & Parncutt, 1999).

All participants were non-musicians and unaware of the concept of musical tension. Consequently, we asked participants to evaluate the fitness between musical events and their expectations to assess musical tension, as the interaction between anticipation and auditory events is pivotal in its generation (Krumhansl, 2002). Specifically, when musical events are in line with listeners' expectations, they can produce feelings of satisfaction and resolution. Conversely, when musical events deviate from listeners' expectations, they can cause a sense of tension or uncertainty (Krumhansl, 2002; Lehne & Koelsch, 2015). Therefore, participants were informed that deviations from their expectations in musical events might lead to sensations of conflict, instability, dissonance, or uncertainty, and *vice versa*. During the EEG recording, participants listened to two stimuli (the original and phase stimuli) in a randomized order, and were instructed to continuously rate the degree of fitness between the musical events and their expectations in mind. Following the EEG recording, the behavioral session started after a short break (about 20-30 min, to prevent fatigue). Participants were instructed to listen to the same stimuli in the same order again, and continuously rate the degree of fitness between the musical events and their expectations using the Continuous Affect Rating and Media Annotation (CARMA) software (Girard, 2014) on a slider scale ranging from -100 (*fit poorly, the far down end of the slider*) to 100 (*fit* *well, the far top end of the slider*). Participants were encouraged to use the full range of the response scale for both sessions. Continuous fitness ratings were recorded at a sampling rate of 4 Hz.

2.4. EEG acquisition and preprocessing

Electrical brain activity was recorded with a sampling rate of 1000-Hz (high pass 0.05 Hz, low pass 100 Hz) using a Neuroscan Quick-Cap (Compumedics Inc., Charlotte, NC, USA) with 64 electrodes to record the EEG activity with the Neuroscan (version 4.3.2) software and a Synamps2 amplifier (Compumedics Inc.). Standard electrode sites were used according to the extension of the international 10-20 system. Impedances were set to be below 10 kΩ for all electrodes. Four electrodes were used to measure horizontal electrooculogram (EOG; placed on the outer canthus of the left and right eyes) and vertical EOG (placed above and below the left eye). All electrodes were referenced to the left mastoid, and a forehead ground was used.

Preprocessing of EEG data was performed with the EEGLAB toolbox v12.0.2 (Delorme & Makeig, 2004) for Matlab 2013b (The Mathworks Inc., Natick, MA, USA). The EEG data were first down-sampled to 250 Hz and then filtered between 1 and 45 Hz with notch (49-51 Hz) using the delay corrected (zero-phase) finite-impulse-response (FIR) filter implemented in the EEGLAB function pop_eegfiltnew. All sensors over the face and four electrodes (M1, M2, CB1, CB2) were excluded from the analysis. The EEG data were then converted to the average reference. Epochs for each stimulus were 9:09 minutes in length, during which eye movement and other artifacts were identified by independent component analysis (ICA) using the Infomax algorithm (Jung et al., 1997), and then removed manually by selecting the component that contributed to the artifact. On average, 2.88 ± .64 components out of 60 ICA components were removed. EEG values with a squared magnitude greater than four standard deviations of their respective channel's mean power were identified as outliers and replaced with Not a Number (NaN).

2.5. Extraction of musical features

To evaluate the contribution of changes in the musical features of the sound to neural synchronization and behavioral tension ratings, five musical features of the original music stimulus were extracted using the MIR toolbox (Lartillot & Toiviainen, 2007) implemented in Matlab (The Mathworks Inc.). The five features were chosen based on earlier findings indicating their impacts on the tension score (Lartillot, 2019). The RMS calculates the signal's immediate energy by taking the square root of the sum of its amplitude squares. Fluctuation peaks estimate the rhythmic periodicity obtained from the spectral analysis of each band of the spectrogram (Pampalk et al., 2002). Key clarity represents the key strength associated with the best key. HCDF measures the flux of the tonal centroid (Harte et al., 2006). Novelty characterizes a specific structural description focused on the temporal progression of moments, each of which has unique melodic characteristics (Foote & Cooper, 2003). All musical features were extracted using a frame-by-frame analysis method to obtain continuous measures of these musical features.

2.6. Data analysis

Reliable Components Analysis. For EEG-ISC calculation, we spatially filtered the EEG data using Reliable Components Analysis (RCA) prior to computing ISC (Dmochowski et al., 2012) to reduce the large data dimensionality and increase the signal-to-noise ratio. RCA involves identifying linear combinations of electrodes that show maximal correlation across participants, which results in the transformation of the electrode-by-time matrices of EEG data into component-by-time matrices. The returned components are ranked by explained reliability in descending order; with the first component, RC1, having the highest ISC in the component-space data, followed by RC2, RC3, *etc*. According to Dmochowski *et al*. (2012), we computed the first three reliable components (RC1-RC3). Further analysis revealed that the ISC values for RC2 and RC3 were not statistically significant compared to the null distribution generated through permutation testing (Figure S1). As such, following Kaneshiro *et al*. (2020) and Dauer *et al*. (2021), we focused our analysis only on the ISC results from the RC1 data. We used a publicly available Matlab implementation (Dmochowski et al., 2015) to compute the RCA and displayed individual components as scalp topographies using forward-model projections of the weight vectors (Parra et al., 2005), as reported by Kaneshiro et al. (2020).

Inter-subject Correlation. On a per-stimulus basis, ISCs in the EEG data were calculated across participants. We first calculated the ISC over the whole length of the stimulus for individual subjects (*i.e.*, Overall ISC). We computed the time-varying ISC in running 5 s windows at 1 s increment following previously described procedures (Dauer et al., 2021; Dmochowski et al., 2012; Kaneshiro et al., 2021). In order to calculate their correlations, the continuous behavioral fitness rating and the frame-to-frame musical features were also down sampled in running 5 s windows at 1 s increments to match the time-varying EEG-ISC. The ISC values presented in this study reflect the degree of concordance between individuals. The ISC computation is identical to previously published methods and may be duplicated using code from http://www.parralab.org/isc/.

The significance of ISC values was assessed using a permutation test (Theiler et al., 1992). As described in detail in previous studies (Dauer et al., 2021; Kaneshiro et al., 2020, 2021), this method involves generating surrogate data by phase-scrambling. We then performed RCA of over 100 different surrogated EEG data to create the null distribution for each stimulus, as previously described (Chang et al., 2015; Cohen & Parra, 2016; Ki et al., 2016). The threshold for statistical significance was set at the 95th percentile. Notably, considering the autocorrelation in the phase-scrambled data accounts for temporal dependencies (Prichard & Theiler, 1994; Theiler et al., 1992), we did not use any cluster correlation on the time-varying ISC.

Relationship between dynamic neural response similarity and musical theme. We examined the effects of the musical theme, focusing on the original version of the music. As the durations of the different musical sections varied, we used *k*-means clustering to divide the EEG-ISC into a smaller number of levels (*i.e.*, clusters of *k*-means clustering) for follow-up comparisons. This clustering was based on EEG-ISC values in conjunction with their 95th percentile threshold, as these parameters collectively represent the magnitude of ISC. In this way, similar EEG-ISC time windows are classified into the same cluster. The optimal cluster number solution (*i.e.*, number of levels or k) was determined by the Elbow method. The Elbow method was used to perform k-means clustering on the dataset using a range of values for k. For each value of k, the sum of squared errors was calculated. The optional number was determined by identifying the position of the elbow (here the number of levels or k = 4). Based on these levels, we constructed a 4 (levels) × 19 (musical sections) contingency table. We then conducted a Fisher's exact test to determine whether the level of ISC could imply the musical structure. Since hierarchical clustering analysis (HCA) is the most common statistical method for identifying homogeneous groups of cases based on measured features (Zhang et al., 2017), we also used Ward's method to perform an HCA on the Euclidean distances of the proportion levels of EEG-ISC. This was undertaken to confirm whether the 19 sections could be grouped into music section clusters (the term *music section clusters* is used here to distinguish it from k-means clusters) (Hair et al., 1995). In line with Bergman (1998), the determination of the number of music section clusters was guided by considerations of interpretability and ease of use. In this study, these aspects were evaluated from the perspective of music theme analysis. We performed a Chi-squared test to corroborate whether there are significant differences in the levels of EEG-ISC between the music section clusters. All analyses were performed at an alpha level of $p < .05$, and the results for Fisher's exact test involving multiple comparisons were corrected using a false discovery rate (FDR-correlation, the corrected alpha level of .05).

3. Results

3.1. Time-varying neural response similarity driven by musical features

Overall similarity of neural responses across listeners. As shown in Figure 2A, RC1 was maximally weighted over the fronto-central region, which is consistent with previous studies on auditory stimuli (Dauer et al., 2021; Kaneshiro et al., 2020, 2021). A low-resolution tomography (LORETA) analysis on the first component suggested a possible source in the cingulate cortex (Dmochowski et al., 2012), which has been suggested to be involved in conflict monitoring (*e.g*., awareness of the violation of expectations) (Bravo et al., 2019), as well as the fundamental process of detecting changes and provides a stimulus for subsequent alterations in behavior (Pearson et al., 2011). The results revealed that the overall EEG-ISC showed a statistically significant response to the original, but not to the phase version. In addition, the original version showed a significantly higher overall EEG-ISC value than the phase version (Figure 2B), $t(40) = 24.18$, $p < .001$, Cohens' $d = 3.776$, 95% CI [2.893, 4.652]. These results suggest that non-musicians have an overall similarity in neural responses to natural music, but not to phase scrambled music.

Insert Figure 2, about here.

Time-varying neural response similarity correlated with behavioral ratings. Since a significant overall similarity of neural responses was observed only in the original but not in the phase version, we focused our analysis on the original version. To examine whether the similarity of neural responses to music varies over time, we computed the time-varying ISC of the time series of neural responses. As shown in the top panel of Figure 2C, the EEG-ISC of the stimulus varied with time. These observed fluctuations were supported by a permutation test that yielded a 77.39% (421/544) significant time windows (ps < .05), suggesting that the neural response similarity across non-musicians varies over time during the whole course of music listening.

To establish whether the time-varying similarity of neural responses is relevant to musical tension, we performed a correlation analysis between the time-varying EEG-ISC and the fitness ratings of musical events (the natural musical tension showed good reliability across subjects, shown in

Supplementary Results). There was a significant negative correlation, $r = -.300$, $p < .001$. As shown in Figure 2D, the lower fitness ratings (*i.e.*, higher musical tension) provided by the participants, the higher the similarity of neural responses they showed. This finding indicates that the similarity of neural responses reflects felt musical tension, with a higher neural response similarity indicating a higher felt musical tension. Additionally, we also found that the more similar the tension ratings between participants (*i.e.*, the lower standard deviations of ratings), the more similar their neural responses were ($r = -.284$, $p < .001$), which further supported the relationship between the similarity of neural responses and musical tension.

Time-varying neural response similarity driven by musical features. We conducted a series of correlation analyses between the time-varying EEG-ISC and the five musical features, namely RMS, fluctuation peaks, key clarity, HCDF, and novelty. As shown in Figure 3A, a significant positive correlation was found between EEG-ISC and RMS ($r = .097$, $p < .05$), but significant negative correlations were found between EEG-ISC and fluctuation peaks, key clarity, and HCDF (*rs.* $< -.125$, *ps.* $< .004$). There was, however, no significant correlation between EEG-ISC and novelty ($p > .05$). These findings indicate that the time-varying neural response similarity is driven by RMS, fluctuation peaks, key clarity, and HCDF, all of which contribute to the experience of musical tension.

We also determined the correlations between the behavioral fitness ratings and the five musical features to confirm the contribution of these properties to musical tension. As shown in Figure 3B, the continuous fitness rating was positively correlated with RMS ($r = .506$, $p < .001$), fluctuation peaks ($r = .503$, $p < .001$), key clarity ($r = .335$, $p < .001$), HCDF ($r = .322$, $p < .001$), and novelty ($r = .144$, $p < .001$), confirming that these properties contribute to musical tension.

Insert Figure 3, about here.

3.2. Effects of musical themes on the patterns of time-varying neural response similarity

Classified levels of the EEG-ISC. We used *k*-means clustering (Quick cluster) to group EEG-ISCs from all time windows into a reduced number of levels (*i.e*., clusters of k-means clustering) based on their EEG-ISC value and 95th percentile threshold. Using the Elbow method, the optimal *k* -value was determined to be $k = 4$, thus EEG-ISC windows were classified into four levels, and the explained variance was about .95 (Figure S2). Among the 19 sections, there were 542 windows, each of which had a 5-s width with 1-s shifting time windows. The first two time-windows were excluded from the analysis, as they could not be assigned to any of the four levels, due to their extreme values. The four levels were sorted in descending order based on their values, with Level 1 having the highest ISC value, followed by Level 2, Level 3, and Level 4 with the lowest ISC value. In contrast, the frequency and proportion of time windows showed an increasing trend, with Levels 1, 2, 4 and 3 comprising 5.90% (32/542), 17.16% (93/542), 35.79% (194/542), and 41.14% (223/542) of the windows, respectively, as shown in Figure 4A.

Insert Figure 4, about here.

Distribution of EEG-ISC levels across the 19 sections. Each of the 19 sections showed different proportions of time windows at the four levels (Table S1). A Fisher's exact test indicated significant differences in the proportion of time windows at each level among these sections (Fisher's exact test: *p*-value < .001). As shown in Figure 4B, *post hoc* pairwise comparisons revealed significant differences in ISC at different levels between sections, as indicated by gray squares (FDR-corrected *p*-values < .05). However, the majority of similar thematic sections produced similar levels of ISC, as indicated by yellow squares (FDR-corrected *p*-values > .05, Table S2 for details). For example, all secondary themes in the exposition and recapitulation (*i.e*., S4, S5, S9, S10, S16, S17) produced comparable levels of ISC, while the primary theme in the exposition (*i.e*., S2, S7) showed similar ISC levels, except for the same theme in the recapitulation (*i.e.*, S15). In addition, transitions (*i.e.*, S3, S8) or closing zones (*i.e.*, S6, S11, S18) in the exposition and recapitulation (if any) showed similar levels of EEG-ISC, respectively. These results indicate that musical thematic materials affect the inter-subject similarity of neural responses.

EEG-ISC-based clustering of the 19 sections. We performed a hierarchical cluster analysis to establish whether the EEG-ISC produced by similar thematic materials can be grouped into the same cluster, in terms of their distribution patterns of the levels of EEG-ISC. Specifically, the frequencies of occurrence at each level were converted into proportions, and then a hierarchical cluster analysis was performed based on the proportion of EEG-ISC levels. The dendrogram and heatmap in Figure 5, depicting the results of the hierarchical clustering analysis, reveal that all sections within each cluster had similar proportions across the four levels, and a four-cluster solution appeared to be the optimal number of clusters. Secondary themes in exposition and recapitulation were grouped into Cluster 1. The primary themes in exposition I and II and development material that differed from primary thematic materials, except for the primary theme in recapitulation, were grouped into Cluster 2. Despite producing a similar level of ISC as indicated by the Fisher's exact test results, the closing zone in exposition I was not grouped with those in exposition II and recapitulation. Therefore, Cluster 3 consisted of the closing zone in exposition II and recapitulation, introduction in development, and coda, while Cluster 4 included introduction, closing zone in exposition I, and transition in exposition I and II. These results indicate that similar thematic materials produce similar levels of EEG-ISC.

Insert Figure 5, about here.

Additionally, the distribution of proportions at the four levels for the four music section clusters is also shown in Figure 5. Cluster 1 had the highest proportion at Level 4, as indicated by the orange-red color, while Cluster 2 had the highest proportion at Level 3. Cluster 3 had a relatively higher proportion at Level 3, while Cluster 4 had a higher proportion at Levels 2 and 3 (Table S3 for details). These frequency differences among the four clusters were subjected to a 4 (ISC levels) × 4 (clusters) Pearson's Chi-squared test, which revealed significant differences between the clusters ($\chi^2(9) = 244.85$, $p < .001$). The subsequently performed Fisher's exact *post hoc* pairwise comparisons (Table 2) revealed that any two clusters were significantly different at least at two EEG-ISC levels (as indicated by the yellow and blue squares in Figure 5). Specifically, for the significant pairwise comparisons, Cluster 1 consistently had higher proportions than the other three clusters at Level 4. Cluster 2 had higher proportions than Cluster 3 or 4 at Level 3, while Cluster 4 had higher proportions than Cluster 3 at Levels 1, 2, and 3. These results suggest that despite the thematic materials within each cluster exhibiting similar levels of EEG-ISC, the thematic materials in different clusters produce different levels of EEG-ISC.

Insert Table 2, about here.

----

4. Discussion

This study demonstrated for the first time that the time-varying similarities of neural responses associated with the experience of musical tension were shaped by both low-level physical features and high-level thematic structures. As predicted, we found that the time-varying similarity of neural responses was associated with the musical tension felt by non-musicians, with greater similarity observed at musical events with higher tension. This neural response similarity was driven by musical features that contribute to tension, as indicated by a positive correlation with RMS and negative correlations with fluctuation peaks, key clarity, and HCDF. The patterns of this time-varying similarity of neural responses were modulated by musical themes, with similar patterns observed for similar thematic materials. These findings have important implications for similarity in understanding.

Consistent with previous studies (Dauer et al., 2021; Kaneshiro et al., 2021), our study revealed a time-varying similarity in response to music. More importantly, our first main finding is that this time-varying similarity was associated with musical tension. Specifically, we found that higher levels of neural response similarity were associated with greater musical tension. This correlation can be explained by attentional engagement. Indeed, EEG-ISC has been suggested as an index of attentional engagement (Dmochowski et al., 2018; Dmochowski et al., 2012; Ki et al., 2016), and unexpected events require high attentional resources (Howard & Holcombe, 2010) and trigger reanalysis (Van de Meerendonk et al., 2010). Thus, the observed time-varying similarity of neural responses suggests that multiple listeners simultaneously display attentional engagement with unexpected musical events, but not with expected events, during the evaluation of musical tension.

Our second main finding is that the time-varying neural response similarity among listeners was driven by musical features that contribute to musical tension. Significant correlations found between behavioral ratings and the five musical features confirmed the contributions of these predictors of musical tension, as reported in a previous study (Lartillot, 2019). This is also consistent with previous studies (Farbood, 2012; Granot & Eitan, 2011; Hjortkjær, 2011) suggesting that musical tension is related to low-level aspects of auditory perception. In our study, the time-varying neural response similarity was positively correlated with RMS and negatively correlated fluctuation peaks, key clarity, and HCDF, all of which contribute to musical tension. These results may be explained by the relationship between neural response similarity, musical tension, and engagement of attention. Explicitly, the more tense participants felt, the higher engagement of attention was required, resulting in a greater level of similarity in brain response (Ki et al., 2016). In contrast to behavioral rating, the neural response similarity was positively correlated with RMS, which suggests that the louder the musical events, the greater the level of similarity in brain responses. Similarly, the neural response similarity, but not behavioral responses, was uncorrelated with novelty. The discrepancy between behavioral and neural responses may reflect different cognitive operations (Sun et al., 2020). The behavioral responses might react to consciously assessing musical event fitness, while neural responses could be tuned to instantly detect violations or fulfillment of musical events. In this case, the novelty of musical structures may be difficult to instantly detect by the brain. This assumption, however, requires further validation in future studies.

Our third main finding is that the patterns of time-varying neural response similarity were affected by musical thematic materials, indicating that the time-varying neural response similarity arose from the perception of musical structure. In particular, similar musical thematic materials, such as all the secondary themes across the exposition and recapitulation sections, had similar patterns of proportion distribution at the four levels, while different thematic materials between clusters, such as the introduction and the primary and secondary themes, had different patterns of proportion distribution at the four levels. This finding is consistent with previous research indicating that when musical materials are similar, group-average performances tend to be similar, and *vice versa* (McAdams et al., 2004). Since a piece of music is constructed by themes through repetitions, variations, and contrast, our study demonstrates the ability of listeners to recognize musical themes. Given that themes convey the essence of music, our findings suggest that time-varying neural response similarity may arise from extracting meaning from musical stimuli.

Our results also revealed some exceptions in which different thematic materials had a similar pattern of time-varying similarity, either in adjacent or nonadjacent structural sections. For instance, the recapitulation-primary theme, an adjacent structural section to recapitulation-secondary theme 1 and 2, achieved similar levels of neural response to the latter two themes. This result may be explained by the repetition of the primary theme. Indeed, the overall neural response similarity decreases over repeated exposures to familiar natural music (Madsen et al., 2019). One potential explanation is that familiar music often evokes personal and contextual associations (Thompson et al., 2023). In our study, since the primary theme was repeated several times, listeners might become more familiar with it, leading to a decrease in the inter-subject similarity of neural responses to the subsequently appearing recapitulation-primary theme. Alternatively, musical context may also explain some exceptions in the present study, as musical prediction is influenced by musical context (Quiroga-Martinez DR et al., 2019). At the beginning of the sonata, for example, the thematic materials in the introduction may be more difficult to predict than those in both the development-introduction and coda sections, even though these thematic materials are similar. As a result, the introduction section may produce a higher level of neural response similarity than those in the latter sections. Furthermore, some nonadjacent sections with different thematic materials produce similar levels of neural responses, such as thematic sections in Cluster 4. These similarities may be attributed to the combined influence of multiple factors, such as melody, harmony (Lehne et al., 2013; Lerdahl & Krumhansl, 2007), and rhythm and metre (Farbood, 2012; Granot & Eitan, 2011), rather than a simple sum of the effects of these factors. However, this hypothesis needs to be tested in future research.

It is worth noting that the EEG-ISC we used allowed us to measure fluctuations in the level of neural response similarity among participants throughout the entire evaluation of musical tension. Although musical understanding involves much more than just tension, musical tension is a critical component of musical emotion and intra-meaning (Koelsch, 2012). From this perspective, the implications of our findings may extend to musical understanding or intersubjectivity, as similarity in understanding is indicative of intersubjectivity (Margulis et al., 2022; Reynaert, 2001). First, our finding that, without any interaction, shared time-varying neural responses associated with musical tension were observed is particularly significant for musical understanding. This is because music lacks the semantic content that natural languages possess, making the acquisition of shared musical understanding through intersubjective interactions more challenging. Our findings, from the perspective of non-musicians, also provide evidence for the theory of musical intersubjectivity proposed by (Schütz, 1951), which posits that musical intersubjectivity is constituted by a shared flux of musical experience in inner time between composers, performers, and listeners. Second, intersubjectivity has long been assumed to depend on social interactions in the fields of philosophy (Husserl, 1977) and social cognition (Schore, 2021; Stone et al., 2012; Trevarthen, 2010). Our study showed that, without any interaction, a time-varying similarity of neural responses associated with musical tension could emerge across listeners. Although a shared external stimulus was found to produce an overall similarity of neural responses during understanding (Nguyen et al., 2019; Yeshurun et al., 2017), our research demonstrated that this similarity of neural responses is dynamic and can be shaped by both physical features and thematic structures. These findings highlight how the attributes of external stimuli contribute to the emergence of intersubjectivity and reveal a previously uncharted underlying neural mechanism of this process.

To conclude, our findings reveal that a time-varying neural response similarity, associated with musical tension, is shaped by the physical features and musical themes. Although musical understandings may vary and be personal, our findings contribute to the current understanding of how dynamic attributes of music can influence shifts in neural response similarity across the entire evaluation of musical tension, even in the absence of social interaction. Given that similarity in understanding has been traditionally believed to rely on social interaction, our findings pave the way for further research on the interplay of similarities in understandings produced by stimulus and social interactions.

Data Availability

All quantitative data have been deposited in OSF (https://osf.io/wxgkm/).

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Grant No. 31470972 to C. J. and F. L., and the European Research Council (ERC) Starting Grant to F. L. and C. J. (ERC-StG-2015, CAASD, 678733), and the Top Talent Cultivation Project for Doctoral Students of Shanghai Normal University to C. X. (209-AC9103-20-368005048). We would like to thank Prof. Guisheng Wang for the help with musical analysis, and thank Meng Du, Danni Wang, Wanqi Wang, and Chen Feng for their help with data acquisition and technical support.

References

Abrams, D. A., Ryali, S., Chen, T., Chordia, P., Khouzam, A., Levitin, D. J., & Menon, V. (2013, May). Inter-subject synchronization of brain responses during natural music listening. *Eur J Neurosci, 37*(9), 1458-1469. https://doi.org/10.1111/ejn.12173

Beethoven, L. v. (1981). Beethoven: Piano Sonata No.8 In C Minor, Op.13 -"Pathétique" - 1. Grave - Allegro di molto e con brio. On *Beethoven: Piano Sonatas*. Deutsche Grammophon GmbH. (published 1799)

Bergman, L. R. (1998). A pattern-oriented approach to studying individual development. In R. B. Cairns, L. R. Bergman, & J. Kagan (Eds.), *Methods and Models for Studying the Individual* (pp. 83–121.). SAGE: Thousand Oaks, CA.

Bigand, E., & Parncutt, R. (1999). Perceiving musical tension in long chord sequences. *Psychol Res, 62*(4), 237-254. https://doi.org/10.1007/s004260050053

Bravo, F., Cross, I., Hopkins, C., Gonzalez, N., Docampo, J., Bruno, C., & Stamatakis, E. A. (2019). Anterior cingulate and medial prefrontal cortex response to systematically controlled tonal dissonance during passive music listening. *Human Brain Mapping, 41*(1), 46-66. https://doi.org/10.1002/hbm.24786

Buber, M. (2012). *I and Thou*. eBookIt.

Chang, W.-T., Jääskeläinen, I. P., Belliveau, J. W., Huang, S., Hung, A.-Y., Rossi, S., & Ahveninen, J. (2015). Combined MEG and EEG show reliable patterns of electromagnetic brain activity during natural viewing. *NeuroImage, 114*, 49-56. https://doi.org/10.1016/j.neuroimage.2015.03.066

Cohen, S. S., & Parra, L. C. (2016). Memorable audiovisual narratives synchronize sensory and supramodal neural responses. *eNeuro*, 0203-0216. https://doi.org/10.1523/ENEURO.0203-16.2016

Dauer, T., Nguyen, D., Gang, N., Dmochowski, J., Berger. J, & Kaneshiro, B. (2021). Inter-subject Correlation While Listening to Minimalist Music: A Study of Electrophysiological and Behavioral Responses to Steve Reich's Piano Phase. *Front. Neurosci., 15:702067*. https://doi.org/10.3389/fnins.2021.702067

Degani, A., Dalai, M., Leonardi, R., & Migliorati, P. (2015). Harmonic change detection for musical chords segmentation. 2015 IEEE International Conference on Multimedia and Expo (ICME),

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*, 9-21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Dmochowski, J. P., Greaves, A. S., & Norcia, A. M. (2015). Maximally reliable spatial filtering of steady state visual evoked potentials. *NeuroImage, 109*, 63-72. https://doi.org/10.1016/j.neuroimage.2014.12.078

Dmochowski, J. P., Ki, J. J., DeGuzman, P., Sajda, P., & Parra, L. C. (2018, Oct 15). Extracting multidimensional stimulus-response correlations using hybrid encoding-decoding of neural activity. *NeuroImage, 180*(Pt A), 134-146. https://doi.org/10.1016/j.neuroimage.2017.05.037

Dmochowski, J. P., Sajda, P., Dias, J., & Parra, L. C. (2012). Correlated components of ongoing EEG point to emotionally laden attention - a possible marker of engagement? *Front Hum Neurosci, 6*, 112. https://doi.org/10.3389/fnhum.2012.00112

Echterhoff, G., Higgins, E. T., & Levine, J. M. (2009). Shared reality: Experiencing commonality with others' inner states about the world. *Perspectives on Psychological Science, 4*(5), 496-521. https://doi.org/10.1111/j.1745-6924.2009.01161.x

Farbood, M. M. (2012). A parametric, temporal model of musical tension. *Music Perception, 29*(4), 387-428. https://doi.org/10.1525/mp.2012.29.4.387

Foote, J. T., & Cooper, M. L. (2003). *Media segmentation using self-similarity decomposition.* (M. M. Yeung, R. W. Lienhart, & C.-S. li, Eds.). https://doi.org/10.1117/12.476302
Garfinkel, H. (1967). *Studies in Ethnomethodology.* Prentice-Hall.
Girard, J. M. (2014). CARMA: Software for continuous affect rating and media annotation. *Journal of Open Research Software, 2*(1), e5. https://doi.org/10.5334/jors.ar
Gómez, E. (2006). Tonal description of music audio signals (PhD thesis). *Universitat Pompeu Fabra, Barcelona, Spain*.
Göncü, A. (1993). Development of Intersubjectivity in Social Pretend Play. *Human Development, 36*(4), 185-198. https://doi.org/10.1159/000278206
Granot, R. Y., & Eitan, Z. (2011). Musical tension and the interaction of dynamic auditory parameters. *Music Perception, 28*(3), 219-246. https://doi.org/10.1525/mp.2011.28.3.219
Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1995). *Hair, J. F., Anderson, R. E., Tatham, R. L., Tatham, R. L. & Black, W. C. (1995). Multivariate Data Analysis with Readings.* . PrenticeHall: Upper Saddle River, NJ.
Hallam, S., & Papageorgi, I. (2016). Conceptions of musical understanding. *Research Studies in Music Education, 38*(2), 133-154. https://doi.org/10.1177/1321103X16671037
Harte, C., Mark Sandler, & Gasser, M. (2006). *Detecting harmonic change in musical audio.* Proceedings of the 1st ACM workshop on Audio and music computing multimedia (AMCMM '06), New York, NY, USA.
Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004, Mar 12). Intersubject synchronization of cortical activity during natural vision. *Science, 303*(5664), 1634-1640. https://doi.org/10.1126/science.1089506
Heidegger, M. (2010). *Being and time.* Suny Press.
Higgins, E. T., Echterhoff, G., Crespillo, R., & Kopietz, R. (2007). Effects of communication on social knowledge: Sharing reality with individual versus group audiences. *Japanese Psychological Research, 49*(2), 89-99. https://doi.org/10.1111/j.1468-5884.2007.00336.x
Higgins, E. T., Rossignac-Milon, M., & Echterhoff, G. (2021). Shared reality: From sharing-is-believing to merging minds. *Current Directions in Psychological Science, 30*(2), 103–110. https://doi.org/10.1177/0963721421992027
Hjortkjær, J. (2011). *Toward a cognitive theory of musical tension* [PhD thesis, Copenhagen University]. Copenhagen, Denmark.
Howard, C. J., & Holcombe, A. O. (2010). Unexpected changes in direction of motion attract attention. *Attention, Perception, & Psychophysics, 72*, 2087–2095 https://doi.org/10.3758/BF03196685
Huron, D. (2006). *Sweet anticipation: music and the psychology of expectation*. Massachusetts Institute of Technology.
Husserl, E. (1977). *Cartesian meditations: An introduction to phenomenology.* (D. Cairns, Trans.). Martinus Nijhoff. (Original work published 1931)
Jung, T. P., Humphries, C., Lee, T. W., Makeig, S., McKeown, M., Iragui, V., & Sejnowski, T. J. (1997). Extended ICA removes artifacts from electroencephalographic recordings. *Advances in neural information processing systems, 10*, 894-900.
Kaneshiro, B., Nguyen, D. T., Norcia, A. M., Dmochowski, J. P., & Berger, J. (2020, Jul 1). Natural music evokes correlated EEG responses reflecting temporal structure and beat. *NeuroImage, 214*, 116559. https://doi.org/10.1016/j.neuroimage.2020.116559
Kaneshiro, B., Nguyen, D. T., Norcia, A. M., Dmochowski, J. P., & Berger, J. (2021). Inter-subject EEG correlation reflects time-varying engagement with natural music. *bioRxiv*. https://doi.org/10.1101/2021.04.14.439913
Ki, J. J., Kelly, S. P., & Parra, L. C. (2016, Mar 9). Attention Strongly Modulates Reliability of Neural Responses to Naturalistic Narrative Stimuli. *J Neurosci, 36*(10), 3092-3101.

https://doi.org/10.1523/JNEUROSCI.2942-15.2016
Koelsch, S. (2012). *Brain and music*. Wiley Blackwell.
Kopiez, R. (2006). Making music and making sense through music: expressive performance and communication. In R. Colwell (Ed.), *MENC Handbook of Musical Cognition and Development* (pp. 189–224). NY: Oxford University Press.
Krumhansl, C. L. (2001). *Cognitive foundations of musical pitch* (Vol. 17). Oxford University Press.
Krumhansl, C. L. (2002). Music: A link between cognition and emotion. *Current Directions in Psychological Science, 11*(2), 45-50. https://doi.org/10.1111/1467-8721.00165
Lahnakoski, J. M., Glerean, E., Jääskeläinen, I. P., Hyönä, J., Hari, R., Sams, M., & Nummenmaa, L. (2014). Synchronous brain activity across individuals underlies shared psychological perspectives. *NeuroImage, 100*, 316-324. https://doi.org/10.1016/j.neuroimage.2014.06.022
Lartillot, O. (2019). *MIRtoolbox 1.7.2 user's manual*. Aalborg University.
Lartillot, O., & Toiviainen, P. (2007). *MIR in Matlab (II): a toolbox for musical feature extraction from audio.* Proc. Intl. Conf. Music Inform.,
Lehne, M., & Koelsch, S. (2015, 2015-February-11). Toward a general psychological model of tension and suspense [Hypothesis and Theory]. *6*. https://doi.org/10.3389/fpsyg.2015.00079
Lehne, M., Rohrmeier, M., Gollmann, D., & Koelsch, S. (2013). The influence of different structural features on felt musical tension; in two piano pieces by mozart and mendelssohn. *Music Perception, 31*(2), 171-185. https://doi.org/10.1525/mp.2013.31.2.171
Lerdahl, F., & Krumhansl, C. L. (2007). Modeling tonal tension. *Music Perception, 24*(4), 329-366. https://doi.org/10.1525/mp.2007.24.4.329
Madsen, J., Margulis, E. H., Simchy-Gross, R., & Parra, L. C. (2019, Mar 5). Music synchronizes brainwaves across listeners with strong effects of repetition, familiarity and training. *Sci Rep, 9*(1), 3576. https://doi.org/10.1038/s41598-019-40254-w
Marchetti, I., & Koster, E. (2014, 2014-January-24). Brain and intersubjectivity: a Hegelian hypothesis on the self-other neurodynamics [Opinion]. *Frontiers in Human Neuroscience, 8*. https://doi.org/10.3389/fnhum.2014.00011
Margulis, E. H., Wong, P. C., Turnbull, C., Kubit, B. M., & McAuley, J. D. (2022). Narratives imagined in response to instrumental music reveal culture-bounded intersubjectivity. *Proceedings of the National Academy of Sciences, 119*(4). https://doi.org/10.1073/pnas.2110406119
Martin, G. M. (1966). John Malcom Tait The significance of musical understanding in music education. *J Current Musicology*(4), 193.
Martínez-Molina, N., Mas-Herrero, E., Rodríguez-Fornells, A., Zatorre, R. J., & Marco-Pallarés, J. (2016). Neural correlates of specific musical anhedonia. *113*(46), E7337-E7345. https://doi.org/doi:10.1073/pnas.1611211113
Mas-Herrero, E., Marco-Pallares, J., Lorenzo-Seva, U., Zatorre, R. J., & Rodriguez-Fornells, A. (2013). Individual Differences in Music Reward Experiences. *Music Perception, 31*(2), 118-138. https://doi.org/10.1525/mp.2013.31.2.118 %J Music Perception
McAdams, S., Vieillard, S., Houix, O., & Reynolds, R. (2004). Perception of Musical Similarity Among Contemporary Thematic Materials in Two Instrumentations. *Music Perception, 22*(2), 207–237. https://doi.org/10.1525/mp.2004.22.2.207
McAuley, J. D., Wong, P. C., Mamidipaka, A., Phillips, N., & Margulis, E. H. (2021). Do you hear what I hear? Perceived narrative constitutes a semantic dimension for music. *Cognition, 212*, 104712. https://doi.org/10.1016/j.cognition.2021.104712
Meyer, L. B. (1956). *Emotion and meaning in music*. University of Chicago.
Moore, B. C. (2012). *An introduction to the psychology of hearing*. Brill.
Nguyen, M., Vanderwal, T., & Hasson, U. (2019). Shared understanding of narratives is correlated with shared neural responses. *NeuroImage, 184*, 161-170.

https://doi.org/10.1016/j.neuroimage.2018.09.010

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia, 9*(1), 97-113. https://doi.org/10.1016/0028-3932(71)90067-4

Pampalk, E., Rauber, A., & Merkl, D. (2002). *Content-based organization and visualization of music archives.* Proceedings of the tenth ACM international conference on Multimedia (MULTIMEDIA '02). New York, NY, USA.

Parra, L. C., Spence, C. D., Gerson, A. D., & Sajda, P. (2005). Recipes for the linear analysis of EEG. *NeuroImage, 28*(2), 326–341. https://doi.org/10.1016/j.neuroimage.2005.05.032

Pearson, J. M., Heilbronner, S. R., Barack, D. L., Hayden, B. Y., & Platt, M. L. (2011). Posterior cingulate cortex: adapting behavior to a changing world. *Trends in Cognitive Sciences, 15*(4), 143–151. https://doi.org/10.1016/j.tics.2011.02.002

Prichard, D., & Theiler, J. (1994). Generating surrogate data for time series with several simultaneously measured variables. *Physical review letters, 73*(7), 951-954. https://doi.org/10.1103/PhysRevLett.73.951.

Quiroga-Martinez DR, Hansen NC, Højlund A, Pearce MT, Brattico E, & P., V. (2019). Reduced prediction error responses in high-as compared to low-uncertainty musical contexts. *Cortex, 120*, 181-200. https://doi.org/10.1016/j.cortex.2019.06.010.

Raymond, C. W. (2019). Intersubjectivity, Normativity, and Grammar. *Social Psychology Quarterly, 82*(2), 182–204. https://doi.org/10.1177/0190272519850781

Reynaert, P. (2001). Intersubjectivity and Naturalism--Husserl's Fifth Cartesian Meditation Revisited. *Husserl Studies, 17*(3), 207. https://doi.org/10.1023/A:1010744200494

Saari, P., Eerola, T., Fazekas, G., & Sandler, M. (2013). Using semantic layer projection for enhancing music mood prediction with audio features. Sound and Music Computing Conference, Stockholm, Sweden,

Sachs, M. E., Habibi, A., Damasio, A., & Kaplan, J. T. (2020). Dynamic intersubject neural synchronization reflects affective responses to sad music. *NeuroImage, 218*, 116512. https://doi.org/10.1016/j.neuroimage.2019.116512

Scheff, T. J., Phillips, B. S., & Kincaid, H. (2006). *Goffman Unbound!: A New Paradigm for Social Science (1st ed.)*. Routledge. https://doi.org/10.4324/9781315634357

Schore, A. N. (2021). The Interpersonal Neurobiology of Intersubjectivity. *Front. Psychol., 12*, 648616. https://doi.org/10.3389/fpsyg.2021.648616

Schutz, A. (1972). *The phenomenology of the social world*. Northwestern university press.

Schütz, A. (1951). Making music together: A study in social relationship. *Social research, 18*, 76-97.

Stone, L. D., Underwood, C., & Hotchkiss, J. (2012). The Relational Habitus Intersubjective Processes in Learning Settings. *Human Development, 55*(2), 65-91. https://www.jstor.org/stable/26764607

Sun, L., Thompson, W. F., Liu, F., Zhou, L., & Jiang, C. (2020). The human brain processes hierarchical structures of meter and harmony differently: Evidence from musicians and nonmusicians. *Psychophysiology, 57*(9), e13598. https://doi.org/10.1111/psyp.13598

Terrace, H. S., Bigelow, A. E., & Beebe, B. (2022). Intersubjectivity and the Emergence of Words. *Frontiers in Psychology, 13*, 693139. https://doi.org/10.3389/fpsyg.2022.693139

Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., & Farmer, J. D. (1992). Testing for nonlinearity in time series: the method of surrogate data. *Phys. Nonlinear Phenom., 58*(1), 77–94. https://doi.org/10.1016/0167-2789(92)90102-S.

Thompson, W. F., Bullot, N. J., & Margulis, E. H. (2023). The psychological basis of music appreciation: Structure, self, source. *Psychological Review, 130*(1), 260-284. https://doi.org/10.1037/rev0000364

Trevarthen, C. (2010). What is it like to be a person who knows nothing? Defining the active intersubjective mind of a newborn human being. *Infant and Child Development, 20*(1), 119–135.

https://doi.org/10.1002/icd.689
Trost, W., Frühholz, S., Cochrane, T., Cojan, Y., & Vuilleumier, P. (2015). Temporal dynamics of musical emotions examined through intersubject synchrony of brain activity. *Social Cognitive and Affective Neuroscience, 10*(12), 1705-1721. https://doi.org/10.1093/scan/nsv060
Van de Meerendonk, N., Kolk, H. H. J., Vissers, C. T. W. M., & Chwilla, D. J. (2010). Monitoring in Language Perception: Mild and Strong Conflicts Elicit Different ERP Patterns. *Journal of Cognitive Neuroscience, 22*(1), 67–82. https://doi.org/10.1162/jocn.2008.21170
Vogeley, K. (2017, Aug 19). Two social brains: neural mechanisms of intersubjectivity. *Philos Trans R Soc Lond B Biol Sci, 372*(1727). https://doi.org/10.1098/rstb.2016.0245
Yeshurun, Y., Swanson, S., Simony, E., Chen, J., Lazaridi, C., Honey, C. J., & Hasson, U. (2017). Same story, different story: the neural representation of interpretive frameworks. *Psychological Science, 28*(3), 307-319. https://doi.org/10.1177/0956797616682029
Zhang, Z., Murtagh, F., Van Poucke, S., Lin, S., & Lan, P. (2017). Hierarchical cluster analysis in clinical research with heterogeneous study population: highlighting its visualization with R. *Annals of translational medicine, 5*(4), 75. https://doi.org/10.21037/atm.2017.02.05

Table 1. The form of Ludwig van Beethoven's Piano Sonata No. 8 in C minor, Op. 13, first movement.

Formal sections	Thematic material	Section number	Measure number	Key
Introduction		S1	1-10	c→♭E→D→c:D_7
‖: Exposition :‖	Primary theme	S2 & S7	11-27	c
	Transition	S3 & S8	28-49	c→♭E:D7
	Secondary theme 1	S4 & S9	50-88	♭e→♭D→♭E
	Secondary theme 2	S5 & S10	89-112	♭E
	Closing zone	S6 & S11	113-132	♭E
Development	Introduction	S12	133-136	d? →g?→e:D7
	Development	S13	137-171	e→g→F?→c:D7
	Retransition	S14	172-194	Prolonged dominant
Recapitulation	Primary theme	S15	195-219	c→?
	Secondary theme 1	S16	220-252	f→c
	Secondary theme 2	S17	253-276	c
	Closing zone	S18	277-294	c
Coda		S19	295-310	c

Note: Question marks denote ambiguity or vagueness of tonality. ‖: :‖ is a repeat sign that indicates a section (*i.e.*, Exposition) should be repeated.

Table 2. Frequency and proportion (in brackets inside the low part of each cell, the unit is %) of music section clusters by EEG ISC levels for each pairwise comparison using Fisher's exact test.

	Level 1	Level 2	Level 1	Level 3	Level 1	Level 4	Level 2	Level 3	Level 2	Level 4	Level 3	Level 4
Cluster 1	0 (0)	1 (100)	0 (0)	62 (100)	0 (0)	129 (100)	1 (1.59)	62 (98.41)	1 (.77)	**129 (99.23)**	62 (32.46)	**129 (67.54)**
Cluster 2	0 (0)	2 (100)	0 (0)	42 (100)	0 (0)	4 (100)	2 (4.55)	42 (95.45)	**2 (33.33)**	4 (66.67)	**42 (91.30)**	4 (8.70)
Cluster 1	0 (0)	1 (100)	0 (0)	**62 (100)**	0 (0)	**129 (100)**	1 (1.59)	**62 (98.41)**	1 (.77)	**129 (99.23)**	62 (32.46)	**129 (67.54)**
Cluster 3	12 (32.43)	25 (67.57)	**12 (16.22)**	62 (83.78)	**12 (20)**	48 (80)	**25 (28.74)**	62 (71.26)	**25 (34.25)**	48 (65.75)	**62 (56.36)**	48 (43.64)
Cluster 1	0 (0)	1 (100)	0 (0)	**62 (100)**	0 (0)	**129 (100)**	1 (1.59)	**62 (98.41)**	1 (.77)	**129 (99.23)**	62 (32.46)	**129 (67.54)**
Cluster 4	20 (23.53)	65 (76.47)	**20 (25.97)**	57 (74.03)	**20 (60.61)**	13 (39.39)	**65 (53.28)**	57 (46.72)	**65 (83.33)**	13 (16.67)	**57 (81.43)**	13 (18.57)
Cluster 2	0 (0)	2 (100)	0 (0)	**42 (100)**	0 (0)	4 (100)	2 (4.55)	**42 (95.45)**	2 (33.33)	4 (66.67)	**42 (91.30)**	4 (8.70)
Cluster 3	12 (32.43)	25 (67.57)	**12 (16.22)**	62 (83.78)	12 (20)	48 (80)	**25 (28.74)**	62 (71.26)	25 (34.25)	48 (65.75)	62 (56.36)	**48 (43.64)**
Cluster 2	0 (0%)	2 (100)	0 (0)	**42 (100)**	0 (0)	4 (100)	2 (4.55)	**42 (95.45)**	2 (33.33)	**4 (66.67)**	42 (91.30)	4 (8.70)
Cluster 4	20 (23.53)	65 (76.47)	**20 (25.97)**	57 (74.03)	20 (60.61)	13 (39.39)	**65 (53.28)**	57 (46.72)	**65 (83.33)**	13 (16.67)	57 (81.43)	13 (18.57)
Cluster 3	12 (32.43)	25 (67.57)	12 (16.22)	62 (83.78)	12 (20)	**48 (80)**	25 (28.74)	**62 (71.26)**	25 (34.25)	**48 (65.75)**	62 (56.36)	**48 (43.64)**
Cluster 4	20 (23.53)	65 (76.47)	20 (25.97)	57 (74.03)	**20 (60.61)**	13 (39.39)	**65 (53.28)**	57 (46.72)	**65 (83.33)**	13 (16.67)	**57 (81.43)**	13 (18.57)

Note: Each 2 x 2 square lattice represents the paired comparison of any two levels and two clusters. Significance levels of paired comparisons are shaded in gray: FDR_adjusted *p*_value < .05 and white: FDR_adjusted *p*_value > .05. Bold text in the gray cell indicates that in this pairwise comparison, the ISC proportion of the current cluster is larger than that of the other cluster in this level.

Figure Captions

Figure 1. The waveforms of the original (A) and the phase (B) stimuli. Nineteen musical sections were generated from five formally-defined musical parts (*i.e*., introduction, exposition, development, recapitulation, and coda). The musical sections (S1-S19) are separated by vertical dashed lines, and the duration of each section is as follows: S1 (introduction: 00:00-1:47), S2 (exposition I-primary theme: 1:48-2:00), S3 (exposition I-transition: 2:01-2:15), S4 (exposition I-secondary theme 1: 2:16-2:49), S5 (exposition I-secondary theme 2: 2:50-3:08), S6 (exposition I-closing zone: 3:09-3:27), S7 (exposition II-primary theme: 3:28-3:39), S8 (exposition II-transition: 3:40-3:55), S9 (exposition II-secondary theme 1: 3:56-4:28), S10 (exposition II-secondary theme 2: 4:29-4:47), S11 (exposition II-closing zone: 4:48-5:10), S12 (development-introduction: 5:11-5:54), S13 (development-development: 5:55-6:17), S14 (development-retransition: 6:18-6:39), S15 (recapitulation-primary theme: 6:40-6:57), S16 (recapitulation-secondary theme 1: 6:58-7:25), S17 (recapitulation-secondary theme 2: 7:26-7:44), S18 (recapitulation-closing zone: 7:45-8:07), and S19 (coda: 8:08-9:09). It is worth noting that the exposition was played twice, and thus exposition I and exposition II were used to differentiate the two presentations of this part. For an enhanced visual representation, the amplitude coordinate scales for the two sound stimuli are not identical. Specifically, the original ranges from -0.15 to 0.15, while the phase ranges from -0.05 to 0.05 [No units provided in MIRtoolbox documentation].

Figure 2. EEG component and overall ISC for each stimulus, as well as time-varying ISC for the original music and its correlation with musical tension. (A) Spatial filtering component RC1 is depicted. (B) EEG ISC is computed across the entire duration of each stimulus. Bar height represents the mean value, and error bar height represents ± SEM. *** denotes significance at a level of p < .001. The shaded gray area indicates the 95th percentile of the null distribution. (C) The shaded gray area in the top panel indicates the 95th percentile of the corresponding null distribution, and the shaded area in the bottom panel shows one SEM across subjects. Dotted lines mark the end of music events. (D) The EEG-ISC is

temporally correlated with the continuous rating of musical tension for the original stimuli, the latter of which is smoothed to match the 5s window of EEG-ISC.

Figure 3. Correlations between ISC/continuous rating and the musical features. (A) Correlations between the time-varying EEG ISC and the five musical features, respectively. (B) Correlations between the continuous fitness rating and the five musical features, respectively. Each musical feature and the continuous rating were downsampled to match the 5s-window of EEG-ISC.

Figure 4. Four EEG-ISC levels and their distribution among the 19 sections. (A) K-means clustering was used to reduce the time-varying EEG-ISC points to a few levels for the purpose of comparing the EEG-ISC distribution of the 19 music sections. Here, a k-means cluster plot of the four EEG-ISC levels created from ISC and the 95th threshold. It uses k-means results and the original data as arguments, and observations are represented by points. The points at each level indicate the frequency of occurrence of time windows. (B) Triangular Heatmap showing pairwise comparisons of the musical sections by the Fisher's exact test. Yellow squares indicate that the two musical sections did not show significant differences at any two-level comparison, and gray squares indicate that the two musical sections were significantly different at the two-level comparison, at least at one pairwise comparison. In order to show whether each two-level pairwise comparison reaches a significant difference, we used the sectors of a sextile circle to represent the results of the six pairs, respectively. The gray sector indicates statistically nonsignificant differences, while the white sector indicates statistically significant differences.

Figure 5. Visualization of hierarchical clustering analysis. The heatmap shows the 19 musical sections projected onto 4 music section clusters identified using hierarchical clustering and displays the distribution of proportions at four levels for the four music section clusters. The rows represent musical sections, while the columns represent ISC levels. The shading of the tiles indicates the proportion of ISC levels: Red color indicates a higher proportion, while blue color indicates a smaller proportion. The

dendrogram was constructed using hierarchical clustering, and *cutree* was used to divide it into 4 music section clusters.

Figure 1

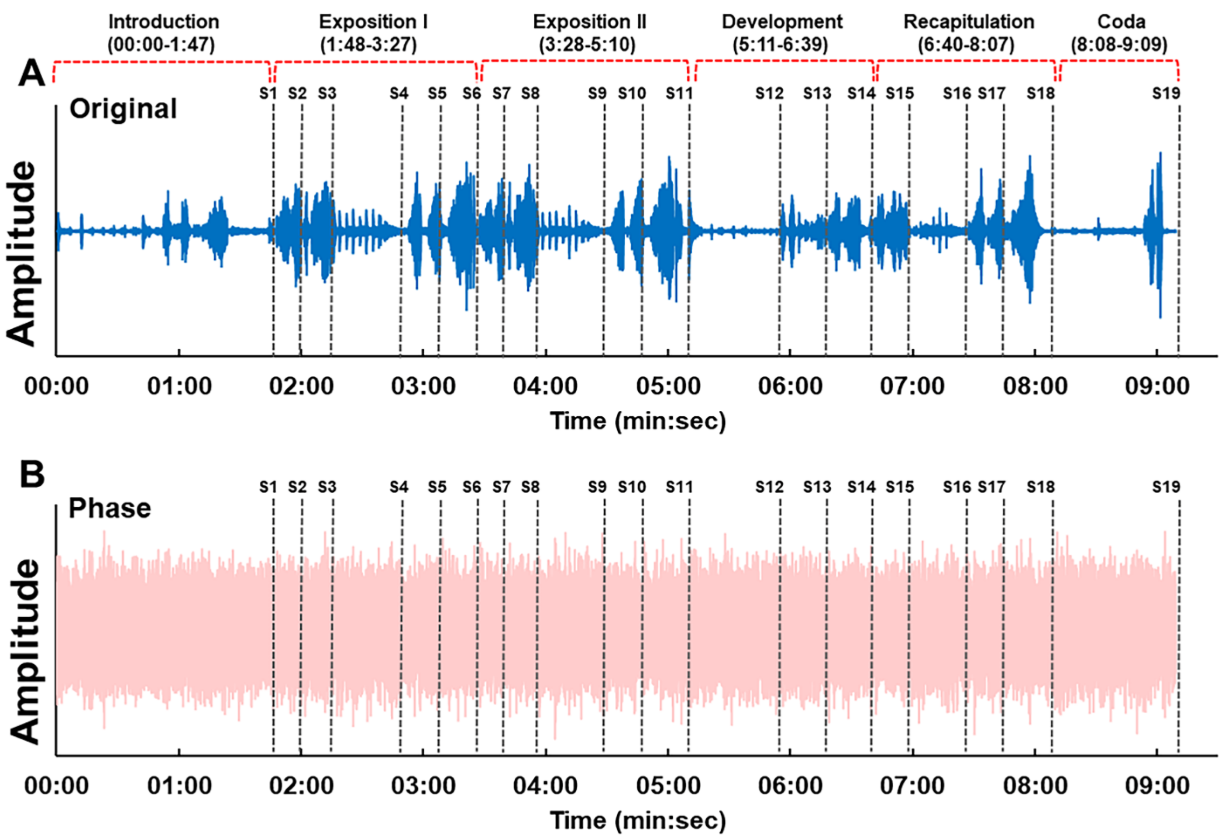

Introduction
(00:00-1:47)
Exposition I
(1:48-3:27)
Exposition II
(3:28-5:10)
Development
(5:11-6:39)
Recapitulation
(6:40-8:07)
Coda
(8:08-9:09)
A
Original
S1 S2 S3 S4 S5 S6 S7 S8 S9 S10 S11 S12 S13 S14 S15 S16 S17 S18 S19
Amplitude
00:00 01:00 02:00 03:00 04:00 05:00 06:00 07:00 08:00 09:00
Time (min:sec)
B
Phase
S1 S2 S3 S4 S5 S6 S7 S8 S9 S10 S11 S12 S13 S14 S15 S16 S17 S18 S19
Amplitude
00:00 01:00 02:00 03:00 04:00 05:00 06:00 07:00 08:00 09:00
Time (min:sec)

Figure 2

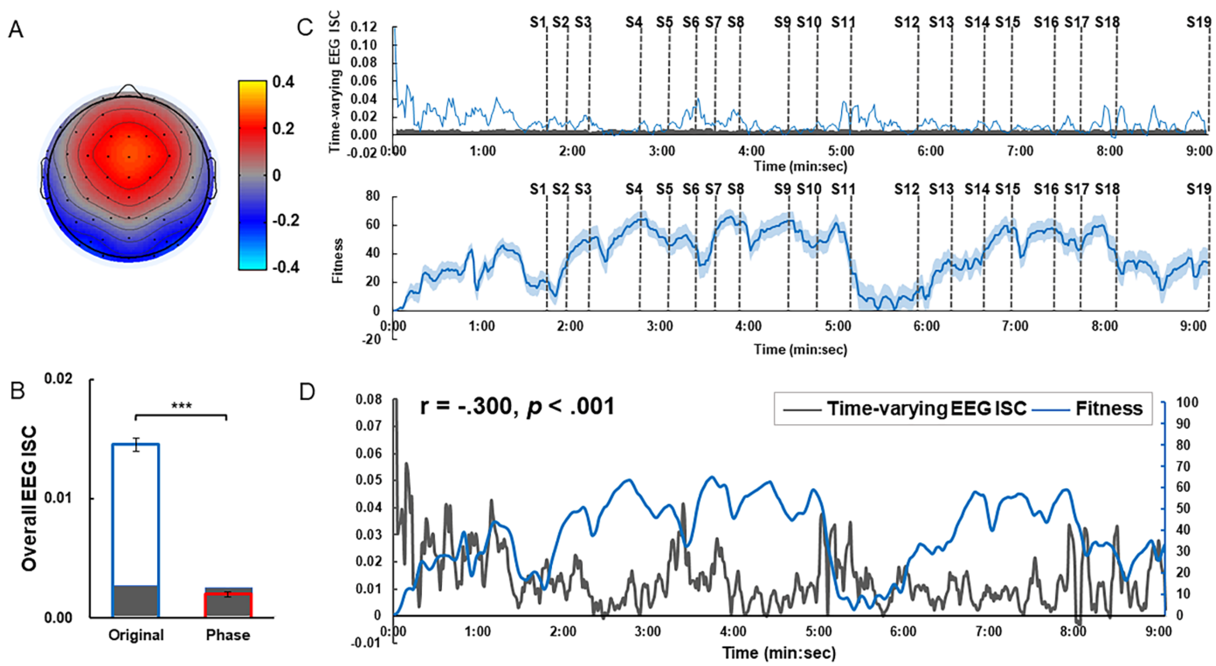
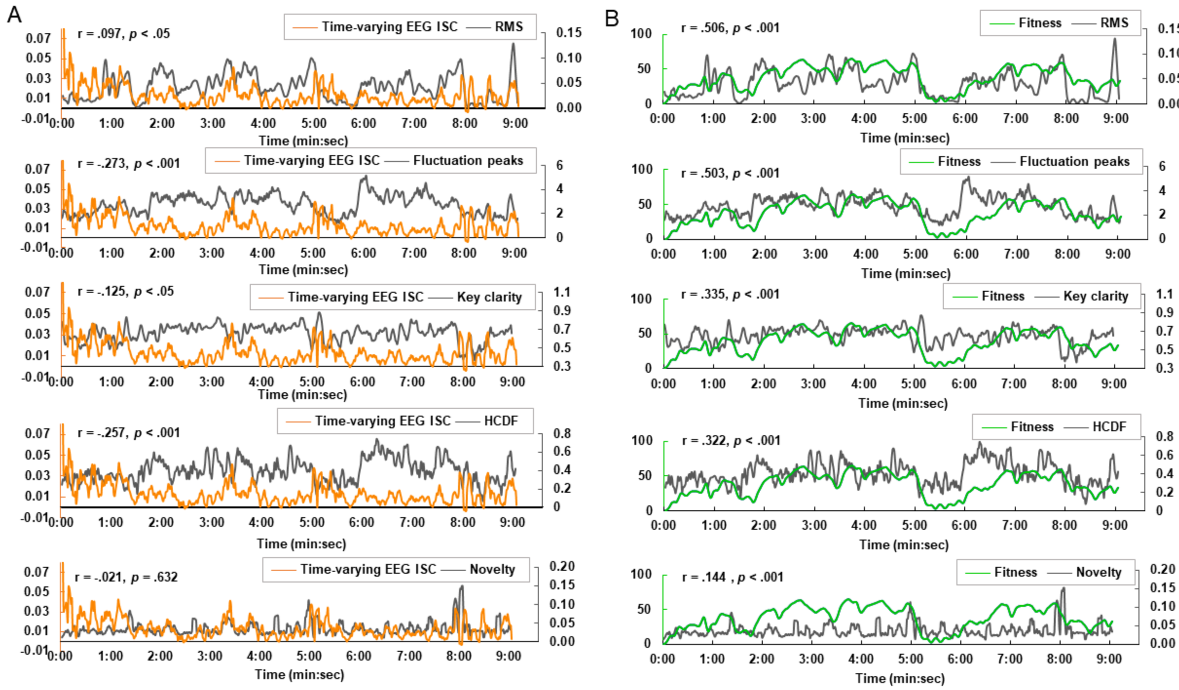
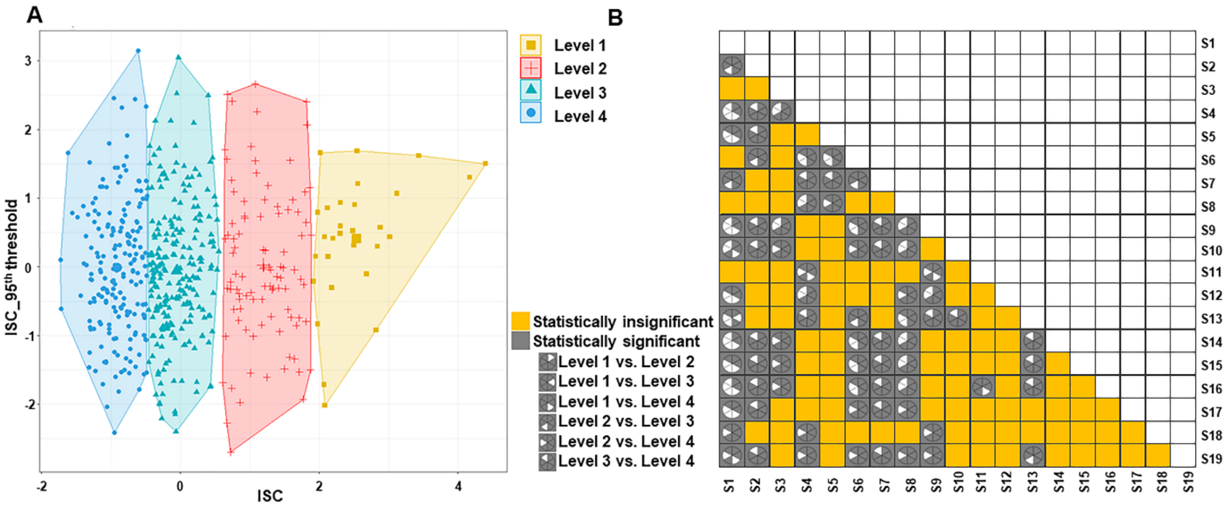

A
0.4
0.2
0
-0.2
-0.4
C
Time-varying EEG ISC
S1 S2 S3 S4 S5 S6 S7 S8 S9 S10 S11 S12 S13 S14 S15 S16 S17 S18 S19
Time (min:sec)
Fitness
B
Overall EEG ISC
***
Original
Phase
D
r = -.300, p < .001
Time-varying EEG ISC
Fitness
Time (min:sec)

Figure 3



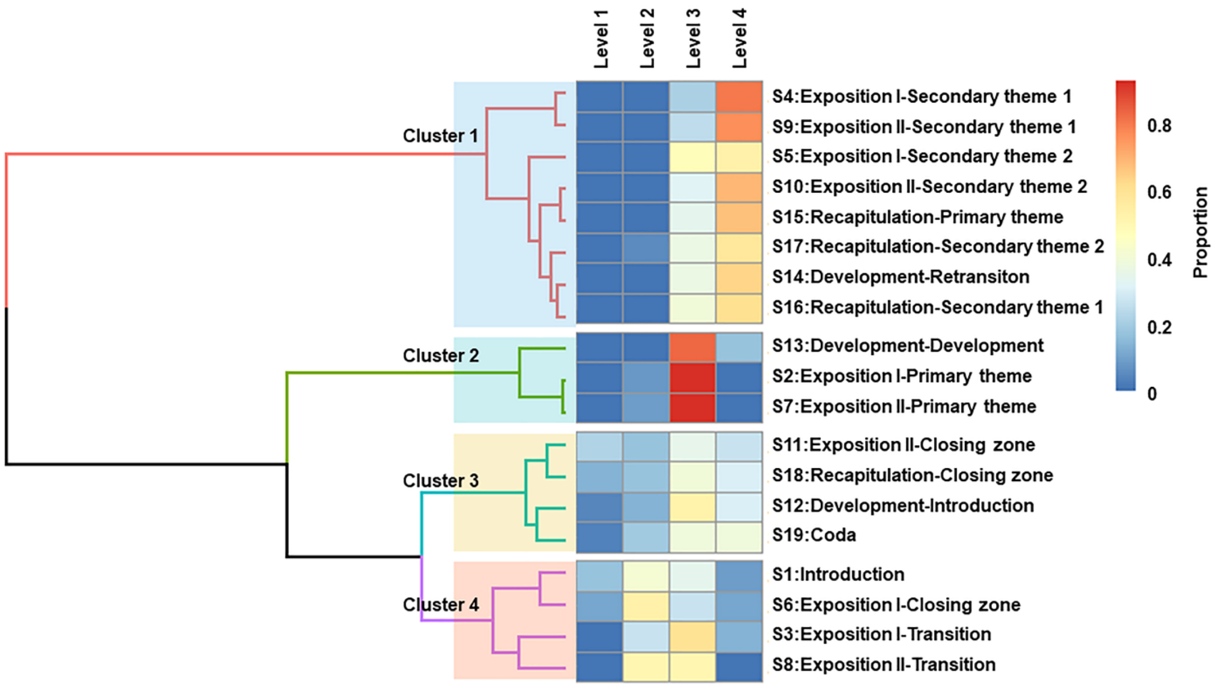
A
r = .097, p < .05
Time-varying EEG ISC
RMS
Time (min:sec)
r = -.273, p < .001
Time-varying EEG ISC
Fluctuation peaks
Time (min:sec)
r = -.125, p < .05
Time-varying EEG ISC
Key clarity
Time (min:sec)
r = -.257, p < .001
Time-varying EEG ISC
HCDF
Time (min:sec)
r = -.021, p = .632
Time-varying EEG ISC
Novelty
Time (min:sec)
B
r = .506, p < .001
Fitness
RMS
Time (min:sec)
r = .503, p < .001
Fitness
Fluctuation peaks
Time (min:sec)
r = .335, p < .001
Fitness
Key clarity
Time (min:sec)
r = .322, p < .001
Fitness
HCDF
Time (min:sec)
r = .144 , p < .001
Fitness
Novelty
Time (min:sec)

Figure 4



A
B
Level 1
Level 2
Level 3
Level 4
ISC_95th threshold
ISC
Statistically insignificant
Statistically significant
Level 1 vs. Level 2
Level 1 vs. Level 3
Level 1 vs. Level 4
Level 2 vs. Level 3
Level 2 vs. Level 4
Level 3 vs. Level 4
S1
S2
S3
S4
S5
S6
S7
S8
S9
S10
S11
S12
S13
S14
S15
S16
S17
S18
S19

Figure 5



Level 1
Level 2
Level 3
Level 4
Cluster 1
Cluster 2
Cluster 3
Cluster 4
S4:Exposition I-Secondary theme 1
S9:Exposition II-Secondary theme 1
S5:Exposition I-Secondary theme 2
S10:Exposition II-Secondary theme 2
S15:Recapitulation-Primary theme
S17:Recapitulation-Secondary theme 2
S14:Development-Retransiton
S16:Recapitulation-Secondary theme 1
S13:Development-Development
S2:Exposition I-Primary theme
S7:Exposition II-Primary theme
S11:Exposition II-Closing zone
S18:Recapitulation-Closing zone
S12:Development-Introduction
S19:Coda
S1:Introduction
S6:Exposition I-Closing zone
S3:Exposition I-Transition
S8:Exposition II-Transition
0.8
0.6
0.4
0.2
0
Proportion

Supplementary Results

Reliability of continuous fitness rating

To investigate the reliability of subjects' continuous fitness ratings, we used two-way mixed, absolute agreement, average-measures intra-class correlation coefficients (ICC) with 95% confidence intervals (CI) to estimate the inter-rater reliability among subjects (Hallgren, 2012). The ICC coefficients range between 0 and 1 and were interpreted as having poor (ICC $< .5$), moderate (.5–.75), good (.75–.9), and excellent ($> .9$) reliability (Koo & Li, 2016). The results showed that ICC values for Original music showed a good reliability, the average measure ICC was .816 with a 95% confidence interval from .791 to .836 (F(2195,87800, $p < .001$); while ICC values for Phase music showed a moderate reliability, the average measure ICC was .610 with a 95% confidence interval from .511 to .686 (F(2195,87800, $p < .001$). These results indicate that higher similarity in the tension assessment of the original music compared to the phase music.

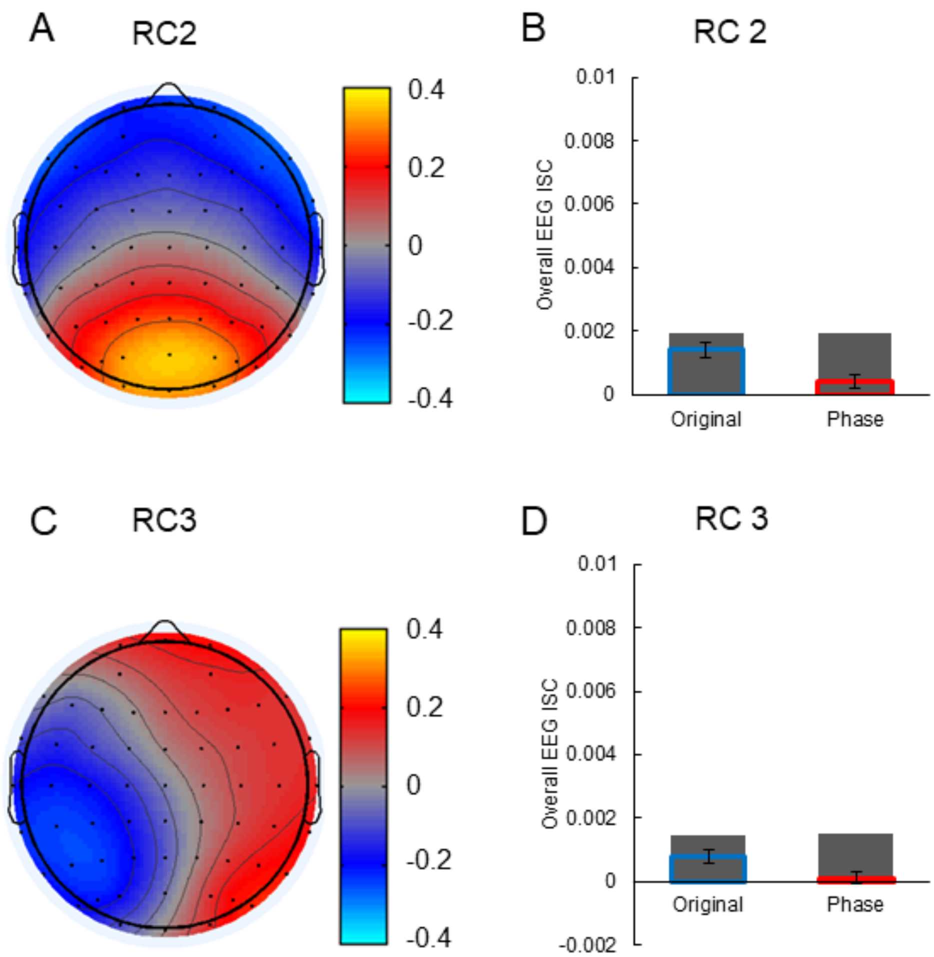


Figure S1. EEG components and overall ISCs for each stimulus. (A) Spatial filtering components RC2 (A) and RC3 (C) are visualized. EEG ISC was computed across the entire duration of each stimulus on RC2 (B) and RC3 (D). Bar height represents the mean value, and error bar height represents ± SEM. The shaded gray area denotes the 95th percentile of the null distribution.

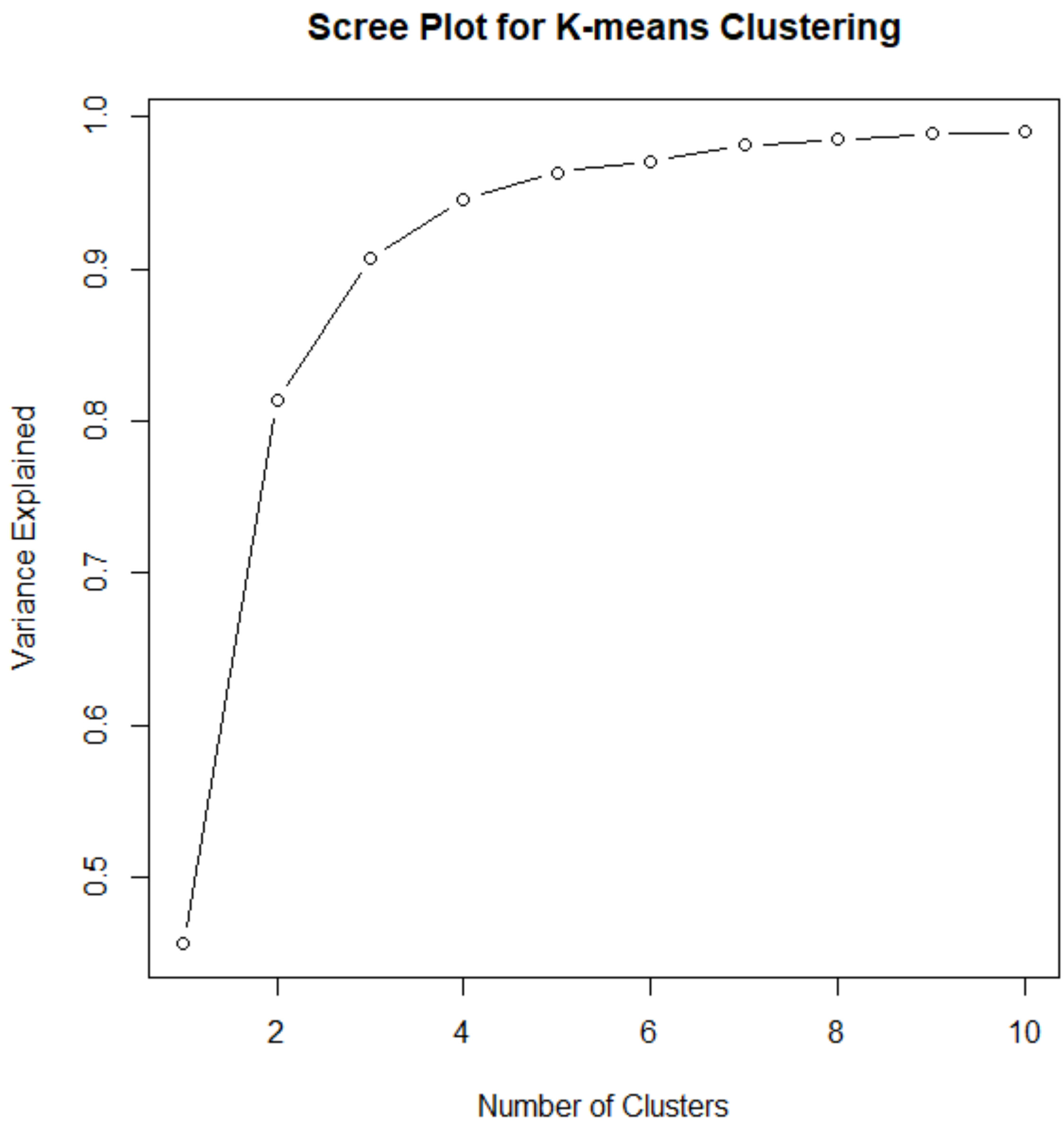


Figure S2. Explained variance by number of clusters for k-means clustering. According to the Elbow method optimal k = 4, and the explained variance was about .95.

Table S1. Frequency and proportion of musical sections by EEG ISC levels.

Musical sections	Level 1	Level 2	Level 3	Level 4	Total
S1: Introduction	18(17.14%)	43(40.95%)	35(33.33%)	9(8.57%)	105(100%)
S2: Exposition I-Primary theme	0(0%)	1(7.69%)	12(92.31%)	0(0%)	13(100%)
S3: Exposition I-Transition	0(0%)	4(26.67%)	9(60%)	2(13.33%)	15(100%)
S4: Exposition I-Secondary theme 1	0(0%)	0(0%)	7(20.59%)	27(79.41%)	34(100%)
S5: Exposition I-Secondary theme 2	0(0%)	0(0%)	9(47.37%)	10(52.63%)	19(100%)
S6: Exposition I-Closing zone	2(10.53%)	10(52.63%)	5(26.32%)	2(10.53%)	19(100%)
S7: Exposition II-Primary theme	0(0%)	1(8.33%)	11(91.67%)	0(0%)	12(100%)
S8: Exposition II-Transition	0(0%)	8(50%)	8(50%)	0(0%)	16(100%)
S9: Exposition II-Secondary theme 1	0(0%)	0(0%)	8(24.24%)	25(75.76%)	33(100%)
S10: Exposition II-secondary theme 2	0(0%)	0(0%)	6(31.58%)	13(68.42%)	19(100%)
S11: Exposition II-Closing zone	5(21.74%)	4(17.39%)	8(34.78%)	6(26.09%)	23(100%)
S12: Development-Introduction	2(4.55%)	6(13.64%)	23(52.27%)	13(29.55%)	44(100%)
S13: Development-Development	0(0%)	0(0%)	19(82.61%)	4(17.39%)	23(100%)
S14: Development-Retransition	0(0%)	0(0%)	8(36.36%)	14(63.64%)	22(100%)
S15: Recapitulation-Primary theme	0(0%)	0(0%)	6(33.33%)	12(66.67%)	18(100%)
S16: Recapitulation-Secondary theme 1	0(0%)	0(0%)	11(39.29%)	17(60.71%)	28(100%)
S17: Recapitulation-Secondary theme 2	0(0%)	1(5.26%)	7(36.84%)	11(57.89%)	19(100%)
S18: Recapitulation-Closing zone	3(13.04%)	4(17.39%)	9(39.13%)	7(30.43%)	23(100%)
S19: Coda	2(3.51%)	11(19.30%)	22(38.60%)	22(38.60%)	57(100%)

Table S2. Results of post-hoc pairwise for the composition of the musical sections by ISC levels.

	L1:L2	L1:L3	L1:L4	L2:L3	L2:L4	L3:L4
S1:S2	1	0.147143	1	0.02301	1	0.528954791
S1:S3	1	0.22687	0.4384976	0.438498	0.778276765	1
S1:S4	1	0.339472	1.06514E-05	0.047525	9.48266E-11	1.85353E-05
S1:S5	1	0.22687	0.007425668	0.027675	6.29655E-05	0.110826646
S1:S6	1	1	1	1	1	1
S1:S7	1	0.144357	1	0.03659	1	0.526959781
S1:S8	0.363161	0.339472	1	1	0.807860085	0.781119147
S1:S9	1	0.339472	1.12952E-05	0.047525	2.31874E-10	0.000115987
S1:S10	1	0.495415	0.001949125	0.082842	6.21583E-06	0.008333558
S1:S11	0.452365	1	0.722730658	0.60076	0.072599296	0.488890484
S1:S12	1	0.099673	0.016991256	0.023559	0.002142744	0.4384976
S1:S13	1	0.023406	0.130980205	0.000153	0.023497701	1
S1:S14	1	0.339472	0.000846246	0.047525	2.70765E-06	0.013490802
S1:S15	1	0.495415	0.001999585	0.082842	1.06514E-05	0.013490802
S1:S16	1	0.144357	0.000202926	0.009092	1.79153E-07	0.013490802
S1:S17	1	0.339472	0.003703124	0.150066	0.000130449	0.031938328
S1:S18	1	1	0.27804878	0.438498	0.036589501	0.363160978
S1:S19	1	0.134431	0.000885301	0.196446	0.000261264	0.057758737
S2:S3	1	1	1	0.780389	1	0.600759869
S2:S4	1	1	1	1	0.181400283	8.30398E-05
S2:S5	1	1	1	1	0.339472188	0.038120921
S2:S6	1	0.421405	1	0.024035	1	0.421404682
S2:S7	1	1	1	1	1	1
S2:S8	1	1	1	0.126391	1	1
S2:S9	1	1	1	1	0.191560866	0.000195388
S2:S10	1	1	1	1	0.291974957	0.003228211
S2:S11	1	0.192513	1	0.48933	1	0.112905384
S2:S12	1	1	1	0.93347	0.826752619	0.127553953
S2:S13	1	1	1	0.934557	0.57126111	0.704153482
S2:S14	1	1	1	1	0.27804878	0.00530826
S2:S15	1	1	1	1	0.304728617	0.006717029
S2:S16	1	1	1	1	0.243589744	0.0062857
S2:S17	1	1	1	1	0.481238274	0.010004044
S2:S18	1	0.60076	1	0.780389	0.952115813	0.079099486
S2:S19	1	1	1	0.438498	0.826752619	0.022000014
S3:S4	1	1	1	0.657671	0.007232687	0.008655878
S3:S5	1	1	1	0.402589	0.066126672	0.420569319
S3:S6	1	0.522832	1	0.431748	1	1
S3:S7	1	1	1	0.780389	1	1
S3:S8	1	1	1	1	0.49741435	1
S3:S9	1	1	1	0.435506	0.008655878	0.016853019

S3:S10	1	1	1	0.666816	0.03745838	0.127553953
S3:S11	0.376302	0.23787	1	1	1	0.634285714
S3:S12	1	1	1	1	0.522831832	1
S3:S13	1	1	1	0.126391	0.304728617	1
S3:S14	1	1	1	0.435506	0.032085561	0.143808138
S3:S15	1	1	1	0.666816	0.044117647	0.127553953
S3:S16	1	1	1	0.357804	0.022000014	0.162568133
S3:S17	1	1	1	1	0.128308573	0.23671701
S3:S18	0.638182	0.625371	1	1	0.799036542	0.628993196
S3:S19	1	1	1	1	0.528954791	0.339472188
S4:S5	1	1	1	1	1	0.263911816
S4:S6	1	1	0.091935484	0.047525	1.06514E-05	0.106223822
S4:S7	1	1	1	1	0.181400283	0.000153176
S4:S8	1	1	1	0.235975	7.26552E-06	0.001511824
S4:S9	1	1	1	1	1	1
S4:S10	1	1	1	1	1	1
S4:S11	1	0.400925	0.013490802	0.651478	0.032623211	0.125525371
S4:S12	1	1	0.421286031	0.778277	0.030957495	0.006415675
S4:S13	1	1	1	1	1	0.000262715
S4:S14	1	1	1	1	1	0.625371429
S4:S15	1	1	1	1	1	0.799036542
S4:S16	1	1	1	1	1	0.48933008
S4:S17	1	1	1	1	0.75887574	0.57126111
S4:S18	1	0.681818	0.106223822	0.657671	0.041323176	0.128308573
S4:S19	1	1	0.600759869	0.489014	0.010004044	0.074622056
S5:S6	1	0.522832	0.27804878	0.023559	0.00247222	0.908321857
S5:S7	1	1	1	1	0.339472188	0.038120921
S5:S8	1	1	1	0.128309	0.000712021	0.088323188
S5:S9	1	1	1	1	1	0.431395746
S5:S10	1	1	1	1	1	1
S5:S11	1	0.23787	0.18	0.373534	0.334365325	1
S5:S12	1	1	1	0.758402	0.280322763	0.683215235
S5:S13	1	1	1	1	1	0.132181695
S5:S14	1	1	1	1	1	1
S5:S15	1	1	1	1	1	1
S5:S16	1	1	1	1	1	1
S5:S17	1	1	1	1	1	1
S5:S18	1	0.625371	0.593406593	0.402589	0.339472188	1
S5:S19	1	1	1	0.322941	0.206197498	1
S6:S7	1	0.438498	1	0.042831	1	0.4384976
S6:S8	1	0.571261	1	1	1	0.57126111
S6:S9	1	0.571261	0.103146677	0.02934	1.62001E-05	0.147143317
S6:S10	1	1	0.209558824	0.090113	0.000668262	0.346513975
S6:S11	0.489014	1	1	0.431748	0.300138953	1
S6:S12	1	0.571261	0.526959781	0.057134	0.071640171	1

S6:S13	1	0.273948	0.972819216	0.000712	0.066126672	1
S6:S14	1	0.571261	0.192513369	0.02934	0.000424946	0.558676912
S6:S15	1	1	0.231081081	0.090113	0.001817203	0.526959781
S6:S16	1	0.438498	0.152614225	0.011408	0.000153669	0.586353984
S6:S17	1	1	0.249483283	0.147143	0.010491153	0.57126111
S6:S18	0.758402	1	1	0.431748	0.182329004	1
S6:S19	1	0.594333	0.33148165	0.252	0.04793226	0.96623115
S7:S8	1	1	1	0.192513	1	1
S7:S9	1	1	1	1	0.191560866	0.000348666
S7:S10	1	1	1	1	0.291974957	0.006504519
S7:S11	1	0.198211	1	0.778277	1	0.126390572
S7:S12	1	1	1	1	0.826752619	0.128308573
S7:S13	1	1	1	0.900592	0.57126111	0.714336288
S7:S14	1	1	1	1	0.27804878	0.008679997
S7:S15	1	1	1	1	0.304728617	0.007425668
S7:S16	1	1	1	1	0.243589744	0.010958122
S7:S17	1	1	1	1	0.481238274	0.016853019
S7:S18	1	0.60076	1	0.780389	0.952115813	0.128308573
S7:S19	1	1	1	0.438498	0.826752619	0.023010133
S8:S9	1	1	1	0.129666	1.05567E-05	0.003140874
S8:S10	1	1	1	0.233155	0.000202926	0.023497701
S8:S11	0.155549	0.393068	1	1	0.091935484	0.23315508
S8:S12	1	1	1	0.23741	0.023497701	0.322192619
S8:S13	1	1	1	0.009351	0.023558871	1
S8:S14	1	1	1	0.129666	0.000153176	0.029339566
S8:S15	1	1	1	0.233155	0.000290886	0.025525084
S8:S16	1	1	1	0.066127	6.32411E-05	0.033533809
S8:S17	1	1	1	0.52696	0.001879568	0.061171647
S8:S18	0.304729	0.648564	1	1	0.091935484	0.237869702
S8:S19	1	1	1	0.826753	0.012296998	0.103995171
S9:S10	1	1	1	1	1	1
S9:S11	1	0.393068	0.016991256	0.405583	0.038120921	0.211152004
S9:S12	1	1	0.4384976	0.758402	0.03745838	0.019936249
S9:S13	1	1	1	1	1	0.000828317
S9:S14	1	1	1	1	1	0.875454059
S9:S15	1	1	1	1	1	1
S9:S16	1	1	1	1	1	0.69506212
S9:S17	1	1	1	1	0.781119147	0.809943508
S9:S18	1	0.648564	0.120585297	0.435506	0.04793226	0.237234328
S9:S19	1	1	0.635345431	0.322941	0.018376593	0.172283478
S10:S11	1	0.431748	0.082608696	0.651478	0.134431027	0.519933001
S10:S12	1	1	1	1	0.252501867	0.147143317
S10:S13	1	1	1	1	1	0.018578012
S10:S14	1	1	1	1	1	1
S10:S15	1	1	1	1	1	1

S10:S16	1	1	1	1	1	1
S10:S17	1	1	1	1	1	1
S10:S18	1	1	0.280322763	0.666816	0.162568133	0.532356154
S10:S19	1	1	1	0.48889	0.126390572	0.69451219
S11:S12	0.799037	0.17685	0.346513975	0.996705	1	1
S11:S13	1	0.054192	0.628993196	0.106893	0.661359206	0.437652661
S11:S14	1	0.393068	0.069160768	0.405583	0.126390572	0.75887574
S11:S15	1	0.431748	0.097151424	0.651478	0.152614225	0.720798287
S11:S16	1	0.198211	0.042830953	0.343849	0.088323188	0.799036542
S11:S17	1	0.400925	0.18	1	0.4384976	1
S11:S18	1	1	1	1	1	1
S11:S19	0.300139	0.294306	0.127553953	1	1	1
S12:S13	1	1	1	0.281838	1	0.471081546
S12:S14	1	1	1	0.758402	0.147143317	0.252501867
S12:S15	1	1	1	1	0.252501867	0.213050985
S12:S16	1	1	0.594332501	0.494038	0.126390572	0.304728617
S12:S17	1	1	1	1	0.57126111	0.343849144
S12:S18	1	0.758402	0.836008896	1	1	1
S12:S19	1	1	1	0.909898	1	0.678178974
S13:S14	1	1	1	1	1	0.026066713
S13:S15	1	1	1	1	1	0.031938328
S13:S16	1	1	1	1	1	0.036736439
S13:S17	1	1	1	0.746971	1	0.065625297
S13:S18	1	0.227221	1	0.126391	1	0.461273527
S13:S19	1	1	1	0.038121	0.746971157	0.110030941
S14:S15	1	1	1	1	1	1
S14:S16	1	1	1	1	1	1
S14:S17	1	1	1	1	1	1
S14:S18	1	0.648564	0.253458498	0.435506	0.144676851	0.781119147
S14:S19	1	1	1	0.322941	0.126390572	0.978339863
S15:S16	1	1	1	1	1	1
S15:S17	1	1	1	1	1	1
S15:S18	1	1	0.307492507	0.666816	0.187431494	0.753953668
S15:S19	1	1	1	0.48889	0.13063846	0.696679186
S16:S17	1	1	1	0.96	0.949780143	1
S16:S18	1	0.60076	0.198211005	0.357804	0.108791209	0.826752619
S16:S19	1	1	1	0.198556	0.071640171	1
S17:S18	1	0.681818	0.339472188	1	0.483226038	1
S17:S19	1	1	1	0.920734	0.4384976	1
S18:S19	0.73415	0.758876	0.4384976	1	1	1

Note. L1 = level 1; L2 = level 2; L3 = level 3; L4 = level 4. S1-S19 means Section 1 to Section 19. Significance levels are shaded with gray: FDR_adjusted *p*_value <.05 and white: FDR_adjusted *p*_value > .05.

Table S3. Frequency and proportion of music section clusters by EEG ISC levels.

Cluster	Level 1	Level 2	Level 3	Level 4	Total
Cluster 1	0(0%)	1(0.52%)	62(32.29%)	129(67.19%)	192(100%)
Cluster 2	0(0%)	2(4.17%)	42(87.5%)	4(8.33%)	48(100%)
Cluster 3	12(8.16%)	25(17.01%)	62(42.18%)	48(32.65%)	147(100%)
Cluster 4	20(12.90%)	65(41.94%)	57(36.77%)	13(8.39%)	155(100%)

Note. Cluster 1: S4, S5, S9, S10, S14, S15, S16, S17; Cluster 2: S2, S7, S13; Cluster 3: S11, S12, S18, S19; Cluster 4: S1, S3, S6, S8.

References

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in quantitative methods for psychology, 8*(1), 23.

Koo, T. K., & Li, M. Y. (2016, 2016/06/01/). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine, 15*(2), 155-163. https://doi.org/https://doi.org/10.1016/j.jcm.2016.02.012