Inter-Subject EEG Correlation Reflects Time-Varying Engagement with Natural Music

Blair Kaneshiro[*1,2], Duc T. Nguyen[1,2,3], Anthony M. Norcia[4,5], Jacek P. Dmochowski[†3,4], and Jonathan Berger[†1]

[1] *Center for Computer Research in Music and Acoustics, Stanford University, Stanford, CA, USA*

[2] *Center for the Study of Language and Information, Stanford University, Stanford, CA, USA*

[3] *Department of Biomedical Engineering, City College of New York, New York, NY, USA*

[4] *Department of Psychology, Stanford University, Stanford, CA, USA*

[5] *Wu Tsai Neurosciences Institute, Stanford, CA, USA*

April 15, 2021

[*]Corresponding author: blairbo@ccrma.stanford.edu. Center for Computer Research in Music and Acoustics, 660 Lomita Drive, Stanford, CA 94305 USA

[†]JPD and JB are joint senior authors on this work.

Abstract

Musical engagement can be conceptualized through various activities, modes of listening, and listener states—among these a state of focused engagement. Recent research has reported that this state can be indexed by the inter-subject correlation (ISC) of EEG responses to a shared naturalistic stimulus. While statistically significant ISC has been reported during music listening, these reports have considered only correlations computed across entire excerpts and do not provide insights into time-varying engagement. Here we present the first EEG-ISC investigation of time-varying engagement *within* a musical work. From a sample of 23 adult musicians who listened to a cello concerto movement, we find varying levels of ISC throughout the excerpt. In particular, significant ISC is observed during periods of musical tension that build to climactic highpoints, but not at the highpoints themselves. In addition, we find that a control stimulus retaining low-frequency envelope characteristics of the intact music, but little other temporal structure, also elicits significant neural correlation, though to a lesser extent than the original. In all, our findings shed light on temporal dynamics of listener engagement during music listening, establish connections between salient musical events and EEG-ISC, and clarify specific listener states that are indexed by this measure.

Keywords Electroencephalography (EEG) Inter-subject correlation Music cognition Engagement Neural decoding

Conflicts of interest None

1 Introduction

Musical engagement can implicate diverse aspects of the musical experience, from creation to interpretation to reception (Kaneshiro, 2016). Engaging with music is typically a hedonic experience able to evoke affective responses ranging from core affects to implicit empathy (Huron and Vuoskoski, 2020). Most studies of engagement focus on the listener, who can engage with music in a variety of ways including, among others, denotative, connotative, reflexive, and associative modes of listening (Huron, 2002). A broadly encompassing definition of listener engagement with music is suggested by Schubert et al. (2013) as being 'compelled, drawn in, connected to what is happening, [and] interested in what will happen next'. In this paper we seek to identify this temporal process of attraction, attention, and anticipation by studying time-varying engagement with natural music across multiple listeners.

A number of studies have considered self-reports of musical engagement, including retrospective ratings delivered after listening to excerpts (Agres et al., 2017; Kaneshiro et al., 2020), questionnaires elucidating listeners' general tendencies to enter states of engagement or absorption with music (Sandstrom and Russo, 2013), and continuous responses delivered as music plays (Gregory, 1989; Broughton et al., 2019; Olsen et al., 2014). However, these approaches have known limitations. Retrospective reports and general listener profiles will not provide insights into specific stimulus events, while self-reports are subject to bias (Rosenman et al., 2011) and limited by a listener's ability to self-assess (Madsen et al., 1993), and thus may be challenging or distracting to deliver—particularly continuously over the course of a stimulus. On the other hand, a time-varying measure of engagement computed from responses delivered without effort by listeners—such as neural responses—could unobtrusively, and perhaps more objectively, index engagement with music as it is heard.

Over the past decade, a neural measure has been proposed as an index of engagement with naturalistic stimuli, including music. Electroencephalography (EEG) is a popular modality

Abbreviations: False Discovery Rate (FDR); inter-subject correlation (ISC); Reliable Components Analysis (RCA).

for studying neural processing of music, as its temporal resolution is sufficiently high to investigate brain responses to specific musical events. Experimental paradigms requiring repeated presentations of short, often tightly controlled stimuli tend to be incompatible with the use of naturalistic musical works and single stimulus exposures (Kaneshiro, 2016). However, a metric based on inter-subject correlation (ISC)—in which audience members' EEG responses to a shared stimulus are correlated with one another—permits both of these experimental design factors. Recent studies have shown that higher EEG-ISC reflects heightened engagement with real-world film excerpts (Dmochowski et al., 2012) as well as television episodes and commercials (Dmochowski et al., 2014). A recent music study by Madsen et al. (2019) showed that EEG-ISC can be modulated by repeated exposures, musical training, and familiarity of musical genres; while more recent work by Kaneshiro et al. (2020) used ISC and related measures to uncover evidence of temporal stimulus manipulations, beat processing, and repetition effects during music listening.

The use of temporally manipulated stimuli to study neural correlation during music listening dates back nearly a decade. One manipulation used in both fMRI (Abrams et al., 2013) and EEG (Kaneshiro et al., 2020) studies is phase randomization of the music audio spectrum, which leaves aggregate power spectra unchanged but disrupts the temporal structure of the audio. In both of these previous studies, phase-scrambled excerpts elicited lower neural correlation, and implicated different brain areas, than intact natural music. Thus, auditory stimulation alone is not thought to explain the neural correlation observed during music listening. However, as noted by both studies, phase scrambling removes numerous structural elements of music, including but not limited to beat, phrase structure, instrumentation, and variations in loudness. Moreover, Kaneshiro et al. (2020) found that stimuli subjected to other, less extreme forms of temporal manipulation—i.e., stimulus reversal and shuffling at the measure (four-beat) level—elicited responses with comparable and even higher ISC than intact music. Therefore, there remain numerous musical attributes whose contributions to neural correlation have yet to be understood.

One advantage of the EEG-ISC approach is that the temporal resolution of EEG permits ISC to be computed over short time windows, producing a time-varying measure of engagement on the scale of seconds. Dmochowski et al. (2012) applied this approach in their first EEG-ISC study on film viewing, finding that ISC not only varied over the course of a film excerpt, but also peaked during periods of high narrative tension and suspense. Cohen et al. (2017) also found that temporally resolved EEG-ISC predicted real-world engagement with videos. As musical tension is thought to share the same underlying states as narrative tension in film excerpts—of dissonance and uncertainty seeking more stable ground (Lehne and Koelsch, 2015)—it is plausible that time-resolved ISC could highlight specific moments of high engagement during music listening as well. However, current EEG-ISC studies of music listening have considered measures of neural correlation across entire excerpts only, on the order of minutes (Madsen et al., 2019; Kaneshiro et al., 2020). While these aggregate measures have provided useful insights into music listening, they overlook temporal dynamics of musical engagement, which likely varies over the course of an excerpt. To the best of our knowledge, no investigation of time-varying musical engagement has yet been attempted using EEG-ISC.

In the present study, we build off these recent works and use EEG-ISC to study time-varying engagement with natural music. As stimuli we used a real-world musical work whose buildups to climatic events we consider conducive to Schubert et al. (2013)'s definition of engagement, in conjunction with a control stimulus providing incrementally more temporal structure than phase-scrambled stimuli used previously (Abrams et al., 2013; Kaneshiro et al., 2020). We took an exploratory approach based on past investigations of temporally resolved EEG-ISC during film viewing (Dmochowski et al., 2012) and conducted ISC analyses over three maximally correlated EEG components. We hypothesized first that the ISC would vary over the course of the original excerpt and peak at preselected musically salient events such as buildups to musical 'highpoints' (Agawu, 1984), which we consider analogous to periods of tension and suspense in more explicitly narrative stimuli such as films (Dmochowski

et al., 2012; Lehne and Koelsch, 2015). Second, based on the importance of the amplitude envelope in speech processing (Van Tasell et al., 1987; Drullman et al., 1994; Shannon et al., 1995; van der Horst et al., 1999; Aiken and Picton, 2008), we hypothesized that our control stimulus—which applies the envelope of the intact music onto a phase-scrambled version—would elicit responses with higher neural correlation than phase scrambling alone. However, we also predicted that fluctuations in audio amplitude envelope would play a role in, but not fully explain, neural correlation during music listening, and thus expected that intact music would elicit higher ISC.

2 Methods

2.1 Ethics Statement

This research was approved by the Institutional Review Board at Stanford University. All participants delivered written informed consent prior to participating in the experiment.

2.2 Data Availability Statement

The data that support the findings of this study (i.e., raw, continuous EEG recordings; preprocessed, aggregated EEG trials; and behavioral ratings with anonymized participant identifiers) are openly available from the Stanford Digital Repository as part of the Naturalistic Music EEG Dataset—Elgar (NMED-E) at `https://purl.stanford.edu/pp371jh5722` (Kaneshiro et al., 2021).

2.3 Stimuli

2.3.1 Stimulus Selection and Characterization

We used two stimuli in this study: The complete opening movement of Edward Elgar's Cello Concerto in E minor, Op. 85, performed by Jacqueline du Pré and the London Symphony

Orchestra, conducted by Sir John Barbirolli, as well as a temporally manipulated version of the performance. Composed in 1919, Elgar's concerto is a dramatic work with a wide dynamic and expressive range, and highly contrasting textures. The work comprises two waves of dramatic gestures that sweep to culminating 'highpoints' (Agawu, 1984), which then lead to structural segmentation boundaries. du Pré's 1965 recording is known for reviving the work and bringing it into the standard orchestral repertoire (Solomon, 2009). The concerto, and in particular the recording we chose, is widely noted as being dramatic, highly expressive, and arousing.[1]

The movement of the concerto is in ternary form with the framing sections in E minor, and the middle section primarily in E major. The sections are linked through two similar motivic elements, which both share a long-short-long rhythmic pattern encompassing neighboring melodic motion. The unifying motivic variants pervade the movement varying in tempi, underlying harmony, melody, orchestration, and with highly contrasting levels of intensity. The A and A' sections each include a climactic event preceded by the main theme played by the soloist, followed by textural and dynamic increases that culminate in a melodic ascent to a cadence.

To guide our data analysis, we identified in advance a set of salient musical events (shown as dashed vertical lines in Figure 1A–B), which we conjectured would elicit significantly correlated ISC responses. First, we labeled as E1 and E1' the two introductions of the main theme played by the soloist, which signify the start of each buildup to climactic highpoints in the movement. The theme is then taken up by the strings at E2 and E2', and in E3 is played by the soloist, now definitively in the tonic key of E minor (E3 is not reprised in the second buildup). Finally, an ascending melodic line played by the soloist reaches a climax with an E-minor cadence at E4 and E4', which are the structural highpoints. We predicted that ISC

[1]The present analysis is not focused specifically on arousal. However, the data analyzed here are part of a larger study aimed to characterize both engagement and arousal during music listening (see § 2.4). Thus, the Elgar concerto was also chosen due in part to its past self-selection by experimental participants as an excerpt arousing strong emotions (Grewe et al., 2007) and use in a subsequent study of frisson, or 'chills' (Grewe et al., 2010).

would be significant between E1 and E4—and later between E1' and E4'—because these are regions of rising tension and suspense which ultimately culminate in the cadential highpoints (cf. Schubert et al. (2013)'s definition of engagement and ISC peaks reported by Dmochowski et al. (2012)). Finally, we labeled as E5 the start of the B section of the movement. Here we anticipated that the cessura at the point of arrival and subsequent introduction of new musical themes would elicit high ISC.

2.3.2 Stimulus Preparation

Both stimuli were 480 seconds (8 minutes) in length. We purchased the EMI Records, Ltd. digitally remastered version of the recording (Elgar et al., 1999) in digital format from iTunes. Using Audacity recording and editing software,[2] we remixed the .m4a stereo recording to mono, and exported the result to .wav format. Subsequent audio processing was performed using Matlab software.[3] First, a linear fade-in and fade-out were applied to the first and last 1 second of the recording, respectively; the resulting audio served as the *Original* stimulus.

The audio was processed further to create the *Control* stimulus. First, we performed the phase-scrambling procedure described in Kaneshiro et al. (2020), converting the audio waveform to the frequency domain, randomizing phase values at each frequency bin while preserving conjugate symmetry, and converting the result back to the time domain. Following this, we used the publicly available MIRtoolbox (Lartillot and Toiviainen, 2007) to compute the audio envelope of the Original stimulus, and subsequently scaled the phase-scrambled audio by this envelope. The result is reminiscent of envelope-shaped noise stimuli that have been used in past speech studies (Horii et al., 1971; Van Tasell et al., 1987; Shannon et al., 1995). Finally, we scaled the overall amplitude of the Control waveform so that its global RMS equaled that of the Original.

Both stimuli are visualized in Figure 1. As shown in the Panels A and B, Original (blue) and Control (red) waveforms are similar but not identical. We also visualized the stimuli

[2] `https://www.audacityteam.org/`
[3] `https://www.mathworks.com/`

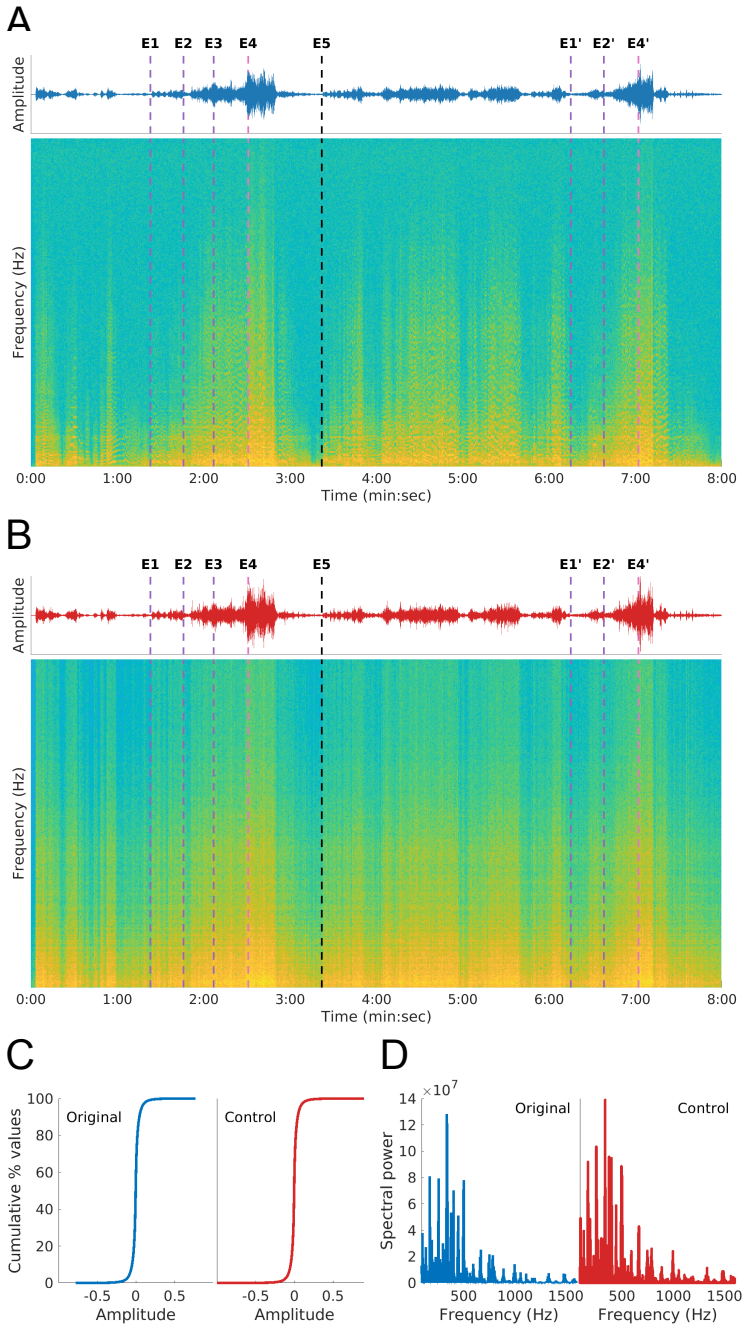

Figure 1: **Experimental stimuli.** Two stimuli were used in the study: **(A)** An influential recording of the first movement of Elgar's Cello Concerto (1919), performed in 1965 by Jacqueline du Pré and conducted by Sir John Barbirolli, and **(B)** a phase-scrambled version of the excerpt which was subsequently scaled by the amplitude envelope of the original. Preselected stimulus events include instances of the main theme played by the solo cello (E1, E1'), strings (E2, E2'), and again by the soloist (E3); structural highpoints (E4, E4'); and start of second section in the movement (E5). **(C)** Cumulative distributions of waveform values are similar across stimuli. **(D)** Stimuli share similar aggregate power spectra.

as spectrograms,[4] which highlight spectral harmonics in the Original but not the Control (e.g., around 1:00). While phase scrambling alone preserves the exact power spectrum of the original audio, the envelope-scaling procedure introduces a similar but not identical power spectrum. Panels C and D show that the two stimuli remain similar on the basis of cumulative distributions of waveform values (computed from histograms with 500 bins each) and aggregate power spectra, respectively.

2.4 Experimental Paradigm

The data analyzed here are part of a larger study on multimodal listener responses to natural music. The larger study comprised a demographic and music experience questionnaire, the Absorption in Music Scale (AIMS) questionnaire (Sandstrom and Russo, 2013), and two experimental blocks of music listening. Each participant's data were collected in a single experimental session: After the participant delivered informed consent, they filled out the questionnaires and then completed the music-listening blocks. In the first block, after completing a short training session to become familiarized with the experimental and task paradigms, the participant heard each stimulus once while EEG, electrocardiogram (ECG), and chest and abdomen respiratory plethysmography were recorded simultaneously. Each stimulus was preceded by a 60-second 'baseline' period to be used during analysis of physiological responses. During this interval, pink noise was played at a low volume while the participant sat quietly and performed no task. Each baseline period was followed by a stimulus, to which the participant listened attentively while avoiding movement and viewing a fixation image shown on a monitor 57 cm in front of them. After each stimulus, participants delivered key-press ratings on a scale of 1 to 9 on the degree of pleasantness, arousal, level of interest, predictability, and familiarity of the stimulus just heard. After the intact excerpt only, participants also used the same response scale to report how often they listened to the genre of classical music. The EEG electrode net and physiological sensors were removed after

[4] `https://www.ee.columbia.edu/~dpwe/LabROSA/projects/`

the first block. In the second block, the participant used a mouse-operated onscreen slider to report their level of engagement over time with each stimulus as it played. In each block, stimuli were presented in a pseudo-random order ensuring an equal distribution of orderings across the original 24 participants. Our present analysis considers demographic information as well as the EEG data and behavioral ratings collected in the first experimental block; thus, further descriptions of data acquisition, preprocessing, and analysis will focus only on these responses.

2.5 Participants

We recruited right-handed participants who were 18–35 years of age, had normal hearing, had no cognitive or decisional impairments, and were fluent in English. As past research has shown that formal musical training is associated with enhanced cortical responses to music (Pantev et al., 1998), we sought participants who had at least five continuous or non-continuous years of formal training in classical music; this could include private lessons in instrumental or vocal performance, composition lessons, and AP- or college-level music theory courses. We recruited participants who enjoyed listening to classical music at least occasionally, but sought participants with no cello experience in order to avoid involuntary motor activations that can occur when hearing music played by one's own instrument (Haueisen and Knösche, 2001). All participants confirmed their eligibility prior to participating.

From 24 participants who completed the experiment, one participant's data were excluded during preprocessing—prior to spatial filtering and ISC analyses—due to gross noise artifacts in the EEG recording (see § 2.6). We thus obtained usable data from $N = 23$ participants (7 male) ranging in age from 19 to 33 years ($m = 23.26$ years, $s = 4.73$ years). From this sample, years of musical training ranged from 5 to 18 years ($m = 10.70$ years, $s = 3.47$ years), and classical-music listening was confirmed through both the verbal eligibility confirmation and a behavioral report delivered during the experiment (see Figure 4).

2.6 Data Acquisition and Preprocessing

Stimulus presentation and key-press response acquisition were programmed using Matlab's Psychophysics Toolbox (Brainard, 1997). Other aspects of stimulus delivery and EEG data acquisition were as described in Kaneshiro et al. (2020): Mono sound files were played through two magnetically shielded Genelec 1030A speakers, which were located 120 cm from the participant in an acoustically and electrically shielded ETS-Lindgren booth. The second audio channel (not played to participants) contained intermittent square-wave pulses, which were sent directly to the EEG amplifier for precise time stamping of stimulus onsets. EEG was recorded from 128-channel nets using the Electrical Geodesics (EGI) GES 300 platform (Tucker, 1993) and Net Station software at a sampling rate of 1 kHz, referenced to the vertex. Responses to both stimuli, their preceding baseline intervals, delivery of ratings, and a short break after the first trial were acquired in a single recording. Electrode impedances were no higher than 60 kΩ at the start of each EEG recording (Ferree et al., 2001).

Continuous EEG recordings were exported to Matlab file format using Net Station software. Subsequent preprocessing and analysis were performed on a per-recording basis, using custom Matlab code as well as publicly available third-party code as described below. First, the continuous recording was highpass (0.3 Hz, 8th-order Butterworth), notch (59–61 Hz, 8th-order Butterworth), and lowpass (50 Hz, 8th-order Chebyshev) filtered using zero-phase filters, and then downsampled by a factor of 8 to a final sampling rate of 125 Hz. Following this, we extracted baseline and stimulus trial labels; event triggers of timing clicks sent from the audio; and behavioral ratings. Baseline and stimulus trials were then epoched using the precise timing triggers sent from the audio: Baseline trials (which were preprocessed but are not analyzed here) were 60 seconds (7501 time samples) in length, and stimulus trials were 480 seconds (60001 time samples) in length. We performed a median-based DC correction on each of the four epochs and concatenated them in time. EOG channels were computed from electrodes above and/or below the eyes (VEOG) and at the sides of the eyes (HEOG). We retained electrodes 1–124 for further analysis, excluding electrodes on the face;

bad electrodes—identified during the experimental session, by visual inspection of the raw data, or with at least 10% of voltage magnitudes exceeding 50 μV across the recording after ICA artifact removal—were also removed at this point, reducing the number of rows in the data matrix. We used the EEGLAB Toolbox (Delorme and Makeig, 2004) implementation of extended Infomax ICA to identify and remove ocular and ECG artifacts from the data in a semi-automatic fashion (Bell and Sejnowski, 1995; Jung et al., 1998). Ocular artifacts were removed according to EOG correlation thresholds described in Kaneshiro et al. (2020), while ECG artifacts were removed by visual inspection of the first 30 components' time-domain waveforms. Subsequent preprocessing steps are as described in Kaneshiro et al. (2020), including identification and exclusion of recording- and trial-wide bad electrodes; replacement of noisy transients with NaNs; replacement of missing rows from bad electrodes with rows of NaNs (ensuring the same number of rows in each data trial); addition of the vertex reference channel and conversion of the data matrix to average reference; replacement of NaNs with the average of data from neighboring electrodes; and a final, mean-based DC correction. During this process, one participant was excluded from further analysis due to gross noise artifacts throughout their recording.

Following preprocessing of all recordings, the trials for each stimulus were aggregated across participants. This produced an electrode-by-time-by-trial matrix for each stimulus, the size of which was $125 \times 7501 \times 23$ for baseline stimuli and $125 \times 60001 \times 23$ for Original and Control stimuli.

2.7 Analysis of EEG Data

We analyzed EEG responses to the Original and Control stimuli (excluding the baseline intervals) on a per-stimulus basis, using a two-step process first introduced by Dmochowski et al. (2012). First, the EEG data matrices were factorized to compute three spatial components in which ISC was maximized. Following this, we computed ISC across entire excerpts and in a time-resolved fashion.

2.7.1 Reliable Components Analysis

As preparation for the ISC calculations, we first spatially filtered the data—that is, we computed optimal spatial 'components' (linear combinations of electrodes) in which ISC of the projected EEG data would be maximized. For this step, we used Reliable Components Analysis (RCA), which was first introduced in the EEG-ISC study by Dmochowski et al. (2012). This procedure concentrates the most-correlated activity across trials—which might be spatially distributed across the electrode montage—into a small number of components. In the case of time-domain analyses as used here, RCA provides ISC-optimized temporal responses while reducing the spatial dimensionality of the data from 125 electrodes to three component vectors. RCA is similar to PCA in that it involves an eigenvalue decomposition; therefore, the procedure returns multiple component weight vectors (eigenvectors)—which for RCA are sorted in descending order of ISC—and accompanying coefficients (eigenvalues). Following Dmochowski et al. (2012), we analyzed the first three Reliable Components (RC1–RC3). Therefore, each trial of data—an electrode-by-time matrix (125×60001)—was multiplied by an electrode-by-component (125×3) weight matrix to produce a component-by-time matrix (3×60001). ISC was then computed separately over each component vector (1×60001).

We used a publicly available Matlab implementation (Dmochowski et al., 2015)[5] to perform RCA using the procedure and parameters reported in Kaneshiro et al. (2020). RCA components were computed separately for each stimulus. As reported in Kaneshiro et al. (2020), we visualized individual components as scalp topographies using forward-model projections of the weight vectors (Haufe et al., 2014; Parra et al., 2005), and report the scalar coefficient of each component.

2.7.2 Inter-Subject Correlation

Inter-subject correlations (ISCs) were computed across participants on a per-stimulus basis. Each ISC calculation involved $N = 23$ single trials of time-domain response data from a

[5] `https://github.com/dmochow/rca`

single RC. We computed ISC in two ways: First, over the entire duration of the stimulus, as was done in past music studies (Madsen et al., 2019; Kaneshiro et al., 2020), and second, in a temporally resolved fashion. For the latter case we followed the procedure of Dmochowski et al. (2012), computing ISC over a 5-second window that advanced in 1-second increments. This produced an ISC time course with a temporal resolution of 1 second, totalling 476 time windows for each 480-second stimulus. Each participant's ISC was computed as the mean correlation of their trial with all other trials. Following that, we computed the mean ISC across participants.

2.8 Statistical Analyses

Statistical significance of the EEG results was computed in two ways. First, we assessed the standalone statistical significance of each RCA coefficient and ISC using permutation testing. As described in Kaneshiro et al. (2020), this involved 1,000 RCA and subsequent ISC calculations for each stimulus; in each calculation every electrode-by-time trial of data input to RCA was first phase scrambled, which disrupts across-trials temporal covariance while retaining aggregate power spectra and autocorrelation characteristics of individual trials (Prichard and Theiler, 1994). This produced a null distribution of 1,000 surrogate observations for each analysis, the 95th percentile of which served as the threshold for statistical significance. In reporting the significance of RCA coefficients, as well as ISC computed across entire stimulus durations, we corrected for multiple comparisons using False Discovery Rate (FDR; Benjamini and Yekutieli (2001)). For this we report adjusted p_{FDR} values, corrected across 3 comparisons (components) on a per-stimulus basis. We present the time-resolved ISC for each stimulus and RC in relation to the 95th percentile of its null distribution, which is also a time series. Here we do not perform any cluster correction, as temporal dependencies are accounted for through the preservation of autocorrelation in the 1,000 phase-scrambled records (Theiler et al., 1992; Prichard and Theiler, 1994; Lancaster et al., 2018).

The second form of statistical analysis involved across-condition comparisons of ISC and

behavioral ratings. For ISC, we focused on RC1 results only, as this was the only component with statistically significant ISC for both stimulus conditions. For the behavioral ratings and ISC computed across entire stimulus durations, we performed paired, two-tailed Wilcoxon signed-rank tests across the distributions of $N = 23$ participants. As the behavioral analysis involved responses to multiple questions, we corrected for 5 comparisons and report p_{FDR} values. Finally, we wished to know whether the temporally resolved ISC differed significantly according to stimulus. To test this, we performed a paired, two-tailed Wilcoxon signed-rank test across the two vectors of across-participant mean ISC values ($N = 476$ time points per vector). Finally, we correlated the RC1 mean ISC time series across stimulus conditions to determine whether their activations were similar, even if their level differed by condition.

3 Results

3.1 Original and Control Stimuli Elicit Correlated Components

In this study we used inter-subject correlation (ISC) to analyze EEG responses to an intact excerpt of natural music and a temporally manipulated control. Prior to computing ISC, we spatially filtered the data, concentrating activity recorded from 125 electrodes to three components in which ISC was maximized. The RC1–RC3 component topographies, associated coefficients, and ISC computed across the entire duration of each stimulus are shown in Figure 2. As shown in Panel A of the figure, the first component (RC1) is broadly similar across stimulus conditions, with a fronto-central topography.[6] Panel B shows the RCA coefficients (eigenvalues) for the first three RCs in line plots, with the respective 95th percentile of the permutation testing null distribution represented as the height of each shaded gray area. As suggested by these plots, the RC1 coefficient is statistically significant for the Original, but not the Control stimulus (permutation test, Original RC1 $p_{FDR} < 0.001$, Control RC1

[6]The RC1 topography for the Control stimulus was multiplied by -1 prior to plotting. This is due to a known arbitrary sign issue with RCA, which does not impact correlation.

$p_{FDR} = 0.23$, corrected for 3 comparisons per condition). The topographies for RC2 and RC3 differ by stimulus condition; however, their coefficients are not statistically significant (all $p_{FDR} \geq 0.23$).

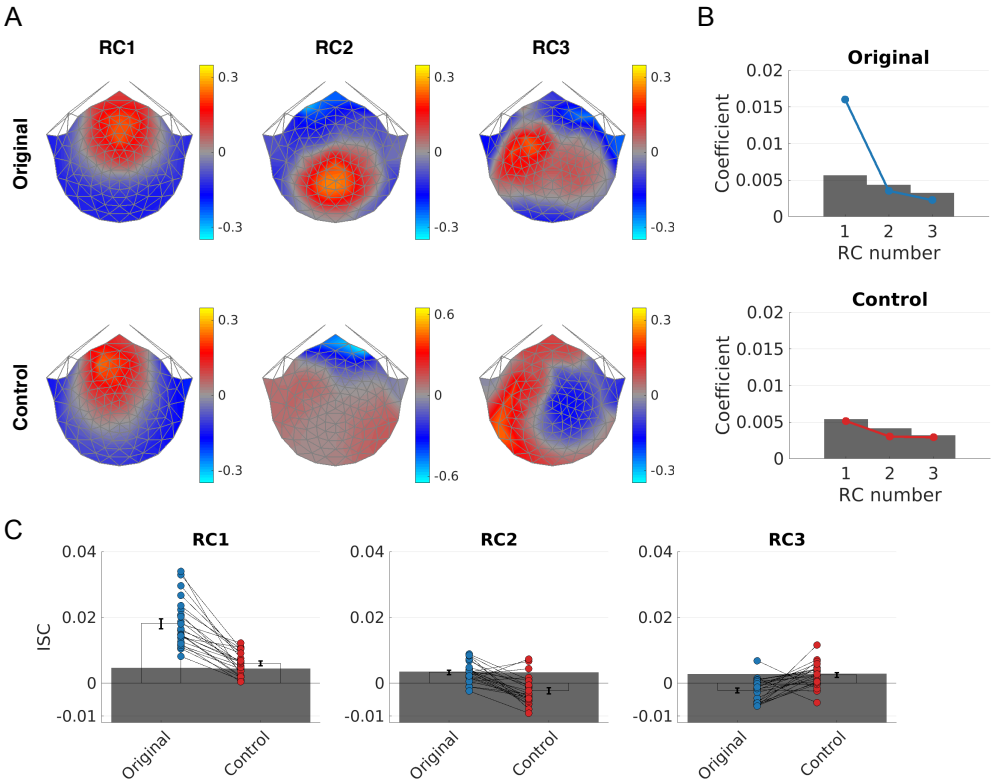

Figure 2: **Aggregate measures of neural correlation. (A)** Spatial filtering components RC1–RC3, computed separately for each stimulus. **(B)** RCA component coefficients with shaded gray area denoting the 95th percentile of the null distribution. The RC1 coefficient for the Original stimulus is statistically significant (permutation test, FDR corrected). **(C)** ISC computed across the entire duration of each stimulus. Error bars represent ± SEM, and points represent individual participants. RC1 ISC is statistically significant for both stimuli (permutation test, FDR corrected) and differs significantly between stimuli (Wilcoxon signed-rank test), with higher ISC for the Original excerpt.

3.2 Aggregate ISC Is Higher for Intact Music

We computed ISC of RC1–RC3 across the entire duration of each stimulus; results for each component and condition are shown in Figure 2C. For RC1, mean ISC exceeds the 95th percentile of the null distribution (upper edge of shaded gray area) for both stimuli (permutation test, Original and Control $p_{FDR} < 0.001$, corrected for 3 comparisons per condition).

However—perhaps unsurprisingly given the RCA coefficients shown in Panel B—all-time ISC is not significant in RC2 or RC3 for either stimulus (all $p_{FDR} \geq 0.12$). A follow-up paired comparison between stimulus conditions for RC1 indicates that ISC differs significantly according to stimulus condition (two-tailed Wilcoxon signed-rank test, $W = 276$, $p < 0.001$), with higher mean ISC for the Original stimulus.

3.3 Temporally Resolved ISC Implicates Salient Musical Events

For each stimulus and RC component, we next computed ISC over 5-second windows, with a temporal resolution (window shift size) of 1 second. Results are shown in Figure 3. As can be seen in Panel A, RC1 ISC of the Original stimulus often exceeds the permutation testing significance threshold (top of shaded gray area), with a total of 37.18% significant time windows (permutation test, uncorrected; see § 2.8). As shown in Panels B and C, however, time-varying ISC rarely reaches statistical significance for RC1 of the Control stimulus, or for RC2 or RC3 of either stimulus.

We observe significant ISC around some, but not all, of the preselected points in the performance: In the first leadup to the highpoint at E4, ISC is significant after each demarcated event (E1, E2, E3). ISC is also significant at the start of the second section of the movement (E5). However, ISC is lower overall in the shorter leadup (E1', E2') to the second highpoint at E4', and is not significant at either highpoint arrival (E4, E4'). Finally, additional notable ISC peaks are observed at the beginning and end of the movement.

Figure 3D shows the overlaid RC1 time-resolved ICS for both stimulus conditions (as noted in Panel C, RC1 was the only component for which the percentage of significant ISC windows exceeded the α level of 0.05 for both stimuli). As suggested by the percentages of significant RC1 ISC reported for each condition, the distributions of temporally resolved ISC differed according to stimulus condition (two-tailed Wilcoxon signed-rank test, $W = 99342$, $p < 0.001$), with higher overall ISC for the Original condition. Finally, we note that while both stimuli elicited high ISC in the early time windows, the two RC1 ISC time series are

not highly correlated ($r = 0.3572$ over all time windows; $r = 0.0497$ when the first four time windows are excluded).

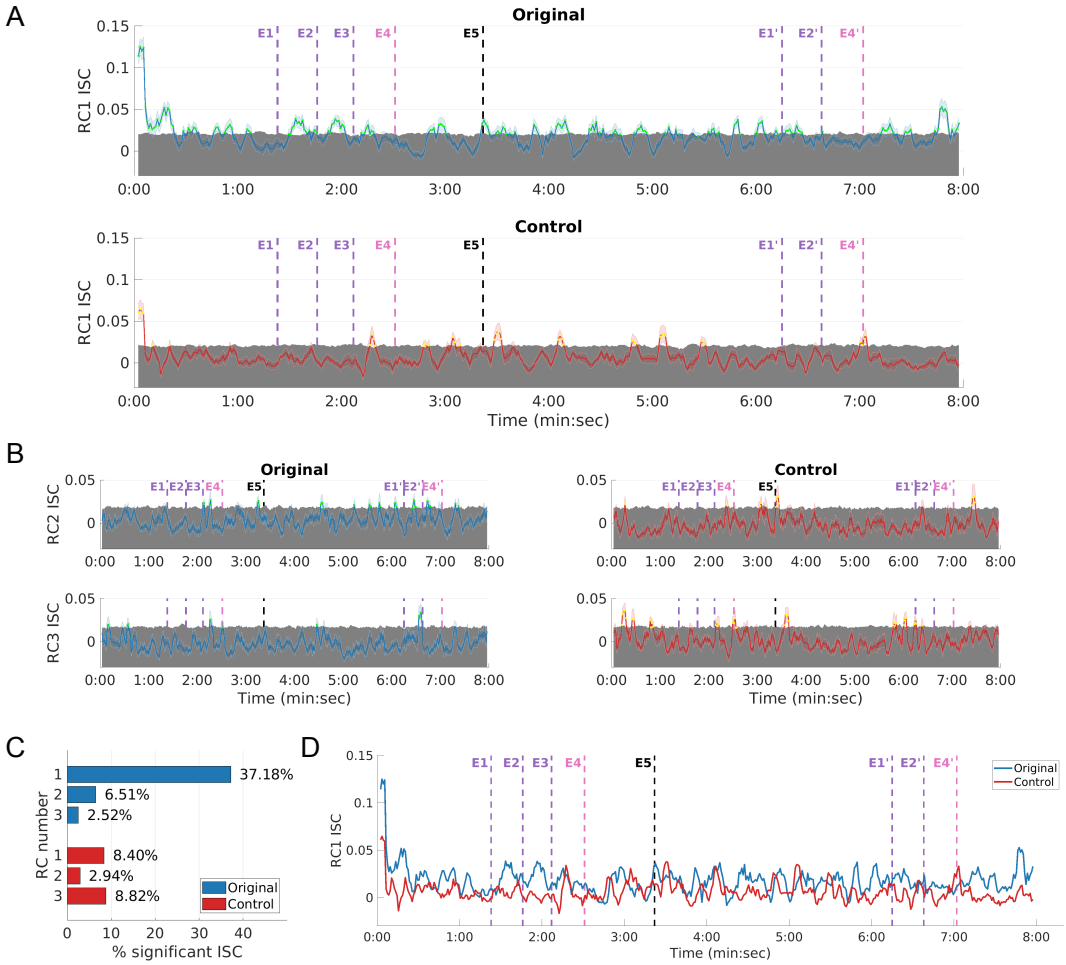

Figure 3: **Time-varying ISC.** ISC was computed in 5-second windows advancing in 1-second increments (colored lines) and plotted in relation to the 95th percentile of the null distribution (shaded gray area) and musically salient events (dashed vertical lines; see Figure 1). **(A)** Temporally resolved ISC for RC1, with line plots representing mean ISC and error bars representing ± SEM. **(B)** Temporally resolved ISC for RC2 and RC3. **(C)** Percentage of statistically significant ISC for each stimulus and RC (permutation test, uncorrected). **(D)** Overlay of mean RC1 ISC for each condition. The Original stimulus elicits higher time-resolved ISC than the Control (Wilcoxon signed-rank test), and time courses are not strongly correlated ($r = 0.3572$ all time windows; $r = 0.0497$ first four time windows excluded).

3.4 Behavioral Ratings Are Higher for Intact Music

Participants reported how pleasant, arousing, interesting, predictable, and familiar they found each stimulus on a scale of 1 to 9. Ratings are shown in Figure 4; y-values of ordinal responses have been slightly jittered for visualization purposes only. Responses differed significantly according to stimulus for all questions (two-tailed Wilcoxon signed-rank test, all $W \geq 55$, all $p_{FDR} \leq 0.044$, corrected for 5 comparisons), with higher mean responses for the Original stimulus in all cases. Participants also reported how often they listened to the genre of the Original stimulus (i.e., classical music); all responded higher than 'Never', confirming this aspect of their eligibility.

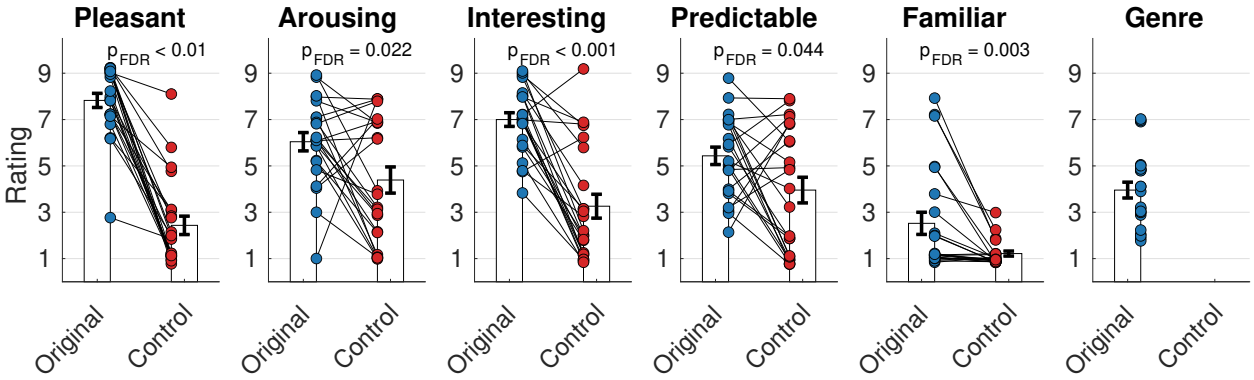

Figure 4: **Behavioral ratings.** Participants rated both stimuli along dimensions of pleasantness, arousal, level of interest, predictability, and familiarity; and reported how often they listen to classical music. Bar heights represent mean values, error bars represent ± SEM, and dots represent individual observations (y-values slightly jittered for visualization only). All between-condition comparisons were statistically significant (Wilcoxon signed-rank test), with higher ratings given for the Original stimulus. All participants listened to classical music at least occasionally.

4 Discussion

Over the past decade, EEG inter-subject correlation (ISC) has been shown to index states of attention and engagement with naturalistic, real-world stimuli such as film excerpts, speeches, and—most recently—musical works. In this study we have for the first time analyzed temporally resolved EEG-ISC in the context of music listening. Through this

approach we have confirmed that the ISC measure of engagement varies over the course of a real-world musical work, as has previously been shown in the context of film viewing (Dmochowski et al., 2012). Moreover, in extending a temporal stimulus manipulation used in past music ISC studies (Abrams et al., 2013; Kaneshiro et al., 2020), we have shown that an audio amplitude envelope characteristic of intact music elicits correlated EEG responses, though not to the extent of intact music.

Whether computed as a single measure over entire stimuli or in shorter time windows, neural correlation methods such as ISC have provided insights into brain states and the stimulus attributes that drive them. EEG-ISC has been shown to index narrative cohesion (Dmochowski et al., 2012; Ki et al., 2016), attentional states (Ki et al., 2016; Cohen et al., 2018), and learning outcomes (Cohen et al., 2018). In addition, EEG-ISC computed over short time windows has revealed connections between periods of heightened engagement and narrative tension. In their first investigation of time-varying EEG-ISC, Dmochowski et al. (2012, Fig. 1) found that temporal peaks in the ISC corresponded to points of high tension and suspense in a film excerpt; Poulsen et al. (2017, Fig. 2) later replicated these findings using mobile EEG. Based on these findings, we expected that ISC would vary during music listening as well, and would similarly peak during periods in which the music induced a state of *engagement*, which we frame using the definition given by Schubert et al. (2013): Being 'compelled, drawn in, connected to what is happening, [and] interested in what will happen next'. Indeed, across an audience of 23 adult musicians listening to the first movement of Elgar's Cello Concerto, we observed temporal variations in ISC with a number of significant peaks (Figure 3).

In designing this study, we prespecified a number of events in the intact stimulus that, based upon the dramatic intensity of the music, we expected would correspond to high ISC (see § 2.3.1). Some but not all of these events corresponded to significant ISC. In the first climactic rise to the structural highpoint (E1–E4), ISC was significantly high leading up to, but not including, the highpoint itself. The significant ISC during periods of escalating

tension is confirmatory with previous findings relating to film viewing (Dmochowski et al., 2012). However, we have observed that at each highpoint's arrival, ISC is not only nonsignificant but also nonsignificant for extended periods of time relative to the rest of the excerpt. While we know of no EEG-ISC study focusing specifically on moments of apotheosis (i.e., the culminating moment of tension and suspense) during film viewing, we posit that the drop in ISC just before the resolution of suspense reflects a state distinct from the preceding, increasingly anticipatory state. We do not suggest that listener engagement ceases or decreases at the musical climax, but rather that the specific nature of engagement at this point is not evident or indexed by ISC. We expect that other types of listener responses—such as the physiological and continuous behavioral responses of our larger study—and EEG analysis techniques could capture this state.

This lack of significant ISC at the very moment and subsequent moments which are deemed climactic may, perhaps, correspond to a conspicuous absence of musicological, music-theoretic, and psychological studies that focus on the affective processes from a point of maximal tension until the actual resolution of tension. Most literature treats this transition as a binary switch in which tension is suddenly resolved. A number of music theorists have sought to nuance the process delineating patterns of fluctuating intensity leading to and following a climax in terms of *intensity* (Berry, 1978) or *narrative* (Childs, 1977; Agawu, 1984). The affective peaks and troughs of these processes are described in terms of intensification, climax, and abatement (Patty, 2009). However there is a considerable lack of consistency and a wide variety of issues, terminology, and analytical approaches, and thus this key aspect of musical engagement begs further study.

The ternary structure of the Original stimulus also provides an opportunity to investigate whether ISC peaks are comparable across repeated periods of musical tension and suspense. 44.29% of the ISC windows in the leadup to the first highpoint (between E1 and E4) are significant—a higher percentage than observed in the second leadup (E1' to E4'; 21.28%). This suggests that repetition effects—in which ISC drops over repeated exposures (Dmo-

chowski et al., 2012; Madsen et al., 2019; Kaneshiro et al., 2020)—may also occur on a local, within-excerpt level. Given the importance of repetition in both musical structure (Margulis, 2012) and formation of musical preferences over repeated listens (Szpunar et al., 2004), this finding highlights avenues for future research with a larger set of stimuli and variety of structural levels (e.g., song, section, phrase) over which musical content is repeated.

Finally, we note that temporally resolved ISC was significant at numerous additional points during the excerpt. Of particular note are the first 21 and last 15 ISC windows. The former case includes two peaks. The first peak, which occurs in the first four time windows centered at 2.5–5.5 seconds inclusive, may be attributable to low-level processing of the onset of the auditory stimulus. However, ISC is markedly higher over these time windows for the Original stimulus compared to the Control, with peak correlations of 0.13 and 0.06, respectively. In addition, the Original stimulus, but not the Control, elicits a second peak at around 19 seconds. As this opening portion of the movement consists of an extended, dramatic solo passage in which the key of the movement is established, we interpret these peaks as reflecting not only low-level processing but also listeners being 'drawn in' to the music and 'interested in what will happen next', consistent with Schubert et al. (2013)'s definition of engagement.

The RC1 topographies for both stimuli are consistent with those reported in recent auditory ISC studies (Cohen and Parra, 2016; Ki et al., 2016; Madsen et al., 2019; Kaneshiro et al., 2020) and other EEG components computed from responses to natural music (Schaefer et al., 2011; Sturm et al., 2015; Gang et al., 2017). In addition, while Kaneshiro et al. (2020) reported an RC1 topography for phase-scrambled stimuli that was dissimilar to components computed from more musical stimuli, here we find that the Control stimulus—in which a phase-scrambled stimulus was scaled by the amplitude envelope of the Original—elicits an RC1 topography that is broadly similar to that computed from intact music.

The Original stimulus elicited higher neural correlation than the Control, as shown by the RCA coefficients (Figure 2B) and ISC (Figure 2C, Figure 3). Even so, the Control elicited

higher neural correlation than did previous phase-scrambled stimuli without amplitude scaling (Kaneshiro et al., 2020), and both stimuli in the present study elicited significant ISC across entire excerpts (Figure 2C). That the present Control stimulus elicits statistically significant ISC in an auditory RC1 topography distinguishes it from previous controls which underwent phase scrambling without envelope scaling (Kaneshiro et al., 2020) and suggests that tracking of the music envelope, even when divorced from other attributes of music, drives correlated cortical responses in auditory brain regions. We acknowledge that one of the four phase-scrambled excerpts reported previously did elicit statistically significant ISC (Kaneshiro et al., 2020, Fig. 3C); however, we cannot make direct comparisons to that finding given the disparate component topography in which it was computed. Future research can further investigate the musical attributes driving neural correlation over a larger set of stimuli and stimulus manipulations.

As shown in Figure 3C, 37.18% of RC1 temporally resolved ISC windows were statistically significant. This quantity is on par with previous reports related to film viewing (between approximately 21% and 33%, Dmochowski et al. (2012); 54.1%, Poulsen et al. (2017)). However, percentages of significant ISC for RC2 and RC3 were markedly lower during music listening compared to film viewing (Dmochowski et al., 2012). This finding is corroborated by sizeable drops in both the RC coefficients (Figure 2B) and all-time ISC (Figure 3C). The drop in RC coefficients has been reported previously by Kaneshiro et al. (2020) and motivated the decision to analyze only RC1 data in that study. Thus, while we took an exploratory approach here in analyzing data from RC1–RC3 (Dmochowski et al., 2012), our between-condition comparisons ultimately focused only on RC1, since RC2 and RC3 ISC were not statistically significant compared to the permutation testing null distribution. The standalone significance of single-component ISC should be carefully assessed before including those values in across-conditions analyses.

Numerous studies have demonstrated that EEG-ISC can derive meaningful measures of brain states related to engagement. Yet at its core, ISC is a group measure, and insights into

individuals are limited to the correlation of one person's responses with the rest of the group. Past work has shown that this one-against-all approach can still reveal significant brain-behavior correlations for individuals (Cohen et al., 2018). Even so, ISC likely misses aspects of time-resolved engagement for individual music listeners. Kaneshiro et al. (2020) have noted that the degree of abstraction embodied by music compared to movie narratives can impact the extent, and musical origin, of correlated responses during music listening. In contrast with film viewing, listeners may be more likely to focus on different musical attributes at different times, which would affect the group-level (i.e., ISC) response. One alternative approach toward indexing individual engagement during music listening is stimulus-response correlation (SRC), which permits investigation of unique experiences by correlating single trials of EEG directly with time-varying stimulus attributes (Dmochowski et al., 2018). Forms of this approach have revealed significant group-level correlations across entire musical works (Gang et al., 2017; Kaneshiro et al., 2020). Future time-resolved applications—perhaps with participant-selected stimuli (Rickard, 2004; Grewe et al., 2007)—could prove useful in deriving single-listener time courses of engagement. Future work can also investigate how the extent of 'narrative' conveyed by music drives ISC. For instance, 'program' music—that is, music conveying an extra-musical narrative, such as Prokofiev's *Peter and the Wolf*, Op. 67 (Prokofiev, 1940)—could bridge between explicit narratives (Ki et al., 2016; Dmochowski et al., 2012), popular music with vocals (Kaneshiro et al., 2020), and 'absolute' (i.e., non-representational) musical works used previously (Madsen et al., 2019) and here.

Neural measures of time-varying musical engagement point to additional directions for future investigation. Our next work will augment present findings with additional responses (ECG, respiratory activity, continuous behavioral reports) collected in the larger study. Other research can extend the analysis of time-varying ISC from our focus—which was on discrete time points corresponding to musical events identified a priori—to continuous measures that can be correlated in time with the ISC to elucidate connections between specific musical events and listener engagement. Such 'second-order' correlations could in-

volve, for example, continuous listener reports of engagement (Broughton et al., 2019; Olsen et al., 2014), computationally extracted feature vectors (Lartillot and Toiviainen, 2007), or time-varying models of musical tension (Farbood, 2012).

Finally, EEG-ISC can contribute not only basic insights into music processing, but also predictive insights toward real-world application (Kaneshiro and Dmochowski, 2015). In addition to permitting the use of ecologically valid stimuli that people experience in real life, the EEG-ISC paradigm has been shown to predict population-level preferences—such as viewership and social media engagement during television viewing, and ratings of commercials (Dmochowski et al., 2014)—from small experimental samples. This speaks to the possibility of predicting real-world *musical* engagement as well. As suggested by current findings, such engagement could be assessed at both the song level—for instance to predict number of streams or overall chart success (Pachet and Roy, 2008; Herremans et al., 2014)—and within-song, e.g., for chorus and hook detection (Van Balen et al., 2015). Recent studies have computed measures of time-varying, within-song engagement from large-scale commercial data using probabilities of Shazam queries (Kaneshiro et al., 2017) and Spotify skips (Montecchio et al., 2020). Thus, future investigations of temporally resolved ISC in conjunction with such reports have the potential to contribute to services aiding both music listeners and creators.

Acknowledgments

The authors thank Tysen Dauer and Daniel Abrams for helpful feedback and discussions on the manuscript, and Kunwoo Kim for contributing to exploratory analyses of the data. This research was funded by the Wallenberg Network Initiative: Culture, Brain Learning (BK, DTN, JB); the Patrick Suppes Gift Fund (BK, DTN); Army Research Laboratory W911-NF-10-2-0022 (JPD); and the Roberta Bowman Denning Fund for Humanities and Technology (BK, JB).

References

Abrams, D. A., Ryali, S., Chen, T., Chordia, P., Khouzam, A., Levitin, D. J., and Menon, V. (2013). Inter-subject synchronization of brain responses during natural music listening. *The European Journal of Neuroscience*, 37(9):1458—1469.

Agawu, V. K. (1984). Structural 'highpoints' in Schumann's 'Dichterliebe'. *Music Analysis*, 3(2):159–180.

Agres, K., Herremans, D., Bigo, L., and Conklin, D. (2017). Harmonic structure predicts the enjoyment of uplifting trance music. *Frontiers in Psychology*, 7:1999.

Aiken, S. J. and Picton, T. W. (2008). Human cortical responses to the speech envelope. *Ear and hearing*, 29(2):139–157.

Bell, A. J. and Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1129–1159.

Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, 29(4):1165–1188.

Berry, W. (1978). *Structural functions in music*. Prentice Hall, Englewood Cliffs, NJ.

Brainard, D. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4):433–436.

Broughton, M. C., Schubert, E., Harvey, D. G., and Stevens, C. J. (2019). Continuous self-report engagement responses to the live performance of an atonal, post-serialist solo marimba work. *Psychology of Music*, 47(1):109–131.

Childs, B. (1977). Time and music: A composer's view. *Perspectives of New Music*, 15(2):194–219.

Cohen, S. S., Henin, S., and Parra, L. C. (2017). Engaging narratives evoke similar neural activity and lead to similar time perception. *Scientific reports*, 7(1):4578.

Cohen, S. S., Madsen, J., Touchan, G., Robles, D., Lima, S. F., Henin, S., and Parra, L. C. (2018). Neural engagement with online educational videos predicts learning performance for individual students. *Neurobiology of Learning and Memory*, 155:60–64.

Cohen, S. S. and Parra, L. C. (2016). Memorable audiovisual narratives synchronize sensory and supramodal neural responses. *eNeuro*, 3(6).

Delorme, A. and Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1):9–21.

Dmochowski, J. P., Bezdek, M. A., Abelson, B. P., Johnson, J. S., Schumacher, E. H., and Parra, L. C. (2014). Audience preferences are predicted by temporal reliability of neural processing. *Nature communications*, 5:4567.

Dmochowski, J. P., Greaves, A. S., and Norcia, A. M. (2015). Maximally reliable spatial filtering of steady state visual evoked potentials. *NeuroImage*, 109:63–72.

Dmochowski, J. P., Ki, J. J., DeGuzman, P., Sajda, P., and Parra, L. C. (2018). Extracting multidimensional stimulus-response correlations using hybrid encoding-decoding of neural activity. *NeuroImage*, 180(Pt A):134–146.

Dmochowski, J. P., Sajda, P., Dias, J., and Parra, L. C. (2012). Correlated components of ongoing EEG point to emotionally laden attention—a possible marker of engagement? *Frontiers in Human Neuroscience*, 6:112.

Drullman, R., Festen, J. M., and Plomp, R. (1994). Effect of temporal envelope smearing on speech reception. *The Journal of the Acoustical Society of America*, 95(2):1053–1064.

Elgar, E., Barbirolli, J., and du Pré, J. (1999). Cello Concerto in E Minor, Op. 85. Cello Concerto: Sea Pictures. EMI Classics.

Farbood, M. M. (2012). A parametric, temporal model of musical tension. *Music Perception*, 29(4):387–428.

Ferree, T. C., Luu, P., Russell, G. S., and Tucker, D. M. (2001). Scalp electrode impedance, infection risk, and EEG data quality. *Clinical Neurophysiology*, 112(3):536–544.

Gang, N., Kaneshiro, B., Berger, J., and Dmochowski, J. P. (2017). Decoding neurally relevant musical features using Canonical Correlation Analysis. In *Proceedings of the 18th International Society for Music Information Retrieval Conference*, pages 131–138.

Gregory, D. (1989). Using computers to measure continuous music responses. *Psychomusicology*, 8(2):127–134.

Grewe, O., Katzur, B., Kopiez, R., and Altenmüller, E. (2010). Chills in different sensory domains: Frisson elicited by acoustical, visual, tactile and gustatory stimuli. *Psychology of Music*.

Grewe, O., Nagel, F., Kopiez, R., and Altenmüller, E. (2007). Listening to music as a re-creative process: Physiological, psychological, and psychoacoustical correlates of chills and strong emotions. *Music Perception*, 24(3):297–314.

Haueisen, J. and Knösche, T. R. (2001). Involuntary motor activity in pianists evoked by music perception. *Journal of Cognitive Neuroscience*, 13(6):786–792.

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., and Bießmann, F. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage*, 87:96—110.

Herremans, D., Martens, D., and Sörensen, K. (2014). Dance hit song prediction. *Journal of New Music Research*, 43(3):291–302.

Horii, Y., House, A. S., and Hughes, G. W. (1971). A masking noise with speech-envelope characteristics for studying intelligibility. *The Journal of the Acoustical Society of America*, 49(6B):1849–1856.

Huron, D. (2002). Listening styles and listening strategies. In *Society for Music Theory 2002 Conference*.

Huron, D. and Vuoskoski, J. K. (2020). On the enjoyment of sad music: Pleasurable compassion theory and the role of trait empathy. *Frontiers in Psychology*, 11:1060.

Jung, T.-P., Humphries, C., Lee, T.-W., Makeig, S., McKeown, M. J., Iragui, V., and Sejnowski, T. J. (1998). Extended ICA removes artifacts from electroencephalographic recordings. *Advances in Neural Information Processing Systems*, pages 894–900.

Kaneshiro, B. (2016). *Toward an Objective Neurophysiological Measure of Musical Engagement*. PhD thesis, Stanford University.

Kaneshiro, B. and Dmochowski, J. P. (2015). Neuroimaging methods for music information retrieval: Current findings and future prospects. In *Proceedings of the 16th International Society for Music Information Retrieval Conference*, pages 538–544.

Kaneshiro, B., Nguyen, D. T., Norcia, A. M., Dmochowski, J. P., and Berger, J. (2020). Natural music evokes correlated EEG responses reflecting temporal structure and beat. *NeuroImage*, 214:116559.

Kaneshiro, B., Nguyen, D. T., Norcia, A. M., Dmochowski, J. P., and Berger, J. (2021). Naturalistic music EEG dataset—Elgar (NMED-E). In *Stanford Digital Repository*.

Kaneshiro, B., Ruan, F., Baker, C. W., and Berger, J. (2017). Characterizing listener engagement with popular songs using large-scale music discovery data. *Frontiers in Psychology*, 8:416.

Ki, J. J., Kelly, S. P., and Parra, L. C. (2016). Attention strongly modulates reliability of neural responses to naturalistic narrative stimuli. *Journal of Neuroscience*, 36(10):3092–3101.

Lancaster, G., Iatsenko, D., Pidde, A., Ticcinelli, V., and Stefanovska, A. (2018). Surrogate data for hypothesis testing of physical systems. *Physics Reports*, 748:1–60. Surrogate data for hypothesis testing of physical systems.

Lartillot, O. and Toiviainen, P. (2007). A Matlab toolbox for musical feature extraction from audio. In *International Conference on Digital Audio Effects*, pages 237–244.

Lehne, M. and Koelsch, S. (2015). Toward a general psychological model of tension and suspense. *Frontiers in Psychology*, 6:79.

Madsen, C. K., Brittin, R. V., and Capperella-Sheldon, D. A. (1993). An empirical method for measuring the aesthetic experience to music. *Journal of Research in Music Education*, 41(1):57–69.

Madsen, J., Margulis, E. H., Simchy-Gross, R., and Parra, L. C. (2019). Music synchronizes brainwaves across listeners with strong effects of repetition, familiarity and training. *Scientific reports*, 9(1):3576.

Margulis, E. H. (2012). Musical repetition detection across multiple exposures. *Music Perception*, 29(4):377–385.

Montecchio, N., Roy, P., and Pachet, F. (2020). The skipping behavior of users of music streaming services and its relation to musical structure. *PLOS ONE*, 15(9):1–16.

Olsen, K. N., Dean, R. T., and Stevens, C. J. (2014). A continuous measure of musical engagement contributes to prediction of perceived arousal and valence. *Psychomusicology: Music, Mind, and Brain*, 24(2):147.

Pachet, F. and Roy, P. (2008). Hit song science is not yet a science. In *Proceedings of the 9th International Conference on Music Information Retrieval*, pages 355–360.

Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., and Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature*, 392(6678):811–814.

Parra, L. C., Spence, C. D., Gerson, A. D., and Sajda, P. (2005). Recipes for the linear analysis of EEG. *NeuroImage*, 28(2):326–341.

Patty, A. T. (2009). Pacing scenarios: How harmonic rhythm and melodic pacing influence our experience of musical climax. *Music Theory Spectrum*, 31(2):235–367.

Poulsen, A. T., Kamronn, S., Dmochowski, J., Parra, L. C., and Hansen, L. K. (2017). EEG in the classroom: Synchronised neural recordings during video presentation. *Scientific reports*, 7:43916.

Prichard, D. and Theiler, J. (1994). Generating surrogate data for time series with several simultaneously measured variables. *Phys. Rev. Lett.*, 73:951–954.

Prokofiev, S. (1940). *Peter and the Wolf*. A. A. Knopf, New York.

Rickard, N. S. (2004). Intense emotional responses to music: A test of the physiological arousal hypothesis. *Psychology of Music*, 32(4):371–388.

Rosenman, R., Tennekoon, V., and Hill, L. G. (2011). Measuring bias in self-reported data. *International Journal of Behavioural and Healthcare Research*, 2(4):320–332.

Sandstrom, G. M. and Russo, F. A. (2013). Absorption in music: Development of a scale to identify individuals with strong emotional responses to music. *Psychology of Music*, 41(2):216–228.

Schaefer, R. S., Farquhar, J., Blokland, Y., Sadakata, M., and Desain, P. (2011). Name that tune: Decoding music from the listening brain. *NeuroImage*, 56(2):843–849.

Schubert, E., Vincs, K., and Stevens, C. J. (2013). Identifying regions of good agreement among responders in engagement with a piece of live dance. *Empirical Studies of the Arts*, 31(1):1–20.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, 270(5234):303–304.

Solomon, J. (2009). Deconstructing the definitive recording: Elgar's Cello Concerto and the influence of Jacqueline du Pré. *Unpublished manuscript*.

Sturm, I., Dähne, S., Blankertz, B., and Curio, G. (2015). Multi-variate EEG analysis as a novel tool to examine brain responses to naturalistic music stimuli. *PLoS ONE*, 10(10):e0141281.

Szpunar, K. K., Schellenberg, E. G., and Pliner, P. (2004). Liking and memory for musical stimuli as a function of exposure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(2):370.

Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., and Farmer, J. D. (1992). Testing for nonlinearity in time series: the method of surrogate data. *Physica D: Nonlinear Phenomena*, 58(1):77–94.

Tucker, D. M. (1993). Spatial sampling of head electrical fields: The geodesic sensor net. *Electroencephalography and Clinical Neurophysiology*, 87(3):154–163.

Van Balen, J. M. H., Burgoyne, J. A., Bountouridis, D., Müllensiefen, D., and Veltkamp, R. C. (2015). Corpus analysis tools for computational hook discovery. In *Proceedings of the 16th International Society for Music Information Retrieval Conference*, pages 227–233.

van der Horst, R., Leeuw, A. R., and Dreschler, W. A. (1999). Importance of temporal-envelope cues in consonant recognition. *The Journal of the Acoustical Society of America*, 105(3):1801–1809.

Van Tasell, D. J., Soli, S. D., Kirby, V. M., and Widin, G. P. (1987). Speech waveform envelope cues for consonant recognition. *The Journal of the Acoustical Society of America*, 82(4):1152–1161.