TOWARD AN OBJECTIVE MEASURE OF LISTENER ENGAGEMENT WITH NATURAL MUSIC USING INTER-SUBJECT EEG CORRELATION

Blair Kaneshiro,[*] *Jacek P. Dmochowski,*[#] *Anthony M. Norcia,*[#] *and Jonathan Berger*[*]

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

[#]Department of Psychology, Stanford University, Stanford, CA, USA

ABSTRACT

This study extends existing research on inter-subject correlations (ISCs) of brain responses as a measure of engagement, with a focus on the relationship between the structural coherence of a musical stream and the listener's degree of engagement with it. EEG was recorded while subjects listened to naturalistic music (popular Hindi songs) presented in original versions and in temporally disrupted, phase-scrambled conditions. ISCs were computed from the EEG data using Reliable Components Analysis. Overall, original versions of songs yielded significantly higher ISCs than phase-scrambled versions, and were also rated as more pleasant, well ordered, and interesting by subjects. The most reliable spatial component extracted from responses to the original songs concurs with past EEG findings involving naturalistic music. The time course of the ISCs is resolved at a musically relevant time scale. The sum of our findings suggests that ISCs show promise toward finding time-critical measures of engagement and attention in typically noisy EEG signals, and, specifically, as a means to find correlations between structural features of music and brain responses in listeners. We discuss possible links between heightened ISCs and regions of musical interest, and implications for future research.

1. INTRODUCTION

Listening to music is generally regarded to constitute a pleasurable endeavor. The sense of pleasure and the degree to which a listener engages with the music are interdependent. Varying degrees of engagement can occur whether the music is the focus of attention (for example when heard in a concert hall), or is the accompaniment to other activities such as socialization or work. Even when music is heard passively as background to another task, arousal and attention levels can fluctuate in response to varying musical and acoustical features. Many of us are familiar with the feeling of engagement with music, whether hearing a song for the first time or the thirtieth. But how is engagement represented in the brain?

The use of inter-subject correlations (ISCs) as a measure of engagement stems from the reasoning that neural activity at any given time comprises exogenous (stimulus-driven) and endogenous (internally generated) components. When engagement with a stimulus increases, so too does the exogenous component of the brain response as peripheral processing decreases; therefore, responses become more correlated across subjects.

To date, ISCs have been examined in a variety of fMRI studies employing naturalistic stimuli including video (Hasson et al., 2004), speech and non-speech sounds (Honey et al., 2012; Boldt et al., 2013), and music (Abrams et al., 2013). In a recent EEG study by Dmochowski et al. (2012), a novel signal-decomposition method was devised for extracting maximally correlated components from neural responses to videos. This method was then used to quantify viewer engagement on a finer time scale, which allowed periods of heightened engagement to be traced back to contextually salient scenes in the videos.

In the present study, we combine the methodology of Dmochowski et al. (2012) and Abrams et al. (2013), using ISCs of EEG responses as a measure of engagement with naturalistic music. We do this by identifying correspondences among brain responses to naturalistic music that is presented in original and control conditions. Behavioral ratings of the stimuli serve as a point of comparison for the EEG results. We employ 'scramble' paradigms similar to those used in past studies (Abrams et al., 2013; Dmochowski et al., 2012; Levitin & Menon, 2003) to disrupt the temporal coherence of the musical stimuli, with a current focus on very small-scale temporal manipulations achieved through phase scrambling. While such low-level temporal units do not necessarily constitute meaningful structural units in a musical sense, we begin our line of research at this level, as a basis for building up to larger structural components of music (e.g., measures, phrases, and song parts).

Our aim in the current study is to validate ISCs of EEG-recorded brain responses as a reliable measure of engagement. Our broader goal aims to establish an objective measure of listener engagement with naturalistic music, which could provide a useful supplement to existing physiological and self-report measures. EEG proves to be a useful modality for this, as its temporal resolution is well matched to that of music, meaning that there is potential to draw connections between heightened ISCs computed over small time windows and corresponding features of the driving stimulus. At the same time, the ISC source-selection technique used here requires a listener to hear each stimulus only once (provided the stimuli are sufficiently long), and does not rely upon event-related averaging (Ben-Yakov et al., 2012). This allows us to include a greater variety of stimuli, and more importantly, captures the real-world experience of engaging with a musical excerpt in a single listen.

2. METHODS

2.1 Stimuli

Song selection. We used an unorthodox stimulus set derived from four songs from recent popular Hindi-language films. We did so because we sought songs in a (relatively) tonal idiom that have appealed to a massive audience and would presumably engage a listener, yet would be unfamiliar to our experiment subjects. While lyrics form an integral component of engagement with popular music, we wanted to avoid in the current study any effects of the semantic content of lyrics. We additionally concluded from a previous pilot study that non-English lyrics would be less unsettling to the subject in control conditions that noticeably disrupted the flow of lyrics.

Songs chosen for the study (summarized in Table 1) were released as singles from their respective films, or featured prominently in the films. All are sung in Hindi dialects with minimal English lyrics;[1] include clear verse and chorus elements; use a steady beat throughout; and comprise regularly structured phrases. We acknowledge some differences in vocalization and instrumentation, but attempted to maximize the 'Westernness' of the song set, in style and instrumentation, given the above constraints.

	Song Name	Film	Year	Length
Song 1	Ainvayi Ainvayi	*Band Baaja Baaraat*	2010	4:27
Song 2	Daaru Desi	*Cocktail*	2012	4:30
Song 3	Haule Haule	*Rab Ne Bana Di Jodi*	2008	4:24
Song 4	Malang	*Dhoom 3*	2013	4:33

Table 1. Songs used in the study.

Control stimuli. The phase-scrambled control stimuli were created as described in Abrams et al. (2013), by randomizing the phase response at each frequency bin in a song's Fourier transform to a value between 0 and 2π, then transforming the signal back to the time domain. This manipulation effectively washes out the temporal structure of a song while preserving its magnitude spectrum. In addition—while we do not cover these results in the present study—beat-shuffled versions of the songs were created using freely available beat-tracking code (Ellis, 2007). Beat onset times were detected in the audio and used as segmentation points for permuting the beats of the songs. This manipulation thus results in a stimulus that retains the steady beat of the original song, but does not provide a coherent narrative for the listener in terms of phrase structure or melodic/harmonic continuity. Both types of control stimuli were created using Matlab.

[1] Songs included at most one- or two-word interjections or one-syllable word substitutions in English.

2.2 Participants and Procedure

Twelve right-handed subjects aged 19-38 (mean age 28.17 years; 2 female) participated in the experiment.[2] Formal musical training ranged from 0-18 years (mean 9.63 years), though all subjects either had formal training or had taught themselves to play an instrument. Participants were unfamiliar with the songs used in the study, and no participants spoke Hindi, listened to Hindi music, or watched Hindi films. All participants listened at least occasionally to popular music in English.

Each stimulus was presented once in its entirety, in random order. Stimuli were delivered at a comfortable listening level through magnetically shielded Genelec 1030A speakers while subjects were seated in a darkened, electrically and acoustically shielded booth. Subjects were instructed to attend to the stimuli but performed no behavioral task while audio was playing. Following each stimulus, subjects answered the following questions on a scale of 1-9:

- How pleasant was the excerpt?
- How well-ordered was the excerpt?
- How much of the excerpt was interesting?

Stimulus delivery and collection of behavioral responses was performed using Neurobehavioral Systems Presentation software. 128-channel EEG (Electrical Geodesics, Inc.) was recorded at a sampling rate of 1 kHz, referenced to the vertex.

2.3 Data Analysis

Preprocessing. Data were passband filtered between 0.3-50 Hz and downsampled by a factor of 4 using EGI's Net Station software; all subsequent analyses were performed in Matlab. Data were epoched, EOG channels were computed, and bad electrodes, as well as electrodes covering the face, were excluded from further analysis. Eye artifacts were removed using EEGLAB's extended Infomax ICA (Jung et al., 1998, Delorme & Makeig, 2004). Following this, DC offset was subtracted from each channel, and data matrices were converted to average reference. Transient samples whose magnitude exceeded 4 standard deviations of its respective channel's mean power were set to NaN.

Inter-subject correlations. ISCs were computed using Reliable Components Analysis (RCA), as described in Dmochowski et al. (2012). This technique is similar to Principal Components Analysis (PCA) and Independent Components Analysis (ICA), differing in that it optimizes reliability of data records (as opposed to variance explained or statistical independence, respectively). Namely, given a set of space-time data records (i.e., one for each subject), the algorithm computes projections of the data exhibiting maximal ISC. In other words, the criterion being maximized is the sum of pairwise ISCs among all unique subject pairs. For the analysis, we computed the Reliable Components (RCs) separately

[2] Data from a thirteenth subject were excluded from analysis due to gross artifacts.

for original and phase-scrambled tracks, pooling covariances from all songs/subjects in the computation.

For each song, we then projected the neural responses of all subjects onto the corresponding first RC and computed the mean (across unique subject pairs) ISC time course in time windows of 10 seconds, with a 2-second shift between successive windows. This yields a time-resolved measure of ISC for all songs. For each time window, we performed a Wilcoxon signed-rank test to compute the probability that the measured ISCs were drawn from a zero-median distribution. This yields a proportion of time windows that exhibited significant ISCs (p<0.05) throughout the song duration.

3. RESULTS

3.1 Behavioral Results

Subjects' behavioral ratings of the stimuli along the dimensions of Pleasantness, Order, and Interestingness are shown in Figure 1. Original versions received higher ratings than phase-scrambled versions along all three dimensions; a nonparametric test of statistical significance finds the difference in ratings between song versions to be significant for each dimension (Mann-Whitney U test, $p<10^{-14}$). We observe a number of outlier ratings of phase-scrambled stimuli, especially for Order; we believe this occurred because we deliberately formulated that prompt in an ambiguous manner, and subjects could interpret the phase-scrambled stimuli as being either highly ordered (lacking variation over time) or not ordered at all (no discernible structure).

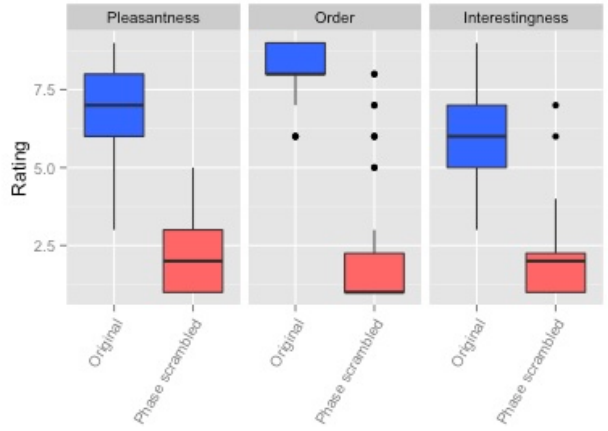

Figure 1. Subjects' behavioral ratings of original (blue) and phase-scrambled (red) versions of songs, along dimensions of Pleasantness (left), Order (middle), and Interestingness (right).

3.2 EEG Results

From the subject- and song-aggregated covariance matrices, we computed the first Reliable Component (RC1). Figure 2 shows the corresponding projections of the extracted activity—the so-called "forward-model" (Parra et al., 2005)—for original (left) and phase-scrambled (right) versions of the songs. RC1 for the original versions of the songs is marked by poles over frontocentral cortex and bilateral temporoparietal regions. Without the availability of detailed anatomical source inverses, it is difficult to speculate on the neural origins of the observed component. However, we do note that its topography is similar to the first Principal Component found in a published EEG study using naturalistic music (Schaefer et al., 2011), and is consistent with bilateral dipoles in temporal cortex. Meanwhile, the first RC stemming from data recorded during phase-scrambled stimuli exhibits a disparate topography with poles in the inferior occipital and occipitotemporal cortices, but lacking physiological plausibility.

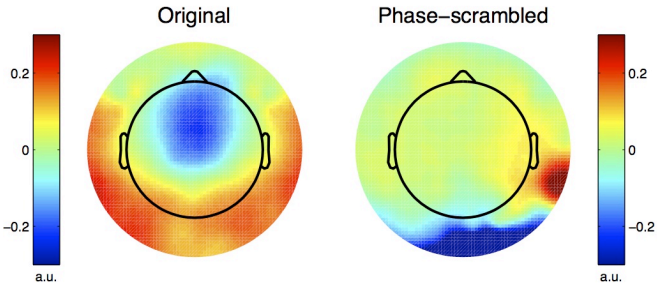

Figure 2. Forward-model projections of RC1 (most reliable component) extracted from responses to the original versions (left) and phase-scrambled versions (right) of stimuli.

The brain responses in the space of these components were used to compute the ISCs as a function of time for each stimulus, using 10-second temporal windows of data that advanced in 2-second increments. Results, grouped by song, are plotted as a function of time in Figure 3. Inspection of the plots suggests that phase-scrambled versions exhibit lower ISCs than original versions, for all songs.

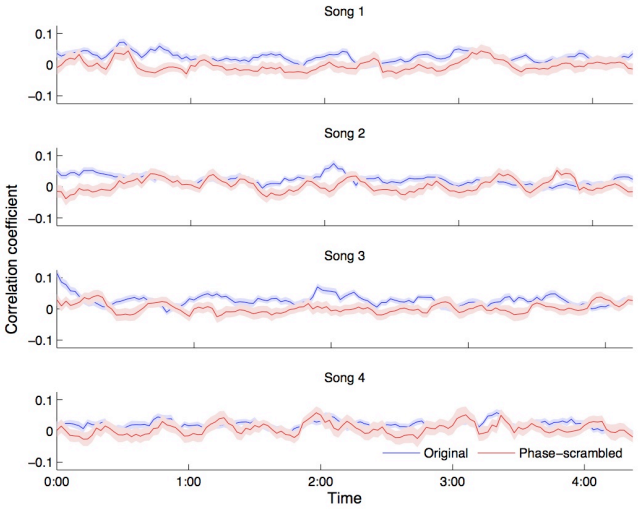

Figure 3. ISCs computed from the most reliable component (RC1) of original (blue) and phase-scrambled (red) versions of individual songs, plotted as a function of time. RC1 was computed separately for the original and phase-scrambled versions in order to maximize the ISCs of all EEG records pertaining to that condition.

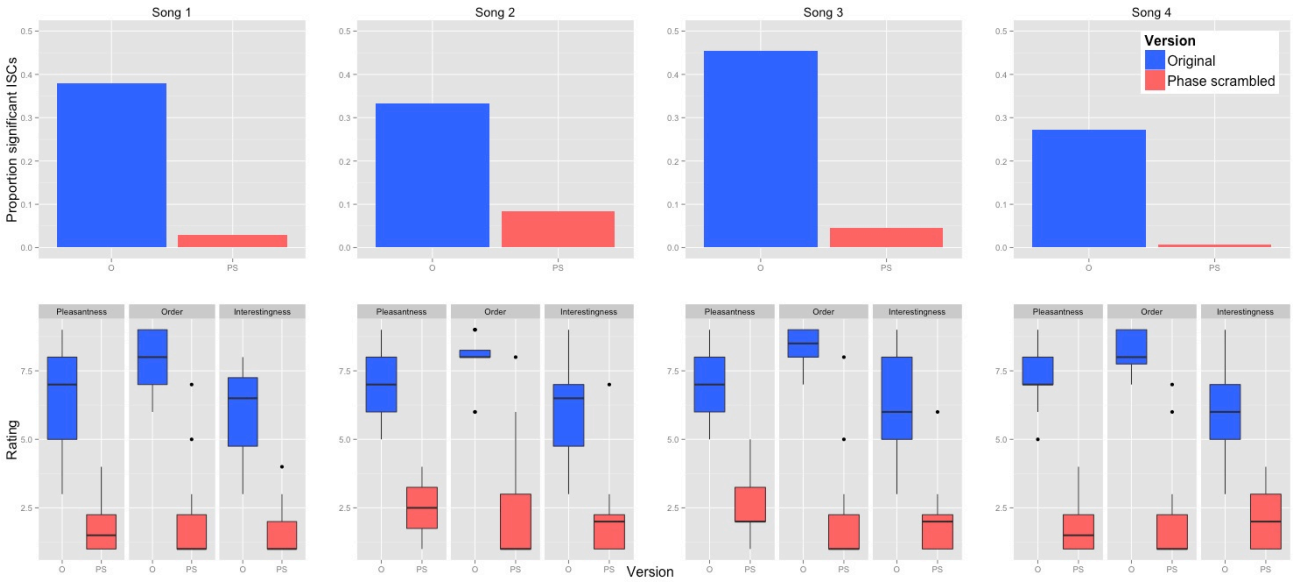

Figure 4. Top: Proportion of significant ISCs in the four songs, for original (blue) and phase-scrambled (red) versions. Bottom: Behavioral ratings of Pleasantness, Order, and Interestingness for each version of the songs.

Finally, we present the proportion of significant ISCs for each stimulus. The top portion of Figure 4 shows the proportion of significant ISCs for each stimulus condition of each song; corresponding behavioral ratings for each song are shown in the bottom portion of the figure. We see that the proportion of significant ICSs of original versions is always higher than the phase-scrambled versions, as are the behavioral ratings.

4. DISCUSSION

This work constitutes an initial step toward developing an objective, time-resolved measure of listener engagement with naturalistic music using ISCs and EEG. In this study, we used a single-listen paradigm to present popular yet novel songs in their entirety, both unaltered as well as temporally distorted through phase scrambling. The most reliable component obtained from EEG responses to original versions of songs using RCA appears to reflect bilateral dipoles in temporal cortex and is consistent with past EEG findings using naturalistic music. Subjects' behavioral ratings of the stimuli concur with ISC results: Original versions of songs were rated as more pleasant, well ordered, and interesting than phase-scrambled versions, while also driving higher ISCs in the brain responses.

We speculate that moments of high engagement, as measured by high temporally resolved ISCs (Figure 3), may correspond to particularly arousing points in the original songs. Indeed, our early observations do suggest a marked co-variation between peaks in the ISCs and structural demarcations of the songs correlating to various musical dimensions.[3] If this proves to be the case, then it may be true that ISCs can serve to index listener engagement with naturalistic music with high temporal resolution, over just a single listen.

One question that arises is whether certain factors may draw attention without necessarily fostering engagement. For example, to what degree do predictable spikes in amplitude of an audio excerpt (as in many metrical accents) drive engagement? As we proceed with our research, we hope to explore and disentangle the interplay of arousal, attention, and more broadly, engagement. Dmochowski et al. (2012) characterize engagement as "emotionally laden attention." By incorporating subjective ratings with objective measures obtained from the EEG data, we hope to pursue this supposition in the context of listening to music.

Broadly speaking, using ISCs to find significant features in noisy EEG data offers a method to use ecologically valid, naturalistic stimuli—stimuli of greater duration and complexity—to study attention, arousal, and engagement. This study provides a referential point of departure for future research aimed at exploration of the effects of temporal organization and coherence on musical engagement. A next step is the analysis of higher-order control stimuli (e.g., re-orderings at the beat and phrase level). Furthermore, ISCs suggest a potential meaningful measure of engagement based upon other musical dimensions. Future studies will explore the effect of repeated hearings on engagement, effects of lyrics and language, and the effect of familiar versus novel songs as stimuli. One intriguing question is whether ISC-based

[3] Examples can be found at http://ccrma.stanford.edu/groups/meri

engagement detection might be used predictively to assess likely audience reaction and behavior on a large scale (following forthcoming work by Dmochowski et al. (2014)). To this end, we intend to create a parallel research track using large-scale datasets tracking music engagement, and integrate statistical observations and models with inter-subject correlations in music engagement experiments.

5. ACKNOWLEDGMENT

This research was supported by the Wallenberg Network Initiative: Culture, Brain, and Learning. The authors wish to thank Sophia Laurenzi and Steven Losorelli for their assistance with data collection, and Shubhabrata Sengupta for his insights into characterizing the stimuli.

6. REFERENCES

Abrams, D. A., Ryali, S., Chen, T., Chordia, P., Khouzam, A., Levitin, D. J., & Menon, V. (2013). Inter-subject synchronization of brain responses during natural music listening. *European Journal of Neuroscience.*

Ben-Yakov, A., Honey, C. J., Lerner, Y., & Hasson, U. (2012). Loss of reliable temporal structure in event-related averaging of naturalistic stimuli. *NeuroImage, 63*, 501-506.

Boldt, R., Malinen, S., Seppä, M., Tikka, P., Savolainen, P., Hari, R., & Carlson, S. (2013). Listening to an audio drama activates two processing networks, one for all sounds, another exclusively for speech. *PLoS One, 8(5)*.

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

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

Dmochowski, J. P., Bezdek, M., Abelson, B., Johnson, J., Schumacher, E., and Parra, L. C. (to appear). Audience preferences are predicted by temporal reliability of neural processing. *Nature Communications*.

Ellis, D. P. W. (2007). Beat tracking by dynamic programming. *Journal of New Music Research, 36(1)*, 51-60. http://labrosa.ee.columbia.edu/projects/beattrack/

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science, 303(5664)*, 1634-1640.

Honey, C. J., Thompson, C. R., Lerner, Y., & Hasson, U. (2012). Not lost in translation: Neural responses shared across languages. *The Journal of Neuroscience, 32(44)*, 15277-15283.

Jung, T. P., Humphries, C., Lee, T. W., Makeig, S., McKeown, M. J., Iragui, V., & Sejnowski, T. J. (1998). Extended ICA removes artifacts from electroencephalographic recordings. *Advances in Neural Information Processing Systems, 10*, 894-900.

Levitin, D. J., & Menon, V. (2003). Musical structure is processed in "language" areas of the brain: A possible role for Brodmann Area 47 in temporal coherence. *NeuroImage, 20*, 2142-2152.

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

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