Inter-subject Correlation While Listening to Minimalist Music: A Study of Electrophysiological and Behavioral Responses to Steve Reich's *Piano Phase*

Tysen Dauer [1,2]*, Duc T. Nguyen [1,3], Nick Gang [1], Jacek P. Dmochowski [4], Jonathan Berger [1] and Blair Kaneshiro [1,3]

[1] Center for Computer Research in Music and Acoustics, Stanford University, Stanford, CA, United States, [2] Rotman Research Institute, Baycrest Centre, Toronto, ON, Canada, [3] Center for the Study of Language and Information, Stanford University, Stanford, CA, United States, [4] Department of Biomedical Engineering, City College of New York, New York, NY, United States

Musical minimalism utilizes the temporal manipulation of restricted collections of rhythmic, melodic, and/or harmonic materials. One example, Steve Reich's *Piano Phase*, offers listeners readily audible formal structure with unpredictable events at the local level. For example, pattern recurrences may generate strong expectations which are violated by small temporal and pitch deviations. A hyper-detailed listening strategy prompted by these minute deviations stands in contrast to the type of listening engagement typically cultivated around functional tonal Western music. Recent research has suggested that the inter-subject correlation (ISC) of electroencephalographic (EEG) responses to natural audio-visual stimuli objectively indexes a state of "engagement," demonstrating the potential of this approach for analyzing music listening. But can ISCs capture engagement with minimalist music, which features less obvious expectation formation and has historically received a wide range of reactions? To approach this question, we collected EEG and continuous behavioral (CB) data while 30 adults listened to an excerpt from Steve Reich's *Piano Phase*, as well as three controlled manipulations and a popular-music remix of the work. Our analyses reveal that EEG and CB ISC are highest for the remix stimulus and lowest for our most repetitive manipulation, no statistical differences in overall EEG ISC between our most musically meaningful manipulations and Reich's original piece, and evidence that compositional features drove engagement in time-resolved ISC analyses. We also found that aesthetic evaluations corresponded well with overall EEG ISC. Finally we highlight co-occurrences between stimulus events and time-resolved EEG and CB ISC. We offer the CB paradigm as a useful analysis measure and note the value of minimalist compositions as a limit case for the neuroscientific study of music listening. Overall, our participants' neural, continuous behavioral, and question responses showed strong similarities that may help refine our understanding of the type of engagement indexed by ISC for musical stimuli.

Keywords: inter-subject correlation (ISC), engagement, continuous behavioral measure, minimalism (music), EEG, music cognition

1. INTRODUCTION

The genre of musical minimalism is (in)famously characterized by highly recurrent, starkly restricted pitch and rhythmic collections. From the early days of scholarship on minimalist, or "repetitive music" as it was often called, commentators described the music's timbral and rhythmic staticity and its limited pitch patterns (Mertens, 1983, p. 12). While many advocates reported what we might call blissing out to this "meditative music" (to use yet another early term for this repertoire), some composers went on record to state their intention that the music should be listened to carefully (Strongin, 1969; Henahan, 1970). For example, the composer Steve Reich wrote in 1968 that he wanted to write works with musical processes that any listener could perceive: works where the process unfolded very gradually in order to "facilitate closely detailed listening" (Reich, 2009, p. 34). Numerous professional musicians and critics have asserted that listeners do not engage in such detailed listening—in part, they argue, because the music is overly simple and has insufficient substance to be cognitively engaging (see summaries of such negative appraisals in Fink, 2005, p. 19; Dauer, 2020, p. 24). Some music scholars have argued that minimalism's simplicity contains complexities upon analysis (Epstein, 1986; Cohn, 1992; Quinn, 2006). Beyond professionally trained listeners, do listeners tend to find the music engaging? If yes, do specific compositional details and techniques drive patterns of engagement?

Reich's *Piano Phase* (1967) offers a case study of how engagement and detailed listening might unfold. The piece, written for two pianos or marimbas, alternates between two distinct and highly repetitive states resulting from a single process. During in-phase sections, the two performers play a short musical unit in rhythmic unison, though varying in pitch alignment (**Figure 1**). In between these in-phase sections, one performer gradually accelerates, resulting in unpredictable note onsets (i.e., phasing sections). Over time these phasing sections lead to a new pitch alignment in the subsequent in-phase section.[1] The driving phasing process offers the listener an

---

[1]The piece begins with one pianist (Pianist 1) playing a twelve-note pattern consisting entirely of 16th notes and containing five unique pitches in the treble register. The pattern can be divided into two groups of six 16th notes, and Reich gave a metronome marking of 72 beats per minute to the dotted quarter note (one group of six 16th notes). The score consists of numbered modules that are repeated an indeterminate number of times: Reich noted approximate ranges for the number of repetitions above each module. After the pattern is established in the first module, the second pianist (Pianist 2) fades in, playing the identical pattern in unison with Pianist 1. After repeating the pattern in unison for some time, Pianist 2 accelerates very slightly while Pianist 1 holds the opening tempo, causing the sound from the two pianos to "wobble" out of sync to varying degrees as the pattern is repeated at different tempos (we call these portions *phasing* sections). Various and unpredictable rhythm and pitch events emerge and disappear in these phasing sections. Eventually Pianist 2's acceleration process culminates in another unison module where each pianist's 16th notes are once again realigned (which we label *in-phase* sections). Although the pianists' rhythms are realigned, the pitch content of the pattern will have shifted: In this example, Pianist 2 aligns the second pitch of the opening pattern with the first pitch of the pattern (played by Pianist 1). *Piano Phase* proceeds by alternating between phasing and in-phase sections, where each successive in-phase section presents the next shifted alignment of the opening, twelve-note pattern (note three aligns with the first note of the pattern, a phasing section occurs, then note four aligns with the first note of the pattern, etc.).

outline of how the piece unfolds at a macro-level while leaving many details unpredictable—for example, rhythms during the phasing sections and accent patterns during in-phase sections. For a listener interested in detailed minutia and slight variation, the work may fascinate; in other moods or with other listening priorities, the piece can bore, confuse, and even anger (Rockwell, 1973). With such a plethora of responses (Dauer, 2020), we aimed this initial study at better understanding engagement in general, operationalized for participants as "being compelled, drawn in, connected to what is happening, and interested in what will happen next" (Schubert et al., 2013).

Recent research using the high temporal resolution of electroencephalography (EEG) has suggested that the correlation of neural responses among participants (inter-subject correlation, or ISC) in response to natural audio-visual stimuli objectively indexes a state of "engagement." Foundational ISC work using fMRI has highlighted across-participant synchronization of neural responses to natural stimuli such as film excerpts (Hasson et al., 2004) and spoken narratives (Simony et al., 2016), and has uncovered relationships between neural synchronization and stimulus characteristics such as emotional arousal (Hasson et al., 2004) and narrative coherence (Lerner et al., 2011). fMRI ISC has also been used to study music processing: Abrams et al. (2013) reported greater synchronization when hearing intact music compared to temporally or spectrally manipulated controls, while Farbood et al. (2015) related hierarchical structural coherence of music to hierarchical neural processing. ISC for EEG was introduced in a film-viewing study by Dmochowski et al. (2012), who found that neural correlation was higher in response to film excerpts containing intact (vs. temporally scrambled) narratives, and peaked during periods of high tension and suspense—leading the authors to frame EEG-ISC as a measure of *engagement*, which they defined as "emotionally laden attention." Dmochowski et al. (2012) note that the brain state of engagement "lacks a rigorous definition" yet can be "readily describe[d] subjectively," and that it implies not only a state of attention, but an attentive state that "entails emotional involvement." The engagement interpretation of EEG ISC in the context of audiovisual processing was further investigated by Dmochowski et al. (2014), who found ISC of an experimental sample to reflect "engagement or interest of a large population" in television viewing. EEG ISC has subsequently been shown to index attentional state (Ki et al., 2016) and to predict memory retention (Cohen and Parra, 2016) and test scores (Cohen et al., 2018).

Ensuing studies have demonstrated how EEG ISC may be a powerful tool for analyzing music listening (Madsen et al., 2019; Kaneshiro et al., 2020, 2021). Madsen et al. (2019) drew on instrumental compositions (19 Western classical musical works in a variety of styles, and one Chinese folk song) to establish that ISCs decrease over repeated exposures to familiar music (though ISCs were sustained for participants with musical training). Kaneshiro et al. (2020) presented popular, Hindi-language songs from "Bollywood" films to participants and reported higher behavioral ratings and ISCs for their original versions when compared with phase-scrambled manipulations. Most recently, Kaneshiro et al. (2021) investigated participants' time-resolved

FIGURE 1 | The opening modules from Steve Reich's *Piano Phase*. Lines under the staff indicate sections: blue lines are in-phase sections and red lines are phasing sections.

ISCs in response to the first movement of Edward Elgar's Cello Concerto in E minor, Op. 85. In contrast to the stimuli used in these previous studies, and true to minimalism's stereotypical characteristics, Reich's *Piano Phase* features a high level of repetition, unchanging timbre, and narrow pitch content.[2]

Other researchers have used minimalist compositions as experimental stimuli, similarly taking advantage of the works' unusual musical properties. Musicologist Keith Potter and computer science colleagues used two early works by Philip Glass to compare information dynamics and musical structure (Potter et al., 2007). Psychologist Michael Schutz worked with percussionist Russell Hartenberger to examine desynchronization among performers of Reich's *Drumming* (Hartenberger, 2016),[3] and Daniel Cameron and colleagues have studied experiences of groove and neural entrainment using Reich's *Clapping Music* (Cameron et al., 2017, 2019). Dauer et al. (2020) examined preattentive cortical responses to various types of formal repetition using synthesized melodies based on early minimalist compositional techniques. The current study takes minimalism as an edge case in the applicability of neural correlation, uniting the repertoire's extreme musical techniques (and unique reception history) with multivariate techniques for analyzing brain data. While we focus on phasing as one important type of musical repetition, we anticipate that some aspects of the results may meaningfully generalize to other repetitive repertoires such as music used to accompany trancing (Becker, 2004). Future work could interrogate such generalizations.

Our primary research question was to uncover whether participants shared engagement patterns (as measured by ISC) while listening to *Piano Phase*. In particular, we hypothesized that phasing sections (sections where one pianist is changing tempo) would be more collectively engaging (i.e., elicit more correlated responses) than in-phase sections, due to phasing

sections' rhythmic variety, rhythmic unpredictability, and a wider variety of pitch interactions (see above and **Figure 1** for musical details about *Piano Phase*). If listeners deployed the hyper-detailed listening strategy described above, phasing sections would offer rich content with which to engage. On the other hand, detailed listening during phasing sections could lead to divergent engagement between listeners as they lock on to different aspects of the music during these more eventful sections. Since ISC depends on time-locked similarities in neural data, these divergent but equally engaged listening experiences may result in lower correlations than in-phase sections. Using ISC as a way to index collective engagement, we explored whether phasing sections contribute to ISC by introducing a manipulation of *Piano Phase* without phasing sections (which we called Abrupt Change). We anticipated that ISC would be lower for this manipulation if phasing sections contributed to ISC in the original version. We also examined whether the gradual nature of the phasing process in *Piano Phase* might be critical for engagement. To this end, we included a manipulation of *Piano Phase* with frequent and random changes in the content (Segment Shuffle). By reshuffling 5-s segments of the original excerpt, we rendered unrecognizable the gradual phasing process and the alternations between in-phase and phasing sections. If the phasing process meaningfully contributes to engagement, we expected lower ISC values for the shuffled version as it lacked gradual phasing. To examine the possibility of listeners being bored or disengaged by the original work, we also introduced a third control stimulus with more extreme repetition that should be less engaging than the original work (Tremolo). Finally, we included a commercial remix of Reich's original work in a popular style (Remix), which we conjectured would reliably engage listeners and elicit EEG ISC comparably to previous experiments with popular music stimuli (Kaneshiro, 2016; Kaneshiro et al., 2020). Remix also provided a stylistic contrast with *Piano Phase*: we expected that the remix would engage listeners more than *Piano Phase* because the remix has more attention-catching musical events. In sum, if the core musical features of *Piano*

---

[2]We note that Madsen et al. (2019) did include Philip Glass's *String Quartet No. 5* (1991): a more popular or "post-minimalist" work by comparison.
[3]https://maplelab.net/reich/

Phase drive engagement, we hypothesized that the manipulated versions (Abrupt Change, Segment Shuffle, and Tremolo) would elicit lower ISC. We expected ISC in response to Remix to be comparable with values found in previous popular-music pieces (Kaneshiro et al., 2020).

In line with recent work, we computed ISCs over entire excerpts and in shorter, overlapping time windows, giving us a sense of overall engagement as well as moment-to-moment patterns shared between audience members (Dmochowski et al., 2012; Kaneshiro et al., 2021). To provide complementary measures of what ISC is reliably indexing, participants rated the stimuli and additionally completed a second experimental block where they continuously reported their level of engagement with the stimuli. This allowed us to compare relationships for both overall and time-resolved neural and behavioral measures. These continuous EEG and behavioral measures allowed us to examine our expectations at a more granular level: We expected significant ISC during phasing sections of the original version and at the onset of new phasing sections in Abrupt Change, scarce ISC for Segment Shuffle, less for Tremolo, and frequent ISC in response to the dramatic musical events in Remix.

2. METHODS

2.1. Stimuli

All five stimuli in the experiment are related to Steve Reich's *Piano Phase*, a much-anthologized example of American minimalism for two pianos or marimbas (**Figure 1**). In the experiment we used pianists Nurit Tilles and Edmund Neimann's 1987 recording on the album *Reich "Early Works"* released by Double Edge (Reich, 1987). The performers take an appropriate tempo (see footnote 1), use detached articulation, and create an overall energetic feel. We used the first 5 min and 5 s (5:05) of the track's 20:26 duration. We refer to this excerpt of *Piano Phase* used in the experiment as the *Original* condition (**Figure 2A**).[4]

Piano Phase offers contrasting sections (phasing and in-phase) with slightly varying musical content for comparison while holding many other musical parameters constant: timbre, dynamics (largely), instrumentation, pitch content, and absence of lyrics or vocal content. These features make it uncommonly amenable to the creation of the stimulus manipulations used in this study.

Using MATLAB software, we created three additional stimulus conditions of equal duration, each based on the content of the excerpt used in the Original condition. First, in the *Abrupt Change* condition, (**Figure 2B**) all phasing sections from the Original excerpt were replaced with exact repetitions of the preceding in-phase material. The stimulus thus presents repetitions of an in-phase motif through the section where the phasing would have occurred, and then shifts abruptly to the next in-phase section as closely as possible to its occurrence in the original recording. For example, the stimulus begins with the in-phase section where Pianist 1 and Pianist 2 align the first notes of the twelve-note pattern. This continues without phasing

---

[4]A meter shift and accompanying pattern change occur later in the piece, but after the excerpt used in the experiment.
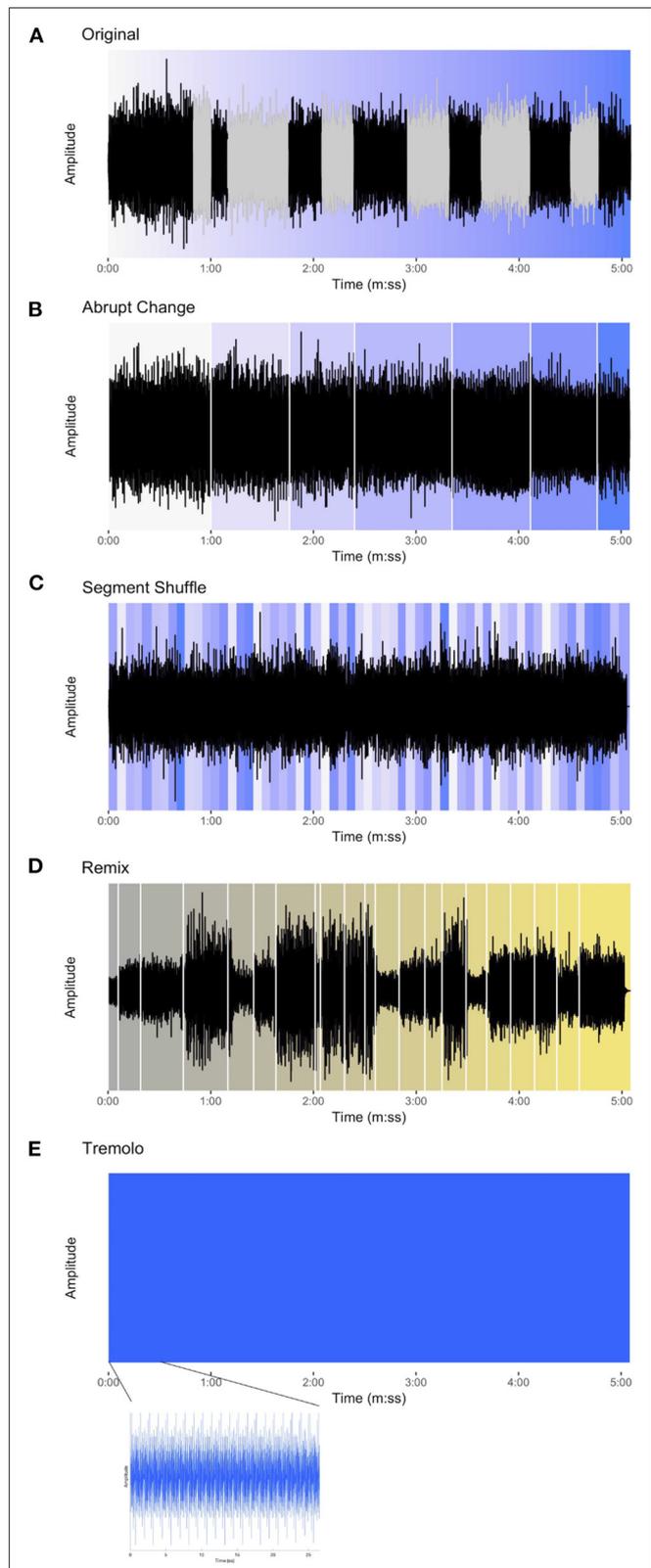

FIGURE 2 | The waveforms for each of the stimuli in the experiment. **(A)** Original, with phasing sections colored gray and the progression of events

(Continued)

FIGURE 2 | represented by the gradual change of color from white to blue. **(B)** Abrupt Change, white lines denoting sudden shift from one in-phase section to the next and background color showing approximate location of in-phase material in the Original condition. **(C)** Segment Shuffle, random re-ordering of 5-s units shown using original color in Original. **(D)** Remix [Winn's *Piano Phase (D*Note's Phased & Konfused Mix)*], gradual progression of events represented with color change from gray to yellow and key musical events beginning with white lines. **(E)** Tremolo, appearing as an unchanging block when zoomed out, but in the lower plot, zoomed in to show the reiterated pitch material.

until suddenly the next in-phase section emerges, where Pianist 2 aligns the second note of the pattern with the first note of the pattern played by Pianist 1. Thus, the Abrupt Change condition is, in essence, form without function: where regular markers of formal sections (i.e., points of arrival at the alignments of in-phase sections) are situated without the functional transitions (i.e., the phasing sections).

As a contrast to the sudden changes embodied by the Abrupt Change condition, we created the *Segment Shuffle* condition (**Figure 2C**). Here we divided the Original audio into 5-s segments and randomly reordered them (i.e., "shuffled" them). In order to avoid abrupt disjunct shifts, the transitions between segments were smoothed by applying a linear crossfade. The 5-s segments included both phasing and in-phase material, meaning that upcoming content was unpredictable for listeners. In contrast with the Abrupt Change condition, Segment Shuffle featured function without form: constant, potentially surprising changes with no overarching formal scheme.

Finally, we synthesized a stimulus with neither form nor function, taking the repetition aspect of minimalist music to an extreme. Our *Tremolo* condition (**Figure 2E**) consisted solely of the aggregated pitch content of *Piano Phase* presented as a block chord, reiterated at Reich's opening tempo marking and lasting the duration of the Original excerpt.

For comparison with the more popular genres of audio materials used in previous ISC studies, we also included Matt Winn's Piano Phase *(D*Note's Phased & Konfused Mix)*, an homage to Reich's piece released on the 1999 *Reich Remixed* album (Reich, 1999); we refer to this condition as *Remix* for short (**Figure 2D**). Winn's dance music group, D*Note, draws on sounds from electronica and jazz, and these influences show up in Remix alongside samples from Reich's piece.[5] The entire track was used in the experiment and its duration (5:05) informed the length of the other stimuli. Listening to Remix, we identified moments (musical events) that we predicted would engage listeners (for a full list, see **Supplementary Table S1**). These events guided our interpretation of time-resolved EEG and continuous behavioral (CB) results.

All stimuli were presented to participants as mono .wav files; the second audio channel was embedded with intermittent square-wave pulses which were used as precise timing triggers (see § 2.3 and Kaneshiro et al., 2020).

---

[5]https://www.mattwinn.co.uk/about

2.2. Participants

We were interested in listeners' initial experiences of Reich's piece and sought participants who were unlikely to have heard the composition before. Participants had to be 18–35 years old, have normal hearing, be right-handed, have no cognitive or decisional impairments, be fluent in English, and have had no individual musical instrument or vocal training, nor musical education after high school (or equivalent).

The participant sample ($N = 30$; 19 female, 11 male) had a mean age of 23.8 years (ranging from 18 to 35 years). Twelve participants reported some formal musical training ranging from 2 to 16 years (average of 4.5 years) including activities such as elementary school band and orchestra and piano lessons in middle school. Only two participants reported ongoing musical activities (one was an amateur ukulele player and another noted participating in occasional jam sessions). All participants reported listening to music regularly, from 0.2 to 8 h a day (average of 2.4 h per day).

2.3. Experimental Paradigm and Data Acquisition

The Stanford University Institutional Review Board approved this research, and all participants gave written informed consent before completing the experiment. After discussing and signing the consent form, each participant completed questionnaires about demographic information and musical experience. Each participant then completed two blocks: one EEG (Block 1) and one behavioral (Block 2), both conducted in an acoustically and electrically shielded ETS-Lindgren booth (**Figure 3**). The participant completed a brief training session to acquaint them with the interface and task before the experimenter donned the EEG net. The participant was told to sit comfortably in front of the monitor and view a fixation image while EEG was recorded. Participants listened to each of the five stimuli once in random order with their eyes open. Participants did not perform any task during the presentation of the stimuli and were told to refrain from moving their body in response to the music: they were told not to tap their feet or hands, or bob their heads. After each stimulus in Block 1, the participant rated how pleasant, well ordered, musical, and interesting the preceding stimulus was on a scale of 1 (not at all) to 9 (very) via key press using a computer keyboard. Participants were permitted to move and take short breaks in between stimuli (during which time a "break" screen appeared). When ready, the participant initiated the next stimulus by pressing the space bar on the keyboard.

The EEG net was removed after Block 1, and the participant returned to the sound booth to complete Block 2. Here the participant heard the same five stimuli (in random order) and this time completed a continuous behavioral task while listening. Their task was to continuously report their level of engagement—which was defined as "being compelled, drawn in, connected to what is happening, and interested in what will happen next" (Schubert et al., 2013)—over the duration of each stimulus. We consider this definition to be aligned with the original definition in EEG-ISC research of "emotionally laden attention" (Dmochowski et al., 2012), while also providing participants a

clearer, more elaborated way of understanding engagement in order to perform the task. This more detailed definition has also been used in previous studies involving continuous reporting of engagement in response to dance (Schubert et al., 2013) and music (Olsen et al., 2014). The definition of engagement was provided only in the second, behavioral block and not in the EEG block as participants in prior EEG-ISC studies of engagement were not informed that their neural responses would be related to this state (Dmochowski et al., 2012, 2014; Madsen et al., 2019; Kaneshiro et al., 2020, 2021).

To perform this task, the participant used a computer mouse to control a slider shown on the computer monitor. The screen displaying the slider contained the prompt "Rate your level of engagement as the excerpt plays," and the endpoints of the slider were labeled "Not at all" and "Very engaged," corresponding to continuous scale values of 0 and 100, respectively. The slider was positioned at the bottom of its range (0 value) at the start of each trial. After each stimulus, the participant rated how engaging they found the preceding stimulus to be overall, using the same 1–9 key press scale used in Block 1. The ordering of blocks was not randomized (i.e., the EEG block always preceded the CB block) because we wanted to ensure that during recording of EEG data in Block 1, participants would not be biased with the definition of engagement and the continuous reporting task that came in Block 2.

The experiment was programmed in MATLAB using the Psychophysics Toolbox (Brainard, 1997). Stimuli were played through two Genelec 1030A speakers located 120 cm from the participant. The stimuli were scaled to a common loudness level based on perceptual assessments from three researchers, all of whom were trained musicians. Stimulus onsets were precisely timed by sending square-wave pulses to the EEG amplifier from a second audio channel (not heard by the participant). We used the Electrical Geodesics, Inc., (EGI) GES 300 platform (Tucker, 1993), a Net Amps 300 amplifier, and 128-channel electrode nets to acquire EEG data with a 1 kHz sampling rate and Cz vertex reference. Before beginning the EEG block, we verified that electrode impedances were below 60 kΩ (Ferree et al., 2001). In the CB block, data were acquired at a sampling rate of 20 Hz.

2.4. Data Preprocessing
Continuous EEG recordings were preprocessed offline in MATLAB after export using Net Station software. The data preprocessing procedure used here is described in detail in Kaneshiro et al. (2021), which itself was adapted from the preprocessing procedure of Kaneshiro et al. (2020). Briefly, data were preprocessed on a per-recording basis: Each recording was highpass (above 0.3 Hz), notch (between 59 and 61 Hz) and lowpass (below 50 Hz) zero-phase filtered before being downsampled from 1 kHz to 125 Hz. Epochs for each stimulus were 5 min (5:00; 37501 time samples) in length; we used a slightly shorter analysis epoch than the length of the stimuli (excluding the last 5 s) because the Remix stimulus included 5 s of silence at the end. Stimulus onsets were precisely timed from the audio pulses. Ocular and EKG artifacts were removed using ICA (Jung et al., 1998): Components whose activity reflected ocular activity (identified according to the procedure described

in Kaneshiro et al., 2020) or EKG artifacts (identified via visual inspection of projected activity in the first 30 components) were zeroed out before projecting data back to electrode space. Finally, data were converted to average reference, and data from bad electrodes or noisy transients were replaced with a spatial average of data from neighboring electrodes. After preprocessing, each trial of data was a 2D electrode-by-time matrix ($125 \times 37, 501$). The matrices contained data from 125 electrodes as we excluded the four sensors over the face (electrodes 125–128) and reconstituted the reference sensor during preprocessing (Kaneshiro, 2016; Losorelli et al., 2017; Kaneshiro et al., 2020, 2021). During preprocessing, participant S08's response to the Tremolo stimulus was flagged as containing excessive noise artifacts; therefore we excluded this trial from further analysis, but retained other trials from this participant.

After preprocessing, we aggregated trials into 3D electrode-by-time-by-participant data matrices for each stimulus. As a result, responses to Original, Abrupt Change, Segment Shuffle, and Remix stimuli were stored in $125 \times 37, 501 \times 30$ matrices, while responses to Tremolo were stored in a $125 \times 37, 501 \times 29$ matrix.
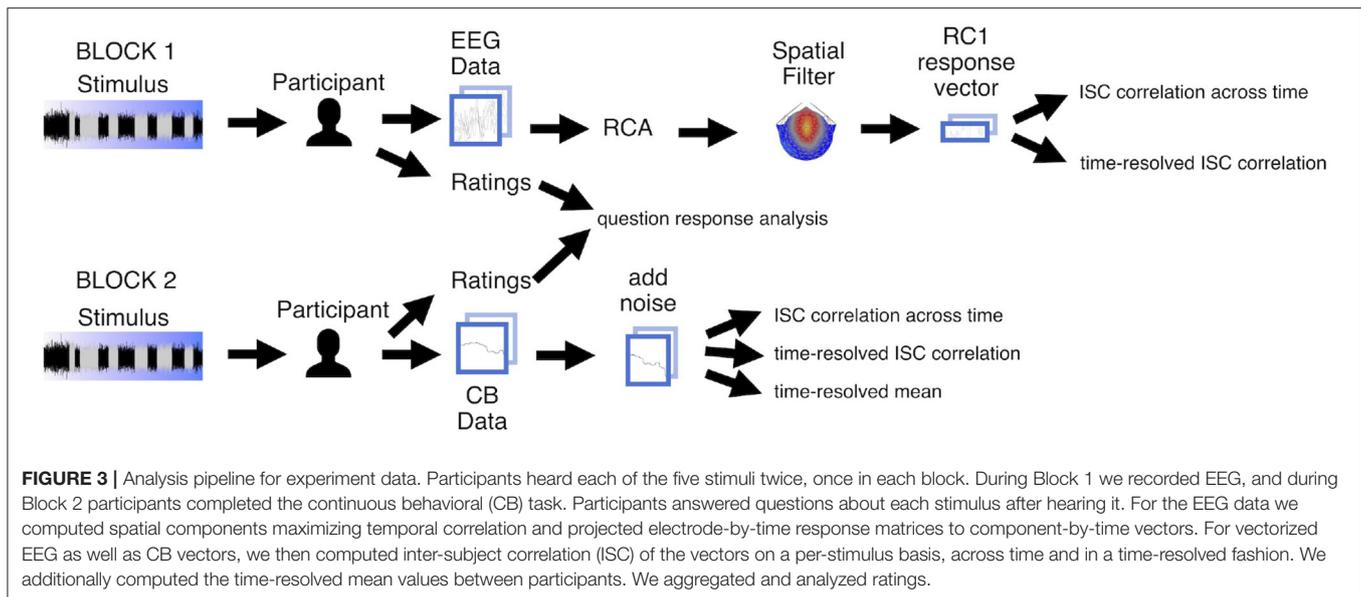
CB data were similarly segmented into 5-min (5:00) epochs and aggregated into a single time-by-participant-by-stimulus matrix. As the data remained at the acquisition sampling rate of 20 Hz for analysis, the matrix was of size $6,000 \times 30 \times 5$. Behavioral ratings from the EEG and CB blocks were aggregated into a single .csv file for statistical analyses.

2.5. Data Analysis
Figure 3 summarizes our analysis pipeline for the EEG and CB data. EEG was recorded from participants in Block 1, and participants provided CB reports of engagement in Block 2. Participants also rated the stimuli in both blocks. We computed ISC of both the EEG and CB measures, and also computed mean CB across participants. Finally, we analyzed the ratings to determine whether they differed significantly according to stimulus.

2.5.1. Spatial Filtering of EEG Data
Previous EEG ISC studies have prepended a spatial filtering operation before computing correlations in order to maximize signal-to-noise ratio of the data while also reducing the dimensionality of each EEG trial from a space-by-time matrix to time vectors from one or a few components (Dmochowski et al., 2012). Therefore, we filtered the EEG data using Reliable Components Analysis (RCA) prior to computing ISC (Dmochowski et al., 2012, 2015). RCA maximizes across-trials covariance of EEG responses to a shared stimulus relative to within-trials covariance, and therefore maximizes correlated activity across trials (i.e., ISC). It is similar to PCA, but maximizes correlation across trials as opposed to variance explained in a single response matrix. Like PCA, RCA involves an eigenvalue decomposition of the data, returning multiple spatial filters as eigenvectors and corresponding coefficients as eigenvalues (Dmochowski et al., 2012). The components are returned in descending order of reliability explained; in other words, the first component RC1 is that in which ISC of component-space data

FIGURE 3 | Analysis pipeline for experiment data. Participants heard each of the five stimuli twice, once in each block. During Block 1 we recorded EEG, and during Block 2 participants completed the continuous behavioral (CB) task. Participants answered questions about each stimulus after hearing it. For the EEG data we computed spatial components maximizing temporal correlation and projected electrode-by-time response matrices to component-by-time vectors. For vectorized EEG as well as CB vectors, we then computed inter-subject correlation (ISC) of the vectors on a per-stimulus basis, across time and in a time-resolved fashion. We additionally computed the time-resolved mean values between participants. We aggregated and analyzed ratings.

is maximized, followed by RC2, RC3, etc. We use the RC1 of Tremolo in what follows but note two important limitations: no RCs for Tremolo are statistically significant, and the topography of RC1 is qualitatively different from the first RCs for the other four stimuli (**Figure 5A** and **Supplementary Figure S1**).

We used a publicly available MATLAB implementation (Dmochowski et al., 2015), computing RCA separately for each stimulus. Following Kaneshiro et al. (2020), we computed the top five reliable components (RCs). We observed a sharp drop in RC coefficients after the first, most-correlated component (RC1); given that past research has reported negligible ISC in subsequent RCs in this scenario (Kaneshiro et al., 2021), we proceeded with ISC analyses using RC1 data only, as was done by Kaneshiro et al. (2020). In presenting the forward-model projections of component weights as scalp topographies (Parra et al., 2005), each weight vector was first multiplied by ±1 such that frontal electrodes were associated with positive weightings; this was for visualization only, and polarity of the projected data does not impact computed correlations.

2.5.2. Inter-subject Correlation Analyses

In computing the EEG ISC of RC1 response vectors, we first computed ISC across the entire duration of each stimulus (Kaneshiro et al., 2020, 2021). Following this, we computed ISC in a time-resolved fashion. Following past research (Dmochowski et al., 2012; Poulsen et al., 2017; Kaneshiro et al., 2021), we used a 5-s window with a 1-s shift between windows. These parameters provide an adequate number of data points (625 EEG samples, 100 CB samples) to produce sufficient signal to noise to measure correlation, while this window length in conjunction with the 1-s hop size leads to an 80% overlap between windows, which smooths the resulting time series and facilitates interpretation. These window length and window shift parameters produced a total of 296 time-resolved ISC points across each stimulus with a temporal resolution of 1 s. ISC for each participant was

computed in a one-against-all fashion (the correlation of each participant's RC1 response vector with every other participant's response vector for a given stimulus). We report the mean ISC across participants and additionally visualize single-participant correlations for all-time ISC and standard error of the mean for time-resolved ISC.

For the CB responses, we computed mean CB at each time sample, as well as CB ISC both across entire excerpts and in the short time windows described above. CB responses were already in vector form for each participant, so we did not perform any operation akin to EEG spatial filtering before computing means and ISC. At times, individual participants did not move the slider in a given 5-s window, which produced missing values when computing correlations. To address this issue, for the CB ISC analyses *only* we added a small amount of noise, uniformly distributed over the interval ±0.001, independently to each CB response vector prior to computing ISC. As with the EEG data, we report means and single-participant values for analyses across entire stimuli, and means with standard error of the mean for time-resolved measures.

2.5.3. Statistical Analyses

Significance of each EEG result was computed using permutation testing. As described in detail in previous studies (Kaneshiro et al., 2020, 2021), we conducted each EEG analysis 1,000 times; in each iteration, the phase spectrum of each EEG trial input to RCA had been randomized (Prichard and Theiler, 1994). The distribution of 1,000 outcomes for each analysis then served as the null distribution for assessing significance of the observed result. We performed a similar procedure to create null distributions for CB ISC, independently phase scrambling each CB response vector prior to computing ISC—also over 1,000 iterations. We compare each observed ISC result to the corresponding null distribution in order to compute p-values, and as effect size d we report the number of standard deviations

between the observed result and the expected result under the null distribution (Nakagawa and Cuthill, 2007).

Behavioral ratings, EEG ISC computed over entire stimuli, and CB ISC computed over entire stimuli were each analyzed using R (Ihaka and Gentleman, 1996; R Core Team, 2019) and the lme4 package (Bates et al., 2012). We performed a linear mixed-effects analysis of the relationship between response values and stimulus conditions, with fixed effect of condition (Original, Abrupt Change, Segment Shuffle, Remix, and Tremolo) and random effect of participant in each model. We then tested each model against a null model without the fixed effect of condition using the anova function in lme4. This produced a chi-squared statistic and associated p-value (Winter, 2013). As in Kaneshiro et al. (2020), ordinal behavioral ratings were treated as approximately continuous (Norman, 2010). Following this we conducted two-tailed pairwise t-tests to assess differences between pairs of stimulus conditions. Effect size (Cohen's D) was also calculated and reported.

Results for analyses involving multiple comparisons were corrected using False Discovery Rate (FDR, Benjamini and Yekutieli, 2001). For discrete results, we corrected for multiple comparisons on a per-stimulus basis (EEG ISC and CB ISC data: 10 paired comparisons over five stimulus conditions; behavioral ratings: 10 paired comparisons per stimulus; RC coefficients: five unpaired comparisons per stimulus). We performed no temporal cluster correction on the time-resolved ISC: as noted by Kaneshiro et al. (2021), temporal dependence was accounted for in the phase-scrambling procedure underlying the permutation testing, which preserves autocorrelation characteristics of the original response data (Prichard and Theiler, 1994; Lancaster et al., 2018).

3. RESULTS

In order to examine engagement with an example of musical minimalism, we used inter-subject correlation (ISC) to analyze EEG and continuous behavioral (CB) responses from 30 adult participants who heard an intact excerpt of Steve Reich's *Piano Phase*, three manipulated control stimuli, and a professional remix of Reich's piece. We analyzed EEG and CB ISC in two ways: an aggregate ISC value for each stimulus (overall EEG ISC, overall CB ISC) and time-resolved ISCs for both EEG and CB data. Each participant also gave ordinal ratings of each stimulus (behavioral ratings).

3.1. Remix Stimulus Garnered Highest Behavioral Ratings

After hearing each stimulus in Block 1, participants used a 1–9 scale to rate how pleasant, musical, well ordered, and interesting they found each excerpt. Later, in Block 2, they used the same scale to report their overall level of engagement with each stimulus. Ratings for all five questions were found to differ significantly by condition (**Figure 4**): pleasant [$\chi^2_{(4)} = 126.03$, $p < 0.001$], musical [$\chi^2_{(4)} = 139.78$, $p < 0.001$], well ordered [$\chi^2_{(4)} = 37.996$, $p < 0.001$], interesting [$\chi^2_{(4)} = 104.29$, $p < 0.001$], and engaging [$\chi^2_{(4)} = 127.92$, $p < 0.001$].
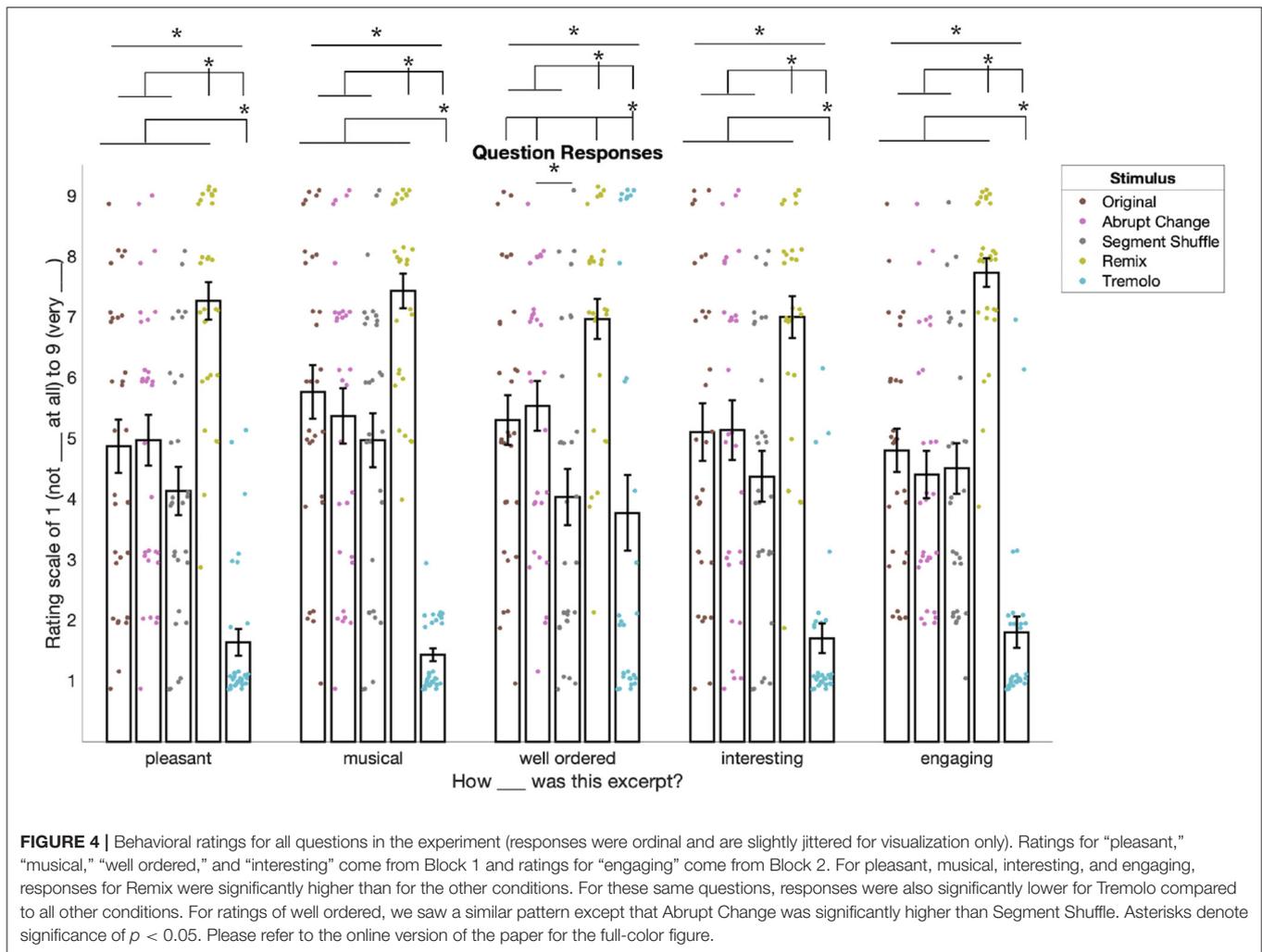
Follow-up pairwise t-tests comparing responses between conditions showed a similar pattern for four of the five questions (see **Supplementary Tables S2–S6** for all p-values and d-values). For pleasant, musical, interesting, and engaging ratings, responses to Remix were significantly higher than to the other four conditions ($p_{FDR} < 0.01$, 10 comparisons) and responses to Tremolo were significantly lower than the other four conditions ($p_{FDR} < 0.01$). However, these ratings did not differ significantly between Original, Abrupt Change, and Segment Shuffle conditions ($p_{FDR} > 0.05$).

Ratings for how "well ordered" the stimuli were followed a slightly different pattern. While Remix was rated significantly higher than all other conditions (see **Supplementary Table S4**), Tremolo was rated significantly lower than all other conditions except Segment Shuffle ($p_{FDR} = 0.719$, $d = 0.065$). In addition, Segment Shuffle was rated significantly lower than Abrupt Change ($p_{FDR} = 0.036$, $d = 0.543$).

3.2. Overall EEG ISC Is Highest for Remix, Lowest for Tremolo

In computing the EEG ISCs, we first spatially filtered the responses for each stimulus in order to reduce their dimensionality from 125 electrodes to a single, maximally correlated spatial component (RC1) for each stimulus. These components are shown in **Figure 5A**. For all but the Tremolo, RC1 was maximally weighted over the fronto-central region. While our spatial filtering technique returned multiple components, we focus only on the first component because it is the only component with statistically significant coefficients for the majority of stimuli: **Figure 5B** demonstrates that RC1 was the only significant component for most stimuli (permutation testing; Original, Abrupt Change, Segment Shuffle, Remix $p_{FDR} < 0.001$; Tremolo $p_{FDR} = 0.379$; see **Supplementary Table S7** for all p-values). Remix also had a significant RC4 and Tremolo had no significant RCs. The topographies and coefficient significance for RC1 are in line with those computed in previous music EEG ISC studies (Kaneshiro et al., 2020, 2021); given that subsequent RCs did not correspond to significant ISC in a closely related study with similar distributions of coefficients (Kaneshiro et al., 2021), here we compute ISC only for RC1.

When computed over the entire duration of a stimulus, EEG ISC was statistically significant in response to Original (permutation test, $p < 0.001$, $d = 2.4$), Abrupt Change ($p < 0.001$, $d = 3.1$), Segment Shuffle ($p < 0.001$, $d = 2.7$), and Remix ($p < 0.001$, $d = 5.1$), but not Tremolo ($p = 0.41$, $d = 0.7$). ISC also differed significantly by condition [$\chi^2_{(4)} = 96.002$, $p < 0.001$]; follow-up pairwise comparisons indicated that Original, Abrupt Change, Segment Shuffle, and Tremolo all significantly differed from Remix ($p_{FDR} < 0.001$), and Original, Abrupt Change, Segment Shuffle, and Remix all differed from Tremolo ($p_{FDR} < 0.001$). **Figure 5C** shows the direction of these significant differences: Remix garnered higher overall EEG ISC values than the other conditions, while Tremolo received the lowest overall values. Despite their structural differences, ISC among Original, Abrupt Change, and Segment Shuffle did not

FIGURE 4 | Behavioral ratings for all questions in the experiment (responses were ordinal and are slightly jittered for visualization only). Ratings for "pleasant," "musical," "well ordered," and "interesting" come from Block 1 and ratings for "engaging" come from Block 2. For pleasant, musical, interesting, and engaging, responses for Remix were significantly higher than for the other conditions. For these same questions, responses were also significantly lower for Tremolo compared to all other conditions. For ratings of well ordered, we saw a similar pattern except that Abrupt Change was significantly higher than Segment Shuffle. Asterisks denote significance of $p < 0.05$. Please refer to the online version of the paper for the full-color figure.

differ significantly from one another when computed over entire excerpts (see **Supplementary Table S8** for a full list of p-values and d-values).
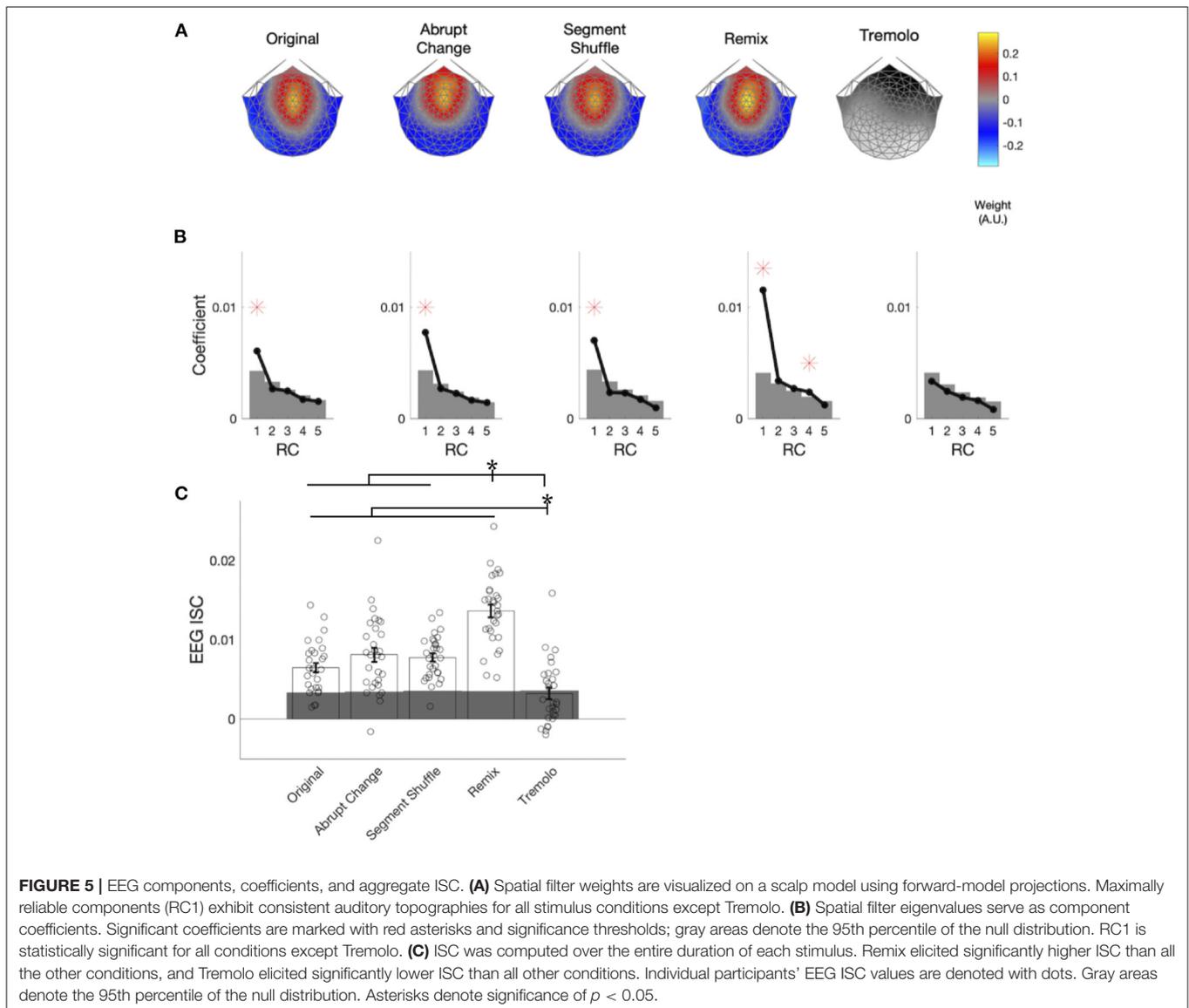
3.3. Overall CB ISC Aligns Broadly With EEG ISC

To analyze the CB ISC values (**Figure 6**), we followed the same procedures used for comparing EEG ISC computed over entire stimuli. Statistically significant CB ISC was observed in responses to Original (permutation test, $p < 0.001$, $d = 6.5$), Abrupt Change ($p < 0.001$, $d = 6.6$), Segment Shuffle ($p < 0.001$, $d = 12.7$), and Remix ($p < 0.001$, $d = 34.7$) stimuli, but not Tremolo ($p = 0.22$, $d = 0.6$). CB ISC significantly differed by condition [$\chi^2_{(4)} = 180.2$, $p < 0.001$]. Pairwise comparisons revealed that Remix had higher ISC than all other conditions, Tremolo had lower ISC than all other conditions, and Segment Shuffle had higher ISC than all conditions except Remix. All condition comparisons were significant except for Original vs. Abrupt Change ($p_{FDR} = 0.87$, $d = 0.06$; all other comparisons, $p_{FDR} < 0.05$, see **Supplementary Table S9** for a full list). Cross-correlations between time-resolved EEG ISC and

CB ISC showed a similar pattern (maximum normalized r-values with a maximum lag of 10 s were: Original: 0.29, Abrupt Change: 0.29, Segment Shuffle: 0.37, Remix: 0.70, and Tremolo: 0.18).

3.4. Time-Resolved Measures Coincide With Musical Events

In addition to calculating the overall ISC for EEG and CB data, we were also interested in observing changes in ISC over the course of the stimuli (see **Supplementary Figure S2** for individual, time-resolved CB responses underlying CB ISC). After computing ISC over short, shifting time windows, we visualized the ISC trajectory over time. Permutation testing provided a time-varying statistical significance threshold, allowing us to see when participants, as a group, had significantly correlated responses. Below we give a qualitative assessment of these results (**Figure 7**). Note that although EEG and CB ISC data had different sampling rates, we used identical time window lengths (5 s) and shifts (1 s) to facilitate comparison. We plot time-resolved ISC at the center of each temporal window. This means significant ISC implicates activity from ±2.5 s around each time point. Because all participants experienced the EEG block first and the CB block

FIGURE 5 | EEG components, coefficients, and aggregate ISC. **(A)** Spatial filter weights are visualized on a scalp model using forward-model projections. Maximally reliable components (RC1) exhibit consistent auditory topographies for all stimulus conditions except Tremolo. **(B)** Spatial filter eigenvalues serve as component coefficients. Significant coefficients are marked with red asterisks and significance thresholds; gray areas denote the 95th percentile of the null distribution. RC1 is statistically significant for all conditions except Tremolo. **(C)** ISC was computed over the entire duration of each stimulus. Remix elicited significantly higher ISC than all the other conditions, and Tremolo elicited significantly lower ISC than all other conditions. Individual participants' EEG ISC values are denoted with dots. Gray areas denote the 95th percentile of the null distribution. Asterisks denote significance of $p < 0.05$.
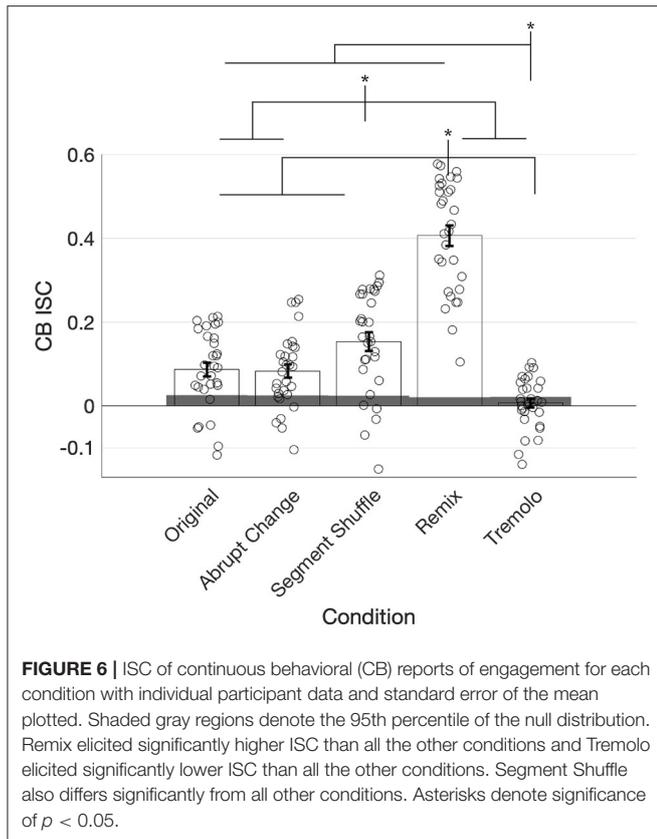
second, differences between the two could be due to repeated exposure (Madsen et al., 2019). In addition, although there is precedent in fMRI ISC research to remove the initial seconds of participants' responses to avoid including an onset response (Wilson et al., 2008), we decided to include the responses for the entire stimulus duration because even these early responses varied by stimulus condition.

Responses to the Original stimulus show small but significant ISC peaks in the EEG data (permutation test $p < 0.05$, uncorrected, see Methods; time-varying effect sizes are given in **Supplementary Figures S3**, **S4**), with statistically significant ISC in 16.9% of the time windows (**Table 1**). The largest ISC peaks appear around the approximate start times of phasing sections, or shortly thereafter. Each of the phasing section onsets (marked in **Figure 7A** with dotted lines) is accompanied by a significant peak with the exception of the third phasing section (which may

have a perceptually smoother transition than the other phasing sections). While phasing elicits ISC peaks relatively consistently, in-phase sections fail to correspond to any significant ISC peaks. Both EEG and CB ISC also contain a significant peak at the start of the excerpt. In the time-resolved CB ISC data, only a handful of small peaks occur above the significance threshold after the initial drop; they seem unrelated to phasing and in-phase musical events (peaks one and four fall in-phase sections, peaks two, three, and five fall in phasing sections), and only 4.7% of the ISC values are significant (**Table 1**). In contrast with phasing sections eliciting consistent peaks in the EEG ISC data, the CB mean data shows an increase in mean engagement rating after the start of each in-phase section. There also appears to be a slight decrease across the length of the stimulus.

EEG ISC data for the Abrupt Change condition shows significant peaks within seconds of the in-phase shifts (shifts

FIGURE 6 | ISC of continuous behavioral (CB) reports of engagement for each condition with individual participant data and standard error of the mean plotted. Shaded gray regions denote the 95th percentile of the null distribution. Remix elicited significantly higher ISC than all the other conditions and Tremolo elicited significantly lower ISC than all the other conditions. Segment Shuffle also differs significantly from all other conditions. Asterisks denote significance of $p < 0.05$.

number two, three, five and six as marked in solid lines in **Figure 7B**; (18.6% of ISC values are significant; see **Table 1**).[6] In contrast with the Original condition, in the Abrupt Change condition, where in-phase sections begin suddenly, they seem to elicit ISC peaks in the EEG data. The other small significance peaks in the EEG data come between in-phase changes, perhaps as participants anticipate stimulus alterations during the long stretches of unchanging material (perhaps something like the hazard function between warning and imperative stimuli (Tecce, 1972; Nobre et al., 2007). After an initial descent, the CB ISC data shows significant peaks around the first two and final two in-phase changes (percentage of significant time-resolved CB ISCs = 7.0%; **Table 1**). The other two significant peaks appear between in-phase changes, perhaps related to the effect noted above. As in the Original condition, time-resolved CB mean data shows slight increases in engagement ratings after all six abrupt changes and an overall decline in engagement.

The perennially unpredictable changes in Segment Shuffle were met with frequent, small bursts of significant ISC correlations in the EEG data (**Figure 7C**; 15.9% significant ISC values; see **Table 1**). Comparing EEG and CB ISC time courses reveals unreliable alignment: After the initial drop in CB data, eight significant peak bursts unfold; about half of them align with EEG peaks (see peaks around time 1:30 and 3:05) while the other half do not (see peaks around time 0:15 and 2:30). CB means show small bumps in engagement ratings in the

---

[6]Smaller peaks are easiest to see in the online version of the paper.

midst of a long-term downward trend (percentage of significant time-resolved CB ISCs = 10.3%; see **Table 1**).

Time-resolved ISCs for the Remix condition give ample opportunity to correlate peaks with musical events, with statistically significant EEG ISC in 45.6% time windows and significant CB ISC in 25.9% of time windows (**Table 1**). We selected the coded events in **Figure 7D** based on moments in the work that we deemed most musically salient (see **Supplementary Table S1** for the timings and descriptions of all 19 events). Note that not all of these events aligned with ISC peaks, but here we discuss some that did. After a sample from *Piano Phase* is presented for the first few seconds of Remix, a dramatic drum machine attack builds into simultaneous entrances for a synth countermelody and marimba riff (0:06). This build up and entrance align with the first and largest peak in the EEG data. The second peak in the EEG data comes at what might be the most dramatic moment in the piece, a beat drop anticipated with a drum machine lick (0:44). Note the potentially related peak in the CB ISC data following this event. But ISC peaks are not always elicited in both EEG and CB data. For example, the neighboring musical moments around minute 2:00 arise from a sudden dropping out of the percussion for a few seconds (2:01), leaving only a low, meandering synth line and a *Piano Phase* sample until the percussion reenters (2:04). This double event seems associated with an EEG ISC peak but no significant CB activity. A similar compositional technique plays out before minute 3:00. Two coded lines before that time (2:36), all instruments drop out except for the *Piano Phase* sample. It goes on, unchanging, until lush pitched percussion (a marimba) and additional synth lines enter at 2:50 (the line just before minute 3:00 in **Figure 7D**). The ISC peaks in both the EEG and CB data anticipate the reentry of additional instrumental lines, possibly in line with the previously mentioned hazard function: an anticipation that something must be coming given the static situation.

We did not expect any significant EEG ISC peaks for Tremolo, with its static, stark content. We see only occasional, small peaks above significance in the EEG ISC, an initial pair of significant points in the CB ISC, and a low and relatively unchanging CB mean (**Figure 7E**; percentage of significant time-resolved EEG ISCs = 7.4%; percentage of significant time-resolved CB ISCs = 1.0%; see **Table 1**). We also note that in contrast to the other stimulus conditions, the time-resolved EEG ISC for this condition does not include a significant peak at the beginning of the excerpt. However, similar to the control condition in Kaneshiro et al. (2020), this RC1 differs in topography from the other conditions (**Figure 5A**) and is not statistically significant (**Figure 5B**), making for uneven comparison between the Tremolo EEG ISC time course and the EEG responses to the other stimuli.

Comparing the present percentages of significant time-resolved ISCs for EEG data in RC1 with those reported by Kaneshiro et al. (2021) shows that our highest EEG ISC (for Remix) eclipses their finding of 37% (in response to Elgar's cello concerto); our Original, Abrupt Change, and Segment Shuffle stimuli elicit higher percentages of significant ISC than their control condition (an envelope-scaled but otherwise temporally unstructured manipulation, 8%); and our Tremolo condition

FIGURE 7 | Time-resolved EEG ISC, CB ISC, and CB means for each condition. The top of each shaded gray region represents the 95th percentile of the corresponding null distribution. **(A)** Original: Dotted lines mark the start of phasing sections, solid lines mark the start of in-phase sections. **(B)** Abrupt Change: Solid lines mark the start of each new in-phase section. **(C)** Segment Shuffle: Light gray lines mark the start of each new segment. **(D)** Remix: Dashed lines mark musical events expected to be significant to listeners. **(E)** Tremolo.

TABLE 1 | For each stimulus: the percentage of statistically significant EEG and CB time-resolved ISC windows, description of statistically significant EEG and CB peaks and CB mean changes, and qualitative assessment of alignment between time-resolved EEG and CB ISC and time-resolved EEG ISC and CB mean.

Stimulus	% sig. EEG ISC	% sig. CB ISC	Desc. of sig. EEG ISC peaks	Desc. of sig. CB ISC peaks	Desc. of CB mean	Alignment: EEG and CB ISC	Alignment: EEG ISC and CB mean
Original	16.9	4.7	Most phasing sections	Unrelated to phasing	increase after each in-phase section, general decrease over time	Weak	Weak
Abrupt change	18.6	7.0	Most in-phase shifts	Most in-phase shifts	Increase after each in-phase section, general decrease over time	Moderate	Moderate
Segment shuffle	15.9	10.3	After many shifts	After many shifts	Infrequent increases after shifts, general decrease over time	Moderate	Moderate
Remix	45.6	26.0	After most musical events	After many musical events	After many musical events	Strong	Strong
Tremolo	7.4	1.0	Infrequent and small	Only at opening	Infrequent and small	Weak	Weak

approximates the percentage found for their control condition. Even the present Remix stimulus elicits a lower percentage of significant ISC windows, however, than RC1 ISC reported by Dmochowski et al. (2012) during film viewing, where over 50% of time windows contained significant ISC.

4. DISCUSSION

We used inter-subject correlation (ISC) as a measure of engagement with Steve Reich's *Piano Phase* and manipulated and remixed versions of the work. We expected the phasing process at the heart of *Piano Phase* to drive electroencephalographic (EEG) and continuous behavioral (CB) ISC. At the overall-level, we found no statistically significant differences between the EEG ISC for the original work and our phasing-related manipulations (Abrupt Change and Segment Shuffle). At the time-resolved level, however, we noted the impact of phasing and in-phase sections in the confluence of the start of phasing sections in the Original with significant EEG ISC and CB mean activity, in-phase sections in Abrupt Change with significant CB ISC and EEG ISC, and Segment Shuffle shifts with significant EEG ISC and CB mean activity. At the overall level, we found that Original, Abrupt Change, and Segment Shuffle had significantly higher EEG ISC levels than Tremolo (the extremely repetitive manipulation). The remixed version, more related to popular music, resulted in the highest ISC. Overall CB ISC results were similar, but Segment Shuffle had significantly higher ISC than Original, Abrupt Change, and Tremolo, and significantly lower than Remix. From this overall stance, EEG and CB ISC values for Original, Abrupt Change, and Segment Shuffle generally align with participants' behavioral ratings (with the single exception of ratings for "well ordered"). In general, we found alignment between behavioral and neural measures of engagement.[7]

---

[7]See **Supplementary Figure S5** for the correlations between mean CB and engagement rating values.

In addition to the overall alignment, we also noticed differences between EEG ISC and CB measures at the time-resolved level (We note that because EEG data collection always preceded CB data collection, it is possible that order effects play a role: perhaps participants focus more on lower-level features in the initial hearing when compared with subsequent hearings). Phasing sections in the Original, with their many and unpredictable onsets, elicited neural ISC but failed to generate significant CB ISC. Participants had higher CB mean ratings at the start of in-phase sections, perhaps returning attention to the stimulus when it emerged from complex phasing sections back toward unison clarity (in-phase sections)—a phenomenon not seen in the time-resolved EEG ISC data. We also noted the mix of alignment and independence between neural and behavioral measures in Abrupt Change, Segment Shuffle, and Remix, again with some significant EEG ISC unaccompanied by behavioral ISC. One way to understand the differences between EEG and CB measures is to connect them with the previous ISC finding that frequently and unexpectedly changing stimuli seem capable of driving correlated neural responses, perhaps pointing to a relationship between ISC and something like the orienting response (voluntary and automatic neural and behavioral responses to novel information, Sokolov, 1990; Sokolov et al., 2002). Dmochowski and colleagues reported relationships between EEG ISC and population ratings of Super Bowl commercials and found that an audio-visual stimulus with "repeated and jarring scene cuts" associated with "relatively strong neural reliability" drove ISC measures above population ratings (this stimulus was ultimately excluded in order to maintain stronger predictive performance of population ratings; Dmochowski et al., 2014, **Supplementary Note 3**). Ki et al. (2016) found that narratives in a foreign language elicited higher ISC than a narrative in the participants' native language. Using two films as stimuli, Poulsen et al. (2017) reported a significant correlation between ISC and average luminance difference, suggesting that ISC for their primary component of interest

"may indeed be driven by low-level visual evoked responses" (p. 5). Finally, Kaneshiro et al. (2020) noted that a stimulus manipulation in which measures of music were randomly re-ordered (and thus musically less meaningful but more surprising) resulted in higher EEG ISC than intact music. If EEG ISC is heavily influenced by such contrastive changes in acoustic features, perhaps some of them "break through" to the behavioral level and others do not. While this could explain the differences in time-resolved EEG ISC and CB ISC findings, we argue that it does not point to a break between ISC and engagement. Rather, the strong *overall* similarities between our EEG, CB, and question response data suggests that such contrastive changes may heavily contribute to participants' feeling of engagement and narrow the type of engagement that ISC indexes. Future studies could also work to remove the possible influence of order effects by alternating or randomizing the sequence of blocks in which EEG and CB data are collected (but without biasing participants about the nature of the experiment in the instructions for the CB block when it comes first).

A previous study reported decreased ISC (i.e., lower correlations of brain data between participants) when familiar music stimuli are repeated (Madsen et al., 2019). The authors argued that because EEG ISC tracks rapid responses to stimuli, it likely indexes more stimulus-driven responses, as opposed to cognitive elaborations that likely occur at longer temporal durations. One explanation of our findings is that highly repetitive music (such as minimalism and Reich's phasing process) will elicit lower engagement, and thus, lower ISC values. Our Tremolo condition offers an extreme test and seeming confirmation of this hypothesis. More varied stimuli still featuring high levels of repetition—i.e., Original, Abrupt Change, and Segment Shuffle—yielded higher EEG and CB ISC than Tremolo. Remix's frequently changing musical parameters resulted in rather high ISC. One could argue that the more repetitive the stimulus was, the less interesting it may have been, and thus, less engaging.

Yet, as some have pointed out (Madsen et al., 2019; Kaneshiro et al., 2021), ISC measures *shared* engagement. Put another way, ISC can only pick up on forms of engagement that unfold similarly between multiple participants. Other types of engagement, be they idiosyncratic, or only shared by a few participants, would not show up. The strongest empirical evidence for such a view of our current data comes from individual CB responses (**Supplementary Figure S2**). In said data, at least two participants (the highest two lines of raw data) show patterns of high and dynamic engagement in the Tremolo condition, a condition where we predicted and found very low EEG and CB ISC. Further supporting the notion of idiosyncratic engagement patterns is the fact that these two participants did not have unusually high EEG ISC responses,[8] nor is their behavior explainable via musical background: one had a musical background and one did not. Previous theoretical and empirical work bolsters the idea of multiple styles of engagement. The transportation and cognitive elaboration framework for

engagement (Green and Brock, 2000) posits two strands of engagement: transportation, where audience members are locked into the content of the art object, tracking details; and cognitive elaboration, where an observer or listener is prompted by the stimulus to reflect on the artwork, drawing connections with other experiences and other knowledge. David Huron's (2002) listening styles offer even more potential types or modes of engagement, ranging from mentally singing along to mentally reminiscing about musically associated memories. ISC would be unlikely to pick up on these listening styles equally, and it would be odd if a single measure could.

Some cognitive science of music scholars have argued that repetition could augment individualized, internally focused experiences by gradually demanding less processing power and attention over time. Such a process may open up reflective space for listeners (Margulis, 2012, 2014).[9] (This is in contrast with the type of engagement that might occur during dramatic moments like the beat drop in the first minute of Remix.) In *Piano Phase*, such a trajectory could be cyclical, with listeners drifting off into individual experience and tugged back into the details of the ongoing external stimulus events by changes in the music. If enough participants were drawn back to the stimulus details at the same time, neural responses could become sufficiently correlated to produce an ISC peak (perhaps something like the peak around minute 3:00 in the Original EEG ISC time-resolved data). In this line of thought, musicologists and music theorists have noted the long trajectories of expectation formation in minimalist music such as Reich's. Cadences in tonal music (i.e., the ends of phrases) often drive and ultimately resolve such expectations (what key are we in? where are we in the phrase? what harmonic and melodic activity is likely to come next?). Cadences and their accompanying harmonic trajectories are also present in minimalism but often in a stretched out form (Fink, 1996). Some listeners may lose interest along the way, while others may be drawn into granular detail and vary in what layer of granularity they are caught up in. Perhaps most move from state to state: For examples of the former situation, two participants in the present study noted that the Tremolo stimulus was difficult to listen to—"intense" in the words of one. Another participant stated that to them the stimuli were "all the same but with different layers."

One potential route forward for this line of research would be to use the current results to hypothesize quantifiable musical features that may be driving time-resolved EEG and CB ISC peaks (Alluri et al., 2012). This could lead to a fruitful exploration of how far specific compositional techniques such

---

[8]We note the caveat that the RC1 for Tremolo was not statistically significant and did not appear to be auditory in nature.

[9]Margulis's account could be framed as a type of cognitive elaboration, where listeners are able to chunk or process ever longer swaths of the highly repetitive stimulus: "...the horizon of involvement widens with additional exposures, so that the music doesn't seem to be coming at the listener in small bits, but rather laying out broader spans for consideration" (Margulis, 2014, p. 9). In the original Reich piece, this process could happen within each in-phase unit: the musical material is short enough that even within a few repetitions, some listeners may "zoom out." If a listener is engaged in detailed listening, phasing sections offer little exact repetition, but, if engaged in a temporally broader manner, the pattern of in-phase to phasing could similarly encourage gradual zooming out as the listener groks the structure of the composition.

as phasing generalize into other repetition-based techniques like Philip Glass's additive and subtractive modular technique (York, 1981) or electronic dance music (Solberg and Dibben, 2019). It also reveals new layers of detail for scholars who work on the repertoire—a testing ground for theories of how the music can function for individuals. On that front, this study suggests important follow up research. For instance, alpha activity is thought to reflect meditative states (Lee et al., 2018). Therefore, alternative approaches to analyzing the EEG data— e.g., by assessing alpha power, or correlation thereof—may prove more appropriate measures for indexing listener states while listening to minimalist music. We might hypothesize that when participants are diversely engaged with a stimulus, a similar psychological state may be shared—but one that is better indexed by other means than EEG ISC. As alpha activity has been shown to index multiple states in varying locations (Nunez et al., 2001; Başar, 2012; Lee et al., 2018), future research could also include interviews with music listeners to provide complementary insights into inter-individual differences in music listening. Such mixed-methods work could reveal patterns for calm vs. bored listeners or time periods of boredom, interest, and relaxation. While we limited ourselves to exploring a general type of engagement, future research could work to distinguish between types of engagement and even diverse forms on non-engagement (distinguishing boredom from confusion, for example).

Our hypothesis that the core compositional feature of Reich's *Piano Phase* would differentially drive engagement (measured using inter-subject correlation, ISC) between an excerpt of that work and conditions that manipulated the phasing process was consistent with time-resolved EEG and behavioral ISC data which showed that the timing of key musical moments often corresponded with these measures of engagement. Overall, our participants' neural, continuous behavioral, and question responses show that a popular-music style Remix of Reich's *Piano Phase* was more engaging than the original work and two conditions that manipulated its core compositional technique. In turn, these three stimuli were more engaging than an intensely repetitive condition that featured no compositional changes. Although research continues to unravel the specifics of what EEG ISC measures when participants are presented with musical stimuli, we found that participants EEG ISC, CB ISC, and question responses broadly align: evidence that *some* type of engagement is tracked by EEG ISC. We propose that the nature of the engagement indexed by EEG ISC with musical stimuli seems to be a mixture of attention, acoustic features, and the level of contrastive change in those acoustic features.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://purl.stanford.edu/kt396gb0630.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Stanford University Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

TD, DN, JB, and BK designed the experiment. TD, NG, JB, and BK created the stimuli. DN and BK created participant interfaces for the experiment. TD and DN collected the data. TD, DN, and BK curated the data. TD, JD, and BK specified formal and statistical analyses. TD and BK analyzed the data, created the visualizations, and drafted the original manuscript. DN, NG, JD, and JB reviewed and edited the manuscript. JB and BK supervised the research. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2021.702067/full#supplementary-material

REFERENCES

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

Alluri, V., Toiviainen, P., Jääskeläinen, I. P., Glerean, E., Sams, M., and Brattico, E. (2012). Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm. *Neuroimage* 59, 3677–3689. doi: 10.1016/j.neuroimage.2011.11.019

Başar, E. (2012). A review of alpha activity in integrative brain function: fundamental physiology, sensory coding, cognition and pathology. *Int. J. Psychophysiol.* 86, 1–24. doi: 10.1016/j.ijpsycho.2012.07.002

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., et al. (2012). *Package 'lme4'*. Vienna: CRAN. R Foundation for Statistical Computing.

Becker, J. (2004). *Deep Listeners: Music, Emotion, and Trancing, Vol. 1.* Bloomington, IN: Indiana University Press.

Benjamini, Y., and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* 29, 1165–1188. doi: 10.1214/aos/1013699998

Brainard, D. (1997). The psychophysics toolbox. *Spat. Vis.* 10, 433–436. doi: 10.1163/156856897X00357

Cameron, D., Potter, K., Wiggins, G., and Pearce, M. (2017). Perception of rhythmic similarity is asymmetrical, and is influenced by musical training, expressive performance, and musical context. *Timing Time Percept.* 5, 211–227. doi: 10.1163/22134468-00002085

Cameron, D. J., Zioga, I., Lindsen, J. P., Pearce, M. T., Wiggins, G. A., Potter, K., et al. (2019). Neural entrainment is associated with subjective groove and complexity for performed but not mechanical musical rhythms. *Exp. Brain Res.* 237, 1981–1991. doi: 10.1007/s00221-019-05557-4

Cohen, S. S., Madsen, J., Touchan, G., Robles, D., Lima, S. F., Henin, S., et al. (2018). Neural engagement with online educational videos predicts learning performance for individual students. *Neurobiol. Learn. Mem.* 155, 60–64. doi: 10.1016/j.nlm.2018.06.011

Cohen, S. S., and Parra, L. C. (2016). Memorable audiovisual narratives synchronize sensory and supramodal neural responses. *eNeuro* 3:ENEURO.0203-16.2016. doi: 10.1523/ENEURO.0203-16.2016

Cohn, R. (1992). Transpositional combination of beat-class sets in steve reich's phase-shifting music. *Perspect. New Music* 30, 146–177. doi: 10.2307/3090631

Dauer, T. (2020). *The Varieties of Minimalist Experience: The Roles of Psychological States in the Reception of American Minimalism During the Long Sixties* (Ph.D. thesis). Stanford University.

Dauer, T., Nerness, B., and Fujioka, T. (2020). Predictability of higher-order temporal structure of musical stimuli is associated with auditory evoked response. *Int. J. Psychophysiol.* 153, 53–64. doi: 10.1016/j.ijpsycho.2020.04.002

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. *Nat. Commun.* 5, 1–9. doi: 10.1038/ncomms5567

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. doi: 10.1016/j.neuroimage.2014.12.078

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? *Front. Hum. Neurosci.* 6:112. doi: 10.3389/fnhum.2012.00112

Epstein, P. (1986). Pattern structure and process in steve reich's" piano phase". *Musical Q.* 72, 494–502. doi: 10.1093/mq/LXXII.4.494

Farbood, M. M., Heeger, D. J., Marcus, G., Hasson, U., and Lerner, Y. (2015). The neural processing of hierarchical structure in music and speech at different timescales. *Front. Neurosci.* 9:157. doi: 10.3389/fnins.2015.00157

Ferree, T. C., Luu, P., Russell, G. S., and Tucker, D. M. (2001). Scalp electrode impedance, infection risk, and EEG data quality. *Clin. Neurophysiol.* 112, 536–544. doi: 10.1016/S1388-2457(00)00533-2

Fink, R. (2005). *Repeating Ourselves*. Berkeley: University of California Press.

Fink, R. W. (1996). *"Arrows of Desire": Long-Range Linear Structure and the Transformation of Musical Energy* (Ph.D. thesis). University of California, Berkeley.

Green, M. C., and Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *J. Pers. Soc. Psychol.* 79, 701. doi: 10.1037/0022-3514.79.5.701

Hartenberger, R. (2016). *Performance Practice in the Music of Steve Reich*. Cambridge: Cambridge University Press.

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., and Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science* 303, 1634–1640. doi: 10.1126/science.1089506

Henahan, D. (1970). Steve Reich presents a program of pulse music at Guggenheim. *The New York Times*.

Huron, D. (2002). "Listening styles and listening strategies," in *Society for Music Theory Annual Conference*. (Columbus, Oh).

Ihaka, R., and Gentleman, R. (1996). R: A language for data analysis and graphics. *J. Computat. Graph. Stat.* 5, 299–314. doi: 10.1080/10618600.1996.10474713

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

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. doi: 10.1016/j.neuroimage.2020.116559

Kaneshiro, B., Nguyen, D. T., Norcia, A. M., Dmochowski, J. P., and Berger, J. (2021). Inter-subject EEG correlation reflects time-varying engagement with natural music. *bioRxiv*. doi: 10.1101/2021.04.14.439913

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

Ki, J. J., Kelly, S. P., and Parra, L. C. (2016). Attention strongly modulates reliability of neural responses to naturalistic narrative stimuli. *J. Neurosci.* 36, 3092–3101. doi: 10.1523/JNEUROSCI.2942-15.2016

Lancaster, G., Iatsenko, D., Pidde, A., Ticcinelli, V., and Stefanovska, A. (2018). Surrogate data for hypothesis testing of physical systems. *Phys. Rep.* 748, 1–60. doi: 10.1016/j.physrep.2018.06.001

Lee, D. J., Kulubya, E., Goldin, P., Goodarzi, A., and Girgis, F. (2018). Review of the neural oscillations underlying meditation. *Front. Neurosci.* 12:178. doi: 10.3389/fnins.2018.00178

Lerner, Y., Honey, C. J., Silbert, L. J., and Hasson, U. (2011). Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *J. Neurosci.* 31, 2906–2915. doi: 10.1523/JNEUROSCI.3684-10.2011

Losorelli, S., Nguyen, D. T., Dmochowski, J. P., and Kaneshiro, B. (2017). "NMED-T: a tempo-focused dataset of cortical and behavioral responses to naturalistic music," in *Proceedings of the 18th International Society for Music Information Retrieval Conference* (Suzhou), 339–346.

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. *Sci. Rep.* 9, 1–8. doi: 10.1038/s41598-019-40254-w

Margulis, E. H. (2012). Musical repetition detection across multiple exposures. *Music Percept.* 29, 377–385. doi: 10.1525/mp.2012.29.4.377

Margulis, E. H. (2014). *On Repeat: How Music Plays the Mind*. Oxford: Oxford University Press.

Mertens, W. (1983). *American Minimal Music: La Monte Young, Terry Riley, Steve Reich, Philip Glass*. London: Kahn & Averill.

Nakagawa, S., and Cuthill, I. C. (2007). Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biol. Rev.* 82, 591–605. doi: 10.1111/j.1469-185X.2007.00027.x

Nobre, A. C., Correa, A., and Coull, J. T. (2007). The hazards of time. *Curr. Opin. Neurobiol.* 17, 465–470. doi: 10.1016/j.conb.2007.07.006

Norman, G. (2010). Likert scales, levels of measurement and the "laws"9D of statistics. *Adv. Health Sci. Educ.* 15, 625–632. doi: 10.1007/s10459-010-9222-y

Nunez, P. L., Wingeier, B. M., and Silberstein, R. B. (2001). Spatial-temporal structures of human alpha rhythms: Theory, microcurrent sources, multiscale measurements, and global binding of local networks. *Hum. Brain Mapp.* 13, 125–164. doi: 10.1002/hbm.1030

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* 24, 147. doi: 10.1037/pmu0000044

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

Potter, K., Wiggins, G. A., and Pearce, M. T. (2007). Towards greater objectivity in music theory: Information-dynamic analysis of minimalist music. *Musicae Sci.* 11, 295–324. doi: 10.1177/102986490701100207

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. *Sci. Rep.* 7, 1–9. doi: 10.1038/srep43916

Prichard, D., and Theiler, J. (1994). Generating surrogate data for time series with several simultaneously measured variables. *Phys. Rev. Lett.* 73, 951–954. doi: 10.1103/PhysRevLett.73.951

Quinn, I. (2006). Minimal challenges: Process music and the uses of formalist analysis. *Contemporary Music Rev.* 25, 283–294. doi: 10.1080/07494460600726537

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Reich, S. (1987). *Reich "Early Works"*. New York, NY: Double Edge.

Reich, S. (1999). *Reich Remixed*. New York, NY: Arthrob/Nonesuch.

Reich, S. (2009). *Writings on Music, 1965-2000*. Oxford: Oxford University Press.

Rockwell, J. (1973). Records: Roiling work: Reich's 'Four Organs,' which created a stir at concert, is on Angel Disk Hyman's Piano. *New York Times*.

Schubert, E., Vincs, K., and Stevens, C. J. (2013). Identifying regions of good agreement among responders in engagement with a piece of live dance. *Empir. Stud. Arts* 31, 1–20. doi: 10.2190/EM.31.1.a

Simony, E., Honey, C. J., Chen, J., Lositsky, O., Yeshurun, Y., Wiesel, A., et al. (2016). Dynamic reconfiguration of the default mode network during narrative comprehension. *Nat. Commun.* 7, 1–13. doi: 10.1038/ncomms12141

Sokolov, E. (1990). The orienting response, and future directions of its development. *Pavlov. J. Biol. Sci.* 25, 142–150.

Sokolov, E. N., Spinks, J. A., Näätänen, R., and Lyytinen, H. (2002). *The Orienting Response in Information Processing*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Solberg, R. T., and Dibben, N. (2019). Peak experiences with electronic dance music: Subjective experiences, physiological responses, and musical characteristics of the break routine. *Music Percept.* 36, 371–389. doi: 10.1525/mp.2019.36.4.371

Strongin, T. (1969). Is timelessness out of style? *New York Times* 21.

Tecce, J. J. (1972). Contingent negative variation (CNV) and psychological processes in man. *Psychol. Bull.* 77, 73. doi: 10.1037/h0032177

Tucker, D. M. (1993). Spatial sampling of head electrical fields: the geodesic sensor net. *Electroencephalogr. Clin. Neurophysiol.* 87, 154–163. doi: 10.1016/0013-4694(93)90121-B

Wilson, S. M., Molnar-Szakacs, I., and Iacoboni, M. (2008). Beyond superior temporal cortex: intersubject correlations in narrative speech comprehension. *Cereb. Cortex* 18, 230–242. doi: 10.1093/cercor/bhm049

Winter, B. (2013). Linear models and linear mixed effects models in r with linguistic applications. *arXiv preprint* arXiv:1308.5499.

York, W. (1981). Form and process in two pages of philip glass. *Sonus* 1, 28–50.